

T H E S E

présentée

**A L'UNIVERSITÉ PARIS-SUD
CENTRE D'ORSAY**

pour obtenir

**LE GRADE DE DOCTEUR INGÉNIEUR
EN AMÉLIORATION DES PLANTES**

par

Louis MARCHAIS

**CONTRIBUTION AUX MÉTHODES D'ANALYSE EN GÉNÉTIQUE QUANTITATIVE
AUX TECHNIQUES DE SÉLECTION
ET A LA CONNAISSANCE DU PETIT MIL
(*PENNISETUM TYPHOIDES*)**

soutenu le 24 octobre 1979 devant la Commission d'examen :

MM. DEMARLY Y.	Président
GALLAIS A.	Rapporteur
PERNES J.	Examineurs
COMBES D.	

**O.R.S.T.O.M.
PARIS
1979**

ORSAY

n° d'ordre : 406

UNIVERSITE DE PARIS-SUD

CENTRE D'ORSAY

THESE

présentée

Pour obtenir

LeTITRE..... de DOCTEUR ..INGENIEUR..

SPECIALITE : AMELIORATION DES PLANTES

PAR

M. MARCHAIS Louis



SUJET :

CONTRIBUTION AUX METHODES D'ANALYSE EN GENETIQUE QUANTITATIVE, AUX TECHNIQUES DE SELECTION ET A LA CONNAISSANCE DU PETIT MIL (pennisetum typhoides)

soutenue le 24 octobre 1979 devant la Commission d'examen

MM. DEMARLY Président

PERNES

COMBES

GALLAIS

.....

CONTRIBUTION AUX METHODES D'ANALYSE EN
GENETIQUE QUANTITATIVE, AUX TECHNIQUES DE SELECTION
ET A LA CONNAISSANCE DU PETIT MIL (P. TYPHOIDES)

L. MARCHAIS

S U M M A R Y

The present thesis reports informations obtained in a pearl millet breeding work accomplished at Bambey (Sénégal) during the period 1972-1975, utilizing diallel and triple test crosses.

These informations are related to the value of dwarf millet, the quantitative genetics in pearl millet, the methodology of selection.

The anova of the diallel table with the Eberhart model and the Hayman hypotheses provides immediately an estimation of the genetical parameters h^2 , D , H_1 , H_2 .

The theoretical relation between heterosis and R_2 distance of Hanson-Casas is obtained that takes into account epistosis between gene pairs et covariances between locus states. It allows a new estimation of Hayman parameters, the detection of epistasis either duplicate or complementary, an understanding of the strong observed correlations between heterosis and distance. That distance appears a very useful tool for prediction of heterosis.

Two breeding methods are presented : one to maximize a synthetic variety, the other to estimate the yield of a genotype from measures on individual plants of grain and vegetation characters.

Most the twelve characters studied show duplicate epistasis. Characters of architecture show partial or complete apparent dominance. Grain characters show superdominance. Gene recombinations inside each chromosome seem to be rather limited.

Inbreeding hinders gene expression and renders worth less a genetic analysis with inbred material.

Dwarf millet offers interesting hopes for breeding.

P L A N

I	INTRODUCTION : Présentation des recherches et du petit mil.	3
II	Le matériel et les conditions d'expérimentation.	7
	1. Le diallèle 9 x 9 et le croisement triple	8
	2. Le diallèle 6 x 6	11
III	Les méthodes d'analyse	17
	A. Le croisement triple modifié	
	B. Analyse de variance du tableau diallèle par le modèle d'Eberhart.	18
	C. La distance R de Hanson et Casas	22
	1. modèle simple	
	2. modèle du diallèle 9 x 9	26
	- Le modèle génétique	
	- Structure génotypique des neuf lignées	
	- Définition et calcul de l'hétérosis	
	- Calcul de la distance	
	- Expression du test d'épistasie du croisement triple	
	- Expression des estimations D dans les essais Triple et Diallèle	
	- Expression du H1 diallèle estimé dans le modèle simple	
	- Analyse de la relation entre l'hétérosis et la distance ajustée	
	- Construction d'une distance générale.	
	3. Estimation des hétérosis et des distances.	42
	4. Modèle n°2 de la Caroline du Nord	43
IV	Les Techniques de sélection	45
	A. Evaluation du rendement/ha sur les caractéristiques de pieds individuels	
	B. Maximisation du rendement d'un synthétique à la première génération	47
V	Résultats	49
	A. Analyse génétique	
	1. Qualité des essais et valeur du terrain	
	2. Variabilité génétique et hétérosis	
	3. Test du modèle additif	
	4. Influence de la densité de culture	
	5. Corrélations distances, hétérosis	
	6. Estimation des paramètres de Hayman	
	7. Les tests d'épistasie	

8.	Comparaisons H1 Triple, H1 Diallèle, $R\bar{a}^2$ max, D Triple et D Diallèle	
9.	Estimation du nombre de gènes N	
10.	Estimation du degré de dominance apparente	
11.	Construction d'une distance générale R^2G	
12.	Analyse des distances dans le diallèle 6 x 6	
B.	Evaluation du rendement sur les caractéristiques de pieds individuels	73
C.	Maximisation d'un synthétique	76
VI	CONCLUSION	79
	Références bibliographiques	
	Annexes	

I INTRODUCTION

La présente étude constitue un bilan des résultats acquis par l'auteur au cours de travaux sur l'amélioration du petit mil qui s'inscrivaient dans le programme d'ensemble de l'équipe ORSTOM de Bambey au Sénégal entre 1972 et 1975.

Ce programme reposait essentiellement sur le pari du mil nain. Une collection de lignées naines avait été fabriquée dans une première phase. Nos travaux consistaient à étudier les aptitudes à la combinaison de ces lignées et à définir des formules de variétés synthétiques.

Les acquisitions de ces travaux se rapportent à trois niveaux distincts de connaissances :

- la connaissance des lignées soumises à la sélection concernant leurs potentialités de rendement et par conséquent une appréciation sur le pari "mil nain".

- la connaissance de la génétique des caractères quantitatifs dans l'espèce des petits mils cultivés. Ceci concerne l'importance de l'hétérosis, le mode d'action des gènes et leur nombre, les relations de pleiotropie entre caractères, l'influence du milieu (ici la densité de semis) sur leur déterminisme génétique.

Cette connaissance est susceptible de guider la sélection sur l'ensemble du matériel génétique de l'espèce.

- la connaissance des méthodes d'analyse génétique et des techniques de sélection.

La méthode d'analyse des croisements diallèles proposée par HAYMAN (1954 a, 1954 b) est ici confrontée à une méthode basée sur le modèle génétique d'EBERHAT (EBERHAT et al. 1966).

La validité des paramètres génétiques estimés par croisements diallèles est appréciée expérimentalement par comparaison avec des estimations indépendantes obtenues avec un croisement triple (KEARSEY et JINKS 1968).

Une distance génétique fondée sur plusieurs caractères quantitatifs a été élaborée dans la ligne de la distance de Hanson et Casas (1968). Elle permet une appréciation du nombre de facteurs génétiques différents entre unités génotypiques, une estimation des paramètres génétiques et une appréciation des relations de pleiotropie entre

caractères.

Deux aspects techniques de la sélection ont été abordés :

. Connaissant les caractéristiques de végétation ainsi que le poids de grain d'un pied individuel de mil, peut-on estimer son rendement/ha ?

Peut-on en outre apprécier la variabilité génétique du Poids de grain par pied, déduction faite de la variabilité des caractères de végétation ?

. La maximisation du rendement d'un synthétique en première génération a été recherchée connaissant les rendements des lignées et des hybrides.

Tout ce travail se situe donc dans le cadre de l'ingénierie de l'amélioration des plantes.

Disons quelques mots sur la plante soumise à l'amélioration des sélectionneurs.

Présentation du petit mil

Le petit mil appelé aussi mil pénicillaire ou mil chandelle est une graminée de la tribu des panicées, du genre *Pennisetum*, subdivisée actuellement en une multitude d'espèces sauvages et cultivées (PORTERES 1951) en voie de regroupement dans une seule espèce sous l'appellation *P. typhoides* (BILQUEZ et al. 1969) ou *P. americanum* (BRUNKEN J. 1977).

Cette plante ne doit pas être confondue avec les nombreuses autres espèces de graminées portant aussi le nom de mils : *setaria italica*, *digitaria exilis*, *panicum miliaceum*, *eleusine caracana*.

Originnaire d'Afrique sans que l'on puisse actuellement préciser si elle provient d'un centre d'origine ou d'un non-centre (HARLAN J. 1971), elle est actuellement cultivée comme céréale principale sur quelques 10 millions d'hectares en Afrique, comme céréale et comme fourrage sur quelques 12 millions d'hectares en Inde et comme fourrage exclusivement aux Etats-Unis sur quelques 600.000 hectares.

Comme céréale, elle assure l'existence de centaines de millions d'êtres humains en climat sahélien là où le maïs et le sorgho ne poussent plus bien faute de pluies suffisantes. Mais les rendements en grain sont dans ces régions très faibles : 500 kg/ha. Il est urgent de les augmenter. Déjà, les sélectionneurs indiens utilisant des

.../...

hybrides commerciaux ont pu avancer des rendements de 2 tonnes/ha (FERRARIS, 1973, BURTON et al. 1968).

Son principal défaut est d'être naturellement plus conforme à un fourrage qu'à une céréale : c'est une herbe annuelle, robuste à longues et larges feuilles, haute de 3 à 4 mètres surmontée d'une inflorescence de taille variable (jusqu'à 2 m) cylindrique couverte de petits grains serrés.

Le rapport paille/grain est voisin de 5.

Cette espèce dont la diversité phénotypique naturelle est très grande et géographiquement structurée est exclusivement diploïde avec un petit nombre de chromosomes ($2n = 14$). Elle appartient aux plantes avec une activité photosynthétique dite en C₄, caractérisée en particulier par l'absence de photorespiration (POUZET, 1971).

Sa manipulation par le sélectionneur est presque aussi aisée que celle du maïs. Comme ce dernier, elle est presque exclusivement allogame. Les épillets contiennent deux fleurs : une staminée, l'autre hermaphrodite mais d'une protogynie accusée (de 1 à 4 jours suivant le milieu). Cette protogynie rend aisée l'hybridation contrôlée tout en n'interdisant pas l'autofécondation.

Malgré des travaux de sélection conduits depuis plusieurs décades, l'étape décisive de transformation de cette quasi-herbe en céréale parfaitement domestiquée reste à franchir.

Cette étape se fait attendre sans doute parce que les recherches scientifiques fondamentales sur cette plante ont été jusqu'ici négligées. Il est révélateur qu'aucun manuel n'existe à l'heure actuelle qui rassemble les connaissances acquises sur cette plante classée cinquième céréale majeure dans le monde.



II LE MATERIEL ET LES CONDITIONS D'EXPERIMENTATION

Les différentes lignées de petit mil ici étudiées provenaient toutes de la collection naine de Bambey.

Cette collection contient des lignées sélectionnées en culture pédigrée à partir d'hybrides comportant comme parent femelle l'une des deux souches naines de Tifton : 23 D₂ B₁ et 239 D₂ B₂ ou une souche naine indienne I 472 et comme parent mâle une souche traditionnelle de l'Afrique de l'Ouest.

La sélection pratiquée a porté sur les objectifs suivants :

- création de structures morphologiques nouvelles dans le but d'améliorer le rapport grain/paille.

L'idéotype recherché comportait des tiges beaucoup plus courtes et beaucoup plus minces que celles des mils traditionnels, un nombre d'étages foliaires réduit, des feuilles étroites et dressées, chandelles longues.

- adaptation au milieu de ces structures en leur conférant un cycle semis-récolte de 75 jours et une bonne résistance au *sclerospora graminicola*.

Les trois géniteurs femelles de la collection contiennent le gène de nanisme D₂ agissant principalement sur le raccourcissement des entrenœuds mais non sur leur nombre.

La souche naine indienne I 472 est elle-même issue du croisement de 23 D₂ B₁ avec une lignée traditionnelle indienne. Elle est insensible au photopériodisme.

Au cours de l'hivernage 1973, neuf lignées issues de I 472 x 1133 ont été étudiées dans un croisement triple et dans un diallèle 9 x 9 à deux densités.

Au cours de l'hivernage 1975, six lignées provenant de 5 hybrides différents ont été étudiées dans un diallèle 6 x 6 à faible densité (1 m x 1 m).

Les caractères étudiés

Tous ces essais ont été étudiés pour 12 caractères sur chaque pied, classés en caractères de végétation et caractères de production de grain.

Caractères de végétation :

Cy cycle compté en nombre de jours du semis à l'épiaison

Lf longueur en cm de la 3ème feuille numérotée à partir de la chandelle en descendant sur la tige principale

lf largeur en mm de la 3ème feuille

LT longueur de la plus haute tige mesurée en cm du sol à la base de la chandelle

PT périmètre en mm du 3ème entrenœud à partir du pédoncule de la chandelle

PROJ longueur en cm de la plus grande projection d'un talle sur le sol, mesurée à l'aide d'un fil à plomb tenu au bout du talle le plus "étalé".

Caractères de production de grain :

NC nombre de chandelles du pied

LC longueur moyenne en cm des chandelles du pied

PC périmètre moyen en mm des chandelles du pied

NE nombre de touffes d'épillets par cm² de la chandelle moyenne

P 500 Poids de 500 graines prises sur la chandelle moyenne

Poids Poids de grain récolté sur le pied.

1. Le Diallèle 9 x 9 et le croisement triple

Neuf pieds F5 issus de l'hybride I 472 x 1133 ont été croisés en diallèle sans les croisements réciproques au cours de la saison sèche 1973. Les croisements et les auto-fécondations ont été semés le 10/7/73 à Bambey en blocs randomisés complets à deux répétitions de 5 pieds et à deux densités : 1 m x 1 m soit sans compétition et 45 x 15. Dans ce dernier cas, les 5 pieds de chaque répétition étaient bordés de toutes parts par d'autres pieds à même densité.

Parallèlement, neuf pieds F5, frères des précédents, issus des mêmes neuf pieds F4 ont été croisés dans un croisement triple (KEARSEY et JINKS, 1968) : chaque pied a été croisé avec I 472, 1133 et auto-fécondé. Ces croisements et auto-fécondations ont été cultivés en juillet 1973 en essai complètement randomisé à 2 répétitions de 5 pieds à la densité 1 m x 1 m, sur une parcelle différente de celle qui porta l'essai diallèle, dans des conditions moins favorables.

Place des neuf lignées dans la descendance de l'hybride I 472 x 1133

Le parent 1133 est une lignée auto-fécondée pendant plus de 10 ans issu d'une variété de l'ouest nigérien : Haini Kirei.

Le tableau 1 montre la généalogie des neuf lignées étudiées, désignées par L1, L2 ... L9, retenues parmi plus d'une centaine de pieds F4 issus de quelques 68 pieds F3 engendrés par seulement 9 pieds F2. Chacune des neuf lignées Li remonte à un pied F2 différent.

Au stade F4, environ une centaine de lignées remontant à une centaine de pieds F3 remontant eux-mêmes à 13 pieds F2 ont été décrites pour la hauteur de plante, la longueur de chandelle et le nombre de chandelles.

L'analyse en composantes principales des valeurs moyennes permet de représenter l'essentiel de la variabilité (84%) sur les deux axes principaux. Sur le graphique n°1 correspondant à ces deux axes, les positions des neuf lignées contenant les pieds F5 utilisés montrent que ces derniers se répartissent à travers l'ensemble de la diversité contenue dans les 13 pieds F2 d'origine.

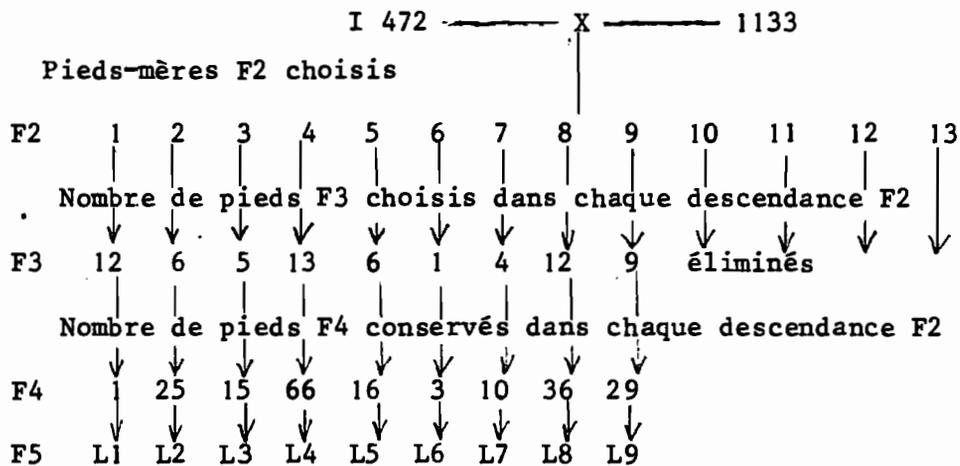


Tableau 1. Généalogie des neuf lignées.

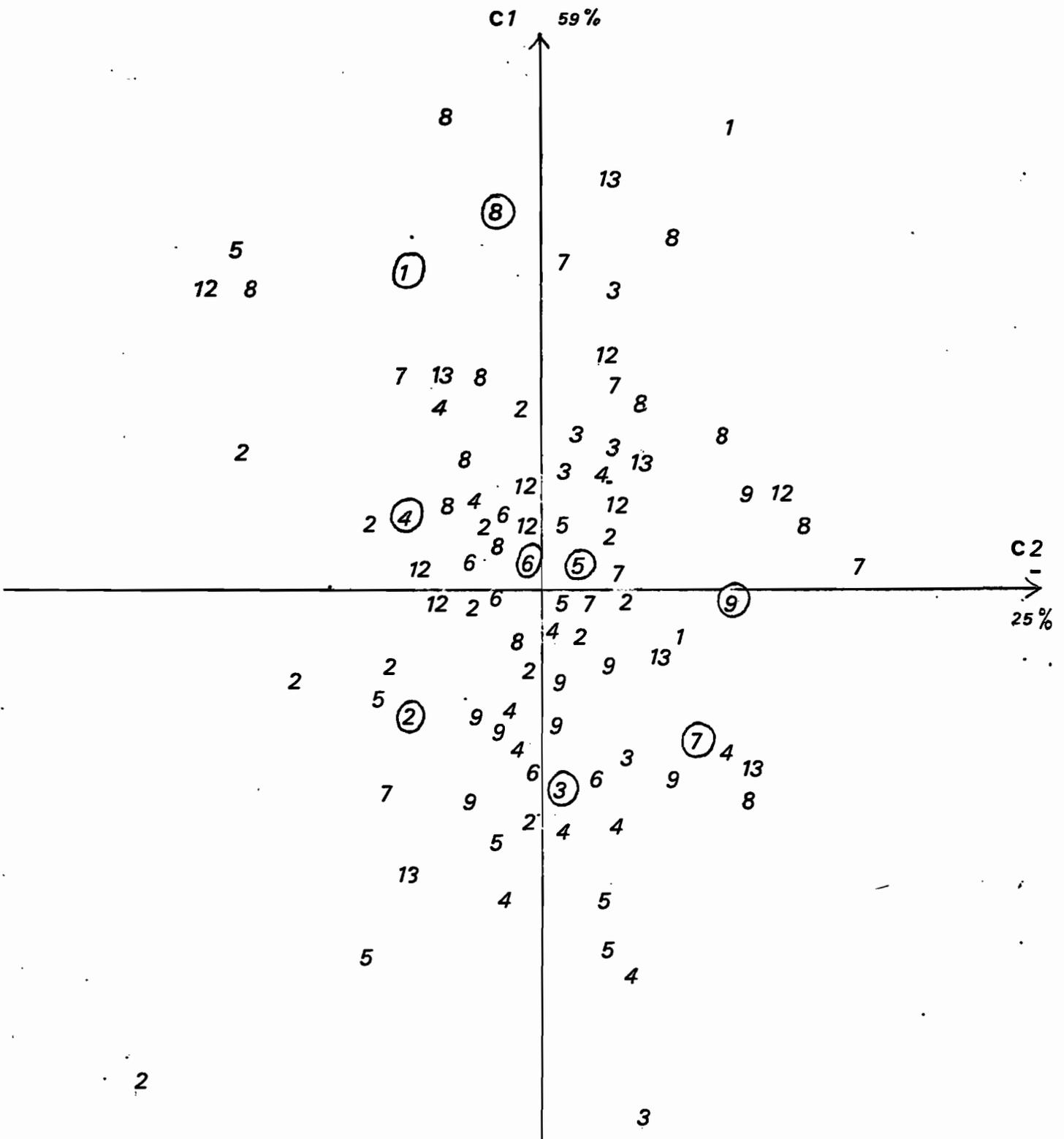


FIG. 1

Analyse en composantes principales. Nuage des lignées sur les deux axes principaux.

Chaque point représente la valeur moyenne de la descendance d'un pied F3 qui porte le numéro du pied F2 dont il est issu. Les pieds encadrés sont les ancêtres F3 des lignées F4 utilisées dans les croisements dialèle et triple.

2. Le Diallèle 6 x 6

Les 6 lignées choisies sont des lignées issues par auto-fécondation jusqu'en F5 et au-delà, de divers croisements parmi lesquels le croisement générateur du diallèle 9 x 9 représenté par 2 lignées parmi les plus différentes.

Lignée	Origine	stade
1	L1 du diallèle 9 x 9	F6
2	L6 du diallèle 9 x 9	F6
3	239 D2 B2 x 1133	F9
4	239 D2 B2 x T E R A	F8
5	I 472 x S O U N A 2	F5
6	I 472 x T E R A	F5

Le diallèle 6 x 6 contient par construction une plus grande diversité géographique que le diallèle 9 x 9.

Les 2 lignées extrêmes du croisement I 472 x 1133 doivent permettre de comparer les diallèles 9 x 9 et 6 x 6. Souna 2 est une population améliorée des mils traditionnels sénégalais. Tera est une souche de mil traditionnel de l'Ouest nigérien voisine de 1133 mais non fixée. 239 D2 B2 est la lignée naine Tifton bien connue, maintenant de stérilité mâle sur le cytoplasme A2. La lignée n°3 issue de 239 x 1133 a été choisie pour une remarquable vigueur en auto-fécondation, un abondant tallage synchronisme et une bonne fertilité.

Les croisements de ces 6 lignées fabriqués en octobre 1974 ont été semés le 18/7/75 en blocs randomisés complets à 4 répétitions de 15 pieds à intervalles de 1m.

A cause des manquants, l'analyse a retenu 14 pieds par répétition pour les F1 et 10 pieds par répétition pour les parents auto-fécondés.

Place des 6 lignées dans la collection naine

En hivernage 1974, un échantillon de 41 lignées naines appartenant à la collection de Bambey comprenant les 6 lignées du diallèle plus un échantillon de mil traditionnel sénégalais Souna 3 ont été étudiées pour les 10 caractères suivants :

.../...

- LF longueur de la 3ème feuille à partir de la feuille paniculaire
- lf largeur de la 3ème feuille
- PF projection horizontale du segment de la 3ème feuille compris entre la base du limbe et le point le plus haut du limbe
- HF projection verticale du segment ci-dessus
- NF nombre de feuilles de la talle principale
- EF distance entre les bases des 2ème et 3ème limbes foliaires
- LT hauteur de la talle principale du sol à la base de l'épi
- PT périmètre de la talle principale au 3ème entrenœud
- NC nombre de chandelles
- LC longueur de la chandelle principale

La liste des 42 échantillons de mil est donnée dans le tableau 2 : on trouve des lignées en F3, F4 ou F5 issues de croisements entre les mils nains Tifton 23 DB, 239 DB ou le mil indien I 472 d'une part et des souches africaines d'une part.

Chaque lignée a été décrite sur 4 pieds. L'ensemble des moyennes de chaque lignée a été soumis à l'analyse factorielle des correspondances de BENZECRI, sur les variables d'origine, homogènes grâce à un choix d'échelle convenable.

Les deux premiers axes traduisent 61% de la variabilité totale et semblent résumer toute la partie interprétable de la variabilité.

Le graphique n°2 plaçant les lignées et les variables sur les deux premiers axes montre les phénomènes suivants :

- Le premier axe tend à placer à gauche les lignées d'origine Haute-Volta et à droite les lignées d'origine Niger.

Les lignées d'origine sénégalaise ont une position intermédiaire.

En effet à gauche du 1er axe, on trouve 10 Haute-Volta et 5 Niger tandis qu'à droite, on trouve 1 Haute-Volta et 18 Niger.

L'influence Haute-Volta se caractérise par des tiges fines et longues, des feuilles penchées, étroites, pas très longues, des chandelles courtes.

L'influence Niger se caractérise par des tiges grosses et courtes, des feuilles nombreuses longues et larges dressées, rapprochées les unes des autres, des chandelles longues.

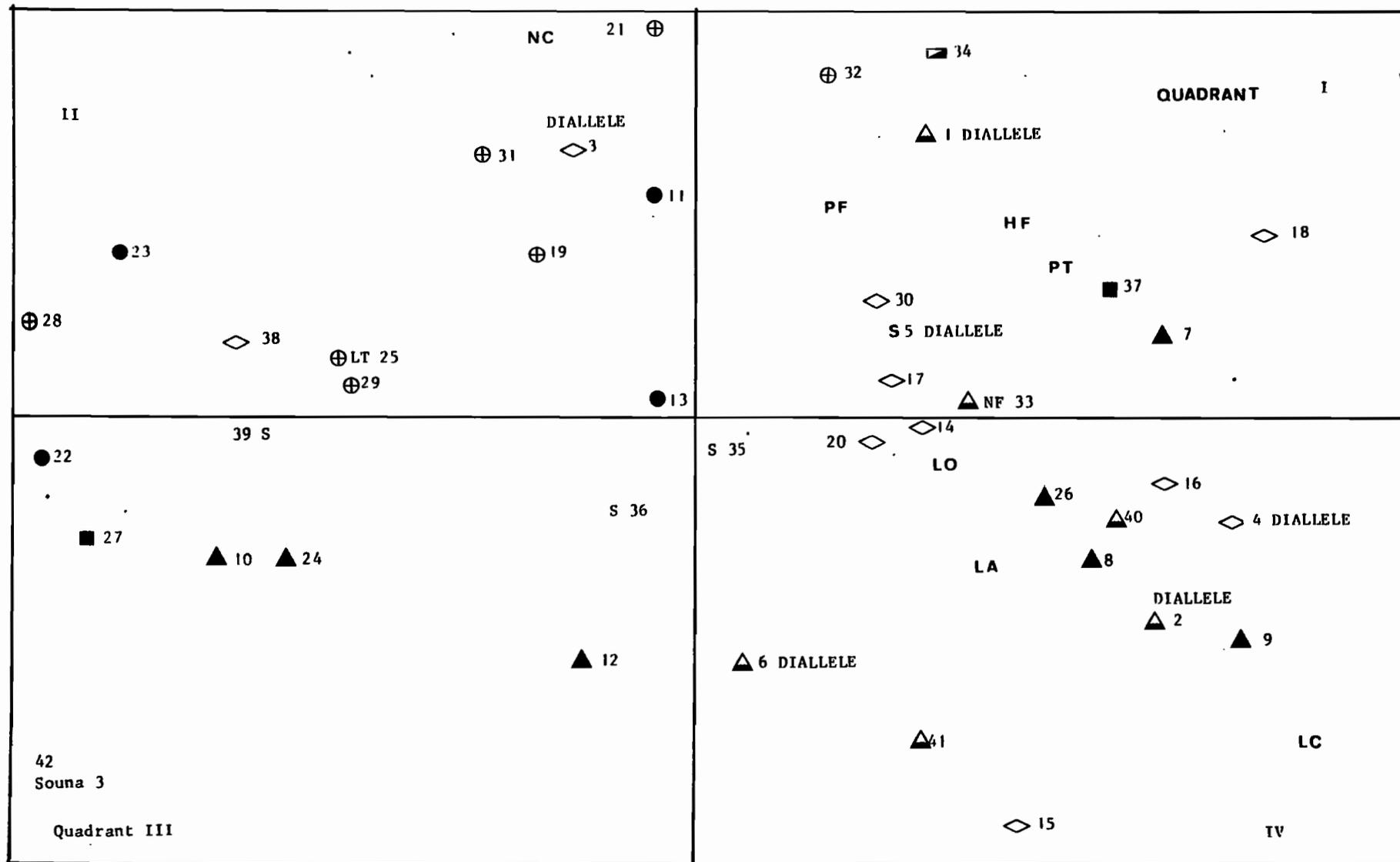
.../...

- Le deuxième axe tend à distinguer en haut les lignées issues de 239 et en bas des lignées issues de 23 ou I 472.

En haut de l'axe 2 on trouve 12 lignées 239, 5 lignées 23 et 3 lignées I 472 tandis qu'en bas on trouve 5 lignées 239, 8 lignées 23 et 6 lignées I 472.

N°	Origine	Stade	N°	Origine	Stade
1	L1 I 472 x 1133		22	23 x Aniata	F4
2	L6 I 472 x 1133		23	23 x Aniata	F4
3	239 x 1133		24	23 x Maïioa	F4
4	239 x Tera		25	239 x Kazouya	F3
5	I 472 x Souna 2		26	23 x Maïioa	F4
6	I 472 x Tera		27	23 x Bandiagara	F4
7	23 x Ba Angouré	F3	28	239 x Aniata	F4
8	23 x Longo	F3	29	239 x Aniata	F4
9	23 x Longo	F3	30	239 x Maïioa	F4
10	23 x Tamangagi	F3	31	239 x Kajouré	F4
11	23 x Kayaoga	F3	32	239 x Kazouya	F4
12	23 x Ibohamane	F3	33	I 472 x HK 5229	F4
13	23 x Titao	F3	34	I 472 x MOPTI	F4
14	239 x Ba Angouré	F3	35	I 472 x Souna 2	F5
15	239 x Ba Angouré	F3	36	I 472 x Souna 2	F5
16	239 x Ba Angouré	F3	37	23 x Goundam	F5
17	239 x Zongo	F3	38	239 x Tera	F8
18	239 x Zongo	F3	39	I 472 x Souna 2	F5
19	239 x Aniata	F3	40	I 472 x Tera	F5
20	239 x Maïwa	F3	41	I 472 x Kolo	F5
21	239 x Kajeuré	F3	42	Souna 3	Population

Tableau 2 : Echantillons de la collection naine de Bambey



- ⊕ 239 x HAUTE-VOLTA
- 23 x HAUTE-VOLTA
- ◇ 239 x NIGER
- ▲ 23 x NIGER
- △ I 472 x NIGER
- 23 x MALI
- ▣ I 472 x MALI
- S I 472 x SOUNA 2

FIG. n°2 : Analyse des correspondances pour les échantillons de la collection naine.

L'influence 23 ou I 472 se caractérise par des feuilles larges bien espacées sur la tige et des chandelles longues peu nombreuses.

L'influence 239 se caractérise par des feuilles étroites empilées les unes sur les autres et des chandelles nombreuses courtes.

Soulignons bien qu'il ne s'agit que de tendances masquées par une fluctuation importante à l'intérieur de chaque hybride : voir par exemple sur le graphique les lignées 4 et 38 éloignées et pourtant issues du même croisement 239 x Tera.

- Les 6 lignées du diallèle occupent environ la moitié du nuage de l'ensemble. Il ne semble donc pas que le diallèle 6 x 6 contienne toute la diversité de la collection naine.

La distance graphique entre les n°1 et 2 soit L1 et L6 du diallèle 9 x 9 est grande et parmi les plus fortes du diallèle 6 x 6. Il ne semble donc pas que le diallèle 6 x 6 apporte une augmentation importante de la richesse génétique en ce qui concerne les caractères d'architecture par rapport au diallèle 9 x 9.

Toutefois, en ce qui concerne la richesse génétique utile au sélectionneur pour le rendement/ha en grain, on peut déjà noter les lignées prometteuses.

Concernant le rendement/ha, le but est de croiser des lignées très voisines pour les caractéristiques désirables de végétation et très différentes pour les caractères de fructification de façon à exalter l'hétérosis grain en étouffant l'hétérosis paille. C'est pourquoi si le diallèle 6 x 6 possède une variabilité restreinte pour la végétation et du même ordre que celle du diallèle 9 x 9, on peut s'en réjouir à condition que sa variabilité pour la fructification soit élevée : cette dernière variabilité ne pourra être appréhendée vraiment que par l'étude du diallèle. Le présent dossier permet surtout de repérer les architectures favorables à un bon rendement/ha.

Le graphique n°2 contient un témoin des mils traditionnels à améliorer, en l'occurrence le Souna 3 (n°42) situé en bas à gauche du quadrant III. Ce mil a une production moyenne de grain et une production élevée de paille et au total un mauvais rapport grain/paille. Le mil recherché est donc à placer du côté opposé, soit en haut à droite du quadrant I, position caractérisée par des tiges courtes et

.../...

grosses, des feuilles érigées et des chandelles nombreuses et longues.

Curieusement, la collection naine de Bambey a effectivement réussi à fuir le Souna 3 pour occuper le quadrant II avec les lignées Haute-Volta et le quadrant IV avec les lignées Niger. Peu de lignées ont atteint le quadrant I et aucune n'y occupe le coin supérieur droit.

Si l'on admet un mode d'action génique simple pour les caractères de végétation à savoir additivité ou dominance incomplète ou complète sans épistasie, on peut prévoir l'effet des croisements.

Le croisement quadrant II x quadrant IV risque fort de produire des hybrides certes lourds en grain mais malheureusement exubérants en végétation.

Les meilleurs croisements sont du type quadrant I x quadrant I avec des lignées d'origine très différentes.

D'après ces considérations, il apparaît que les lignées du diallèle 6 x 6 sont dans l'ensemble bien situées pour produire des hybrides à fort rendement/ha.

Les lignées 1 et 5 devraient donner le meilleur rendement, la lignée 3 devrait donner de bon hybrides avec le défaut cependant de tiges un peu hautes. La lignée 6 devrait produire les plus mauvais hybrides, à cause de tiges un peu hautes et de chandelles peu longues et peu nombreuses.

Il devrait être possible d'avoir des rendements encore meilleurs avec des hybrides comme n°34 x n°18 (I 472 x Mali par 239 x Niger).

III LES METHODES D'ANALYSE

III. A. Le croisement triple modifié

Le terme croisement triple modifié est notre traduction de l'expression de MATHER et JINKS (1971) : the modified triple test cross exposé pour la première fois par JINKS et COLL. (1969). Ce traitement s'applique à un ensemble de I lignées homozygotes issues pour tirages indépendants de l'hybride entre deux lignées homozygotes P1 et P2.

Le schéma de croisement comporte pour chaque lignée le croisement avec P1 (valeur y_{i1}), le croisement avec P2 (valeur y_{i2}) et l'auto-fécondation (valeur y_{i3}). Ce schéma ajoute les auto-fécondations au schéma n°3 de COMSTOCK et ROBINSON (1952).

Il permet de tester l'épistasie en testant :

$$T_i = y_{i1} + y_{i2} - y_{i3} = T \text{ constante pour tout } i$$

Avec les hypothèses traditionnelles de HAYMAN (1954) : absence d'épistasie et distribution indépendantes des gènes dans les lignées du croisement, il est possible d'estimer dans une analyse de variance en effets fixes les paramètres :

$$D = \sum 4 uvd^2 \quad H_1 = \sum 4 uvh^2 \quad \text{définis dans les I lignées}$$

soit $S_i = y_{i1} + y_{i2}$ $d_i = y_{i1} - y_{i2}$

Le modèle posé entraîne :

$$\sum_i (S_i - s.)^2 / I = D \quad \sum_i (d_i - d.)^2 / I = H_1$$

Les contenus génétiques de T_i , des sommes de carrés sont présentés en détail dans la section C concernant la distance R^2 .

Pour un essai randomisé complet à R répétitions, l'analyse de variance d'un plan factoriel à deux traitements s'applique :

$$z_{i1r} = m + m_i + f_1 + (f_m)_{i1} + e_{i1r}$$

$$z_{i2r} = m + m_i + f_2 + (f_m)_{i2} + e_{i2r}$$

$$\text{avec } e_{ijr} (0, \sigma^2), m_i = (s_i - s.) / 2, (f_m)_{i1} = (d_i - d.) / 2$$

Aux paramètres m_i , on associe la somme de carrés

$$2r \sum_i (z_{i..} - z_{...})^2 \text{ à } I - 1 \text{ degrés de liberté et l'espérance du carré moyen } \sigma^2 + \frac{r}{2} D \frac{I}{I-1}$$

Aux paramètres $(f_m)_{ij}$ est associée la somme de carrés

$$\frac{r}{2} \sum_i [z_{i1.} - z_{i2.} - (z_{.1.} - z_{.2.})]^2 \text{ à } I - 1 \text{ degrés de liberté et l'espérance du carré moyen } \sigma^2 + \frac{r}{2} H_1 \frac{I}{I-1}$$

.../...

Une modification apparait avec l'estimation de la variance erreur sur $z_{i 1 r}$, $z_{i 2 r}$ et $z_{i 3 r}$.

$$z_{i 3 r} = m_{i 3} + e_{i 3 r}$$

L'erreur engendre la somme des carrés $\sum_{ijk} (z_{ijk} - z_{ij.})^2$ a $3 I (R - 1)$ degrés de liberté.

L'ensemble des fonctions estimables $T_i - T_j = \psi_{ij}$ se prête au test $\{\psi_{ij} = 0\}$ par un unique test F qui englobe l'ensemble des tests S (SCHEFFE 1959).

$$F \text{ épistasie} = \frac{3 r \sum_i [z_{i 1.} + z_{i 2.} - z_{i 3.} - (z_{.1.} + z_{.2.} - z_{.3.})]^2}{(I - i) s^2}$$

s^2 = Carré moyen erreur.

Cette analyse de variance permet de tester l'absence d'épistasie et dans l'affirmative de mesurer D, H1 et de tester (D = 0) ou (H1 = 0).

III. B. Analyse de variance du tableau diallèle par le modèle d'EBERHART.

Il n'est pas dans notre propos d'exposer et comparer toutes les analyses de variance des croisements diallèles déjà publiées dans de nombreuses publications. Il s'agit ici de justifier le choix du modèle d'EBERHART.

Ce dernier modèle a été choisi pour deux raisons :

. Le choix des paramètres de décomposition des valeurs génotypiques des lignées parentales et des hybrides a une justification génétique alors que les autres modèles choisissent des paramètres pour une raison statistique : obtenir des sommes de carrés orthogonales.

. Le tableau d'analyse de variance suivant le modèle d'EBERHART permet un test clair de l'absence de dominance et d'épistasie, ce qui n'est pas le cas avec les autres modèles.

Il permet en outre directement, l'estimation des paramètres génétiques de HAYMAN (1954 b) D, H_1 , H_2 , F ainsi que les tests D = 0, $h_2 = 0$, $H_1 = H_2$, dans le cas où les hypothèses de HAYMAN sont vérifiées :

2 allèles par locus

lignées parentales homozygotes

absence d'épistasie

.../...

distribution indépendante des gènes à travers les parents.

En effet, le modèle d'EBERHART est posé une fois pour toutes quel que soit le plan de croisement utilisé et les difficultés de tests statistiques rencontrées.

Ce modèle est le suivant :

$$y_{ii} = v + 2 v_i$$

$$y_{ij} = v + v_i + v_j + l + l_i + l_j + l_{ij} \quad i \neq j$$

avec les contraintes $\sum_i v_i = \sum_i v_i = \sum_i l_i = 0 \quad \sum_{j \neq i} l_{ij} = 0$

Ce modèle définit d'abord les unités parentales puis les hybrides par rapport aux parents.

$l + l_i + l_j + l_{ij}$ représente l'hétérosis décomposé en effet moyen, effet général et effet spécifique, d'une grande utilité pour le sélectionneur.

L'interprétation des paramètres a été donnée par EBERHART et GARDNER (1966) dans le cas de populations panmictiques parentales ce qui inclut les lignées homozygotes. Il en ressort que l'absence d'épistasie et de dominance entraîne $l = 0$, $l_i = 0$, $l_{ij} = 0$.

Les autres modèles découlent fondamentalement des 4 modèles de GRIFFING (1956) correspondant aux 4 plans de croisements possibles.

$$y_{ij} = m + g_i + g_j + s_{ij}$$

avec les contraintes $\sum_{j \neq i} s_{ij} + q s_{ii} = 0$

$q = 1, 2, 0, 0$ pour les plans I, II, III, IV

Le changement de plan entraîne le changement des contraintes donc le changement de signification génétique des paramètres g_i et s_{ij} .

KEARNEY (1975) généralisant aux 4 plans le modèle de HAYMAN (1954) applicable lui seulement au plan I a apporté une légère amélioration en remplaçant l'effet s_{ii} par des termes tenant compte des effets spécifiques à l'auto-fécondation donc d'une autre nature que les aptitudes à la combinaison. KEARNEY aboutit à 4 modèles différents assortis des mêmes contraintes : $\sum_i s_i = 0 \quad \sum_i g_i = 0 \quad \sum_{j \neq i} C_{ij} = 0$

$$y_{ij} = m + g_i + g_j - q (D + s_i + s_j) + C_{ij}$$

$y_{ii} = m + 2g_i + (n - 1) D + (n - 2) s_i$ avec toujours $q = 1, 2, 0, 0$ pour les plans I, II, III, IV.

Tous ces modèles présentent l'avantage statistique de permettre la décomposition de la somme des carrés totaux en sommes de carrés

.../...

orthogonales donc chacune associée à un seul groupe de paramètres. Par contre, la signification génétique des paramètres change et n'est jamais claire : les effets d'additivité, de dominance et d'épistasie sont confondus.

Enfin, les paramètres g_i et s_i définissant les valeurs y_{ii} des parents ne peuvent être déterminés en l'absence des hybrides.

Dans l'analyse de variance sur le modèle d'EBERHART du plan II utilisé dans ce travail, les sommes de carrés obtenues confondent les paramètres v et l d'une part, v_i et l_i d'autre part. Les paramètres l_{ij} relèvent d'une somme de carrés spécifique.

Le détail de l'analyse de variance est donné dans l'annexe 1. Elle est analogue à celle de GALLAIS (1970) concernant l'étude des compétitions de génotypes cultivés par paires.

Le tableau 3 présente les sommes de carrés ajustés associées à chaque effet ainsi que l'espérance de ces carrés moyens.

Cette annova permet un test de l'absence simultanée de dominance et d'épistasie : il faut avoir simultanément $l = 0$, $l_i = 0$, $l_{ij} = 0$ donc les 3 tests F non significatifs.

Somme des carrés	Degrés de liberté	Espérance du carré moyen	Signification sous les hypothèses de HAYMAN
v_i ajusté	$p - 1$	$\sigma^2 + 4 krw \sum_i v_i^2 / (p - 1)$	$D = 0$
l ajusté	1	$\sigma^2 + w (p - 1) krp l^2 / (p + 2w - 1)$	$h_2 = 0$
l_i ajusté	$p - 1$	$\sigma^2 + \frac{4 W kr (p - 2) \sum_i l_i^2}{(p - 1)(p + 4 W - 2)}$	$H_1 = H_2$
l_{ij}	$p(p - 3)/2$	$\sigma^2 + 2kr \sum_{i < j} l_{ij}^2 / p(p - 3)$	
Tijk.	$(k - 1)[p(p + 1)/2 - 1]$		
erreur	$p(p + 1)k(r - 1)/2$	σ^2	
$v_i + l_i$	$2(p - 1)$	$\sigma^2 + \frac{kr}{2(p - 1)} [4w \sum_i v_i^2 + (p - 2) \sum_i (v_i + l_i)^2]$	

Tableau 3. ANOVA EBERHART

EBERHART a conçu son modèle seulement pour l'analyse de variance et se refuse à accepter les hypothèses de HAYMAN. Dépassons ses intentions et montrons même si l'intérêt est purement formel que ce tableau d'anova sous les hypothèses de HAYMAN permet l'estimation de D , $H1$, $H2$, $h2$ et les tests indiqués dans la colonne de droite du tableau ($D = 0$, $h2 = 0$, $H1 = H2$).

Le tableau 3 permet d'estimer Σvi^2 , l^2 , Σli^2 et $\Sigma_{i < j} l_{ij}^2$. Relions ces paramètres aux paramètres VP , \bar{Vr} , \bar{Wr} , V_I de MATHER et JINKS (1971) qui donnent les estimations désirés.

$$\text{Il est évident que } \Sigma vi^2/p = \frac{1}{4} Vp = \frac{1}{4} D$$

Considérons le modèle de travail :

$$m_{ij} = m + g_i + g_j + s_{ij} \text{ avec } \Sigma_i g_i = \Sigma_j s_{ij} = 0$$

Ces paramètres sont liés à ceux d'EBERHART :

$$m = v + (p - 1) l/p$$

$$g_i = v_i + (p - 2) l_i/p$$

$$s_{ij} = l/p + 2 (l_i + l_j)/p + l_{ij}$$

$$s_{ii} = - (p - 1) l/p - 2 (p - 2) l_i/p$$

Ils sont liés aussi aux statistiques de MATHER et JINKS :

$$\Sigma_{ij} s_{ij}^2 = (p - 1) l^2 + 4 (p - 2) \Sigma_i l_i^2/p + 2 \Sigma_{i < j} l_{ij}^2 = p^2 V_I$$

$$\Sigma_i g_i^2 = \Sigma_i v_i^2 + (p - 2)^2 \Sigma_i l_i^2/p + 2 (p - 2) \Sigma_i v_i l_i/p = p V \bar{r}$$

$$\Sigma_i v_i g_i = \Sigma_i v_i^2 + (p - 2) \Sigma_i v_i l_i/p = p W \bar{r}/2$$

$$\text{Il découle : } \left(\frac{p-1}{p} l\right)^2 = h_2$$

$$\Sigma_i l_i^2/p = \left(\frac{p}{p-2}\right)^2 \left(\frac{H1 - H2}{4}\right)$$

$$\Sigma_i v_i l_i/p = \frac{-p}{2} \frac{F}{(p-2)} \frac{F}{4}$$

$$\Sigma_{i < j} l_{ij}^2/p^2 = H2 \frac{p+2}{8(p-2)} - \frac{H1}{2(p-2)} - \frac{h2}{2(p-1)}$$

Ces relations portées dans l'espérance des carrés moyens donne les estimations :

$$h2 = (CMI - CM \text{ erreur}) (p - 1) (p + 2w - 1)/krwp^3$$

$$D = (CMvi - CM \text{ erreur}) (p - 1)/krpw$$

$$H2 = 4 A/(p - 2) + 4 h2/(p - 1) + 4 (CM l_{ij} - CM \text{ erreur}) (p - 3)/krp$$

$$H1 = A + H2$$

$$\text{Avec } A = (p - 1) (p - 2) (p + 4w - 2) (CMI - CM \text{ erreur})/krp^3$$

Mentionnons une difficulté rencontrée dans l'Anova du diallèle

.../...

-6 x 6.

Les parents étaient représentés par 10 pieds et les hybrides par 14 pieds dans chaque répétition.

Il a fallu analyser séparément les parents et les hybrides. Puis une analyse conjointe pondérée a été conduite sur les moyennes parentales et hybrides avec les valeurs :

- $kr = 1$
- $w =$ variance erreur d'une moyenne hybride divisée par la variance erreur d'une moyenne parentale.
- $s_e^2 =$ variance erreur d'une moyenne hybride estimée dans l'analyse séparée des hybrides.

III. C. La distance R de HANSON et CASAS

Considérons un croisement diallèle $r \times r$ entre unités génotypiques qui peuvent être des populations, des familles ou des individus.

Soit y_{it} la valeur pour un caractère du croisement $i \times t$. Chaque unité i est représentée par un vecteur de r coordonnées y_{it} . On a l'identité :

$$\sum_t (y_{it} - y_{jt})^2 = r (y_{i.} - y_{j.})^2 + \sum_t [(y_{it} - y_{jt}) - (y_{i.} - y_{j.})]^2$$

Elle peut encore s'écrire :

$$\sum_t (y_{it} - y_{jt})^2 = r (g_i - g_j)^2 + \sum_t (s_{it} - s_{jt})^2$$

où g_i désigne l'aptitude générale à la combinaison $y_{i.} - y_{..}$ et s_{it} l'aptitude spécifique, $s_{it} = y_{it} - y_{i.} - y_{j.} + y_{..}$.

Cette identité peut être interprétée comme la décomposition d'une distance euclidienne totale dans un espace vectoriel de dimension r en une distance le long de la bissectrice de l'espace totale et une distance dans l'espace de dimension $(r - 1)$ orthogonal à la bissectrice.

$$\text{Ecrivons } T^2(i, j) = r G^2(i, j) + R^2(i, j)$$

HANSON et CASAS (1968) ont choisi R^2 pour mesurer la distance entre unités génotypiques. Notons que ce R^2 divisé par r n'est autre que la variance sur l'indice t de $(y_{it} - y_{jt})$.

Cette distance apparaît d'emblée intéressante pour révéler des différences génétiques entre unités phénotypiquement voisines.

Toutefois, à première vue, on ne comprend pas pourquoi la distance T^2 totale ne serait pas retenue.

.../...

HANSON et CASAS ont étudié expérimentalement les propriétés de leur distance dans diverses expériences sur le maïs (HANSON et MOLL 1973, CASAS et al., 1968).

- La distance R était fortement corrélée avec l'hétérosis
- Elle a permis de séparer des populations phénotypiquement voisines.

- La validité de la distance R semblait diminuer au-delà d'une certaine divergence génétique à cause de ce que 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.

La présente étude a cherché d'abord à mesurer la distance R entre les lignées concernées pour mieux connaître celles-ci et à voir si ces distances permettent de prévoir l'hétérosis.

La présente étude a cherché ensuite la signification génétique de la distance R dans plusieurs situations particulières et en particulier dans la situation du diallele 9 x 9. Cette réflexion théorique conduit à mieux connaître les paramètres génétiques d'un caractère (nombre de facteurs génétiques, mode d'action de ces facteurs).

Dans la mesure où la signification génétique de plusieurs distances particulières à plusieurs caractères a pu être précisée, une distance générale a été recherchée exprimant le nombre de facteurs génétiques différents entre unités génotypiques, approchant ainsi la distance minimale de NEI (1975).

Dans la suite de cette étude, les unités testées pour la distance sont désignées par la lettre M et les unités testeurs de distance par la lettre F même si l'ensemble M = ensemble F, comme dans un diallele.

Le terme de distance désigne, pour abrégé, la distance au carré.

La distance est définie comme suit :

$$R^2 (m, m') = \text{variance de } (y_{mf} - y_{m'f}) \text{ sur l'indice } f$$

L'interprétation génétique de l'hétérosis et de la distance est présentée ci-après dans quelques situations théoriques.

1. Modèle simple

Chez les testeurs et les testés, on suppose le biallelisme, l'absence d'épistasie, l'absence de covariance entre les situations

.../...

alleliques aux divers locus.

Le modèle génique de MATHER (1949) s'applique ici, pour décrire la valeur génotypique y_m d'un individu M :

$$y_m = \mu + \sum_a \theta_{am} d_a + \sum_a (1 - \theta_{am}^2) h_a$$

θ_{am} vaut 1, 0, -1 suivant que le locus a est en situation AA, Aa, aa.

Ce modèle simple permet une formulation de l'hétérosis et de la distance commune à des individus quelconques et à des populations panmictiques.

En effet, pour une population panmictique M, on a :

$$\overline{y}_m = \mu + \sum_a \overline{\theta}_{am} d_a + \sum_a (1 - \overline{\theta}_{am}^2) h_a \quad (1)$$

avec $\overline{\theta}_{am} = \text{fréquence AA} - \text{fréquence aa} = p_m - q_m$

$1 - \overline{\theta}_{am}^2 = \text{fréquence Aa} : 2p_m q_m$

p_m et q_m sont les fréquences des allèles A et a dans M.

La valeur génotypique $y_{m \times m'}$, du croisement M x M' s'écrit :

$$y_{\overline{m} \times \overline{m}'} = \mu + \sum_a \theta_{am \times m'} d_a + \sum_a (1 - \theta_{am}^2)_{m \times m'} h_a$$

$\theta_{a, m \times m'} = \text{fréquence AA} - \text{fréquence aa} = p_m p_{m'} - q_m q_{m'} = \frac{\overline{\theta}_{am} + \overline{\theta}_{am'}}{2}$

$(1 - \theta_{am}^2)_{m \times m'} = \text{fréquence Aa} = p_m q_{m'} + p_{m'} q_m = \frac{1 - \overline{\theta}_{am} \overline{\theta}_{am'}}{2}$

Alors :

$$\overline{y}_{m \times m'} = \mu + \sum_a \frac{\overline{\theta}_{am} + \overline{\theta}_{am'}}{2} d_a + \sum_a \frac{1 - \overline{\theta}_{am} \overline{\theta}_{am'}}{2} h_a \quad (2)$$

Dans le cas de populations panmictiques, la relation (2) donne bien

$y_{m \times m} = \overline{y}_m$ parce que $\frac{1 - \overline{\theta}^2}{2} = 1 - \overline{\theta}^2 = 2pq$

La relation (2) s'applique immédiatement à des individus en posant $\overline{\theta}_{am} = \theta_{am}$. Dans ce cas, si les individus sont entièrement homozygotes, on a toujours $y_{m \times m} = y_m$.

Si les individus sont entièrement hétérozygotes, on ne trouve plus $y_{m \times m} = y_m$ comme il est normal.

Définissons l'hétérosis entre M et M' par la relation :

$$H_e(m, m') = y_{\overline{m} \times \overline{m}'} - \frac{y_{m \times m} + y_{m' \times m'}}{2}$$

.../...

L'application de la relation (2) donne :

$$H_e(m, m') = \sum_a \frac{(\bar{\theta}_{am} - \bar{\theta}_{am'})^2}{4} h_a \quad (3)$$

La distance $R^2(m, m')$ s'écrit :

$$R^2(m, m') = \text{Variance de } [y_{m \times f} - y_{m' \times f}] = \sum_a \left(\frac{\bar{\theta}_{am'} - \bar{\theta}_{am}}{2} \right) \bar{\theta}_{af} h_a$$

donc $R^2(m, m') = \sum_a \frac{(\bar{\theta}_{am'} - \bar{\theta}_{am})^2}{4} V_{af} h_a^2 \quad (4)$

V_{af} est la variance de $\bar{\theta}_{af}$

Les expressions (3) et (4) montrent que si les termes de dominance h_a sont de signe constant, la corrélation sera forte entre l'hétérosis et la distance.

Si on remarque que $\sum_m m' (\bar{\theta}_{am} - \bar{\theta}_{am'})^2 / R^2 = 2 V_{am}$ avec V_{am} = variance de $\bar{\theta}_{am}$ dans l'ensemble des M , on peut écrire les moyennes des hétérosis et des distances :

$$\overline{HE} = \sum_a V_{am} h_a / 2$$

$$\overline{R^2} = \sum_a V_{am} V_{af} h_a^2 / 2$$

Avec des populations panmixtiques quelconques, l'analyse ne peut être pensée plus loin, mais avec des lignées homozygotes dans un diallèle, on a $V_{am} = V_{af} = 4 U_a V_a$.

$U_a V_a$ sont les fréquences des génotypes AA et aa dans l'ensemble des lignées.

La variance de l'hétérosis peut être calculée :

$$VHE = \text{variance de l'hétérosis} = \sum_a 2 U_a V_a (1 - 2 U_a V_a) h_a^2$$

$$\begin{aligned} \text{En effet variance } \left(\frac{\theta_{am} - \theta_{am'}}{4} \right)^2 &= \text{variance } \frac{1 - \theta_{am} \theta_{am'}}{2} \\ &= \frac{1}{4} [1 - \bar{\theta}_a^4] = \frac{1}{4} (1 - \bar{\theta}_a^2)(1 + \bar{\theta}_a^2) \\ &= \frac{1}{4} V_a (2 - V_a) \end{aligned}$$

Le calcul de \overline{HE} , $\overline{R^2}$ et VHE permet alors d'estimer les paramètres de HAYMAN (1954) : h_2, H_2, H_1

$$\begin{aligned} h_2 &= \left(\sum_a 2 U_a V_a h_a \right)^2 = \overline{HE}^2 & \} \\ H_2 &= \sum 16 U_a^2 V_a^2 h_a^2 = 2 \overline{R^2} & \} \quad (5) \\ H_1 &= \sum 4 U_a V_a h_a^2 = \overline{R^2} + 2 VHE & \} \end{aligned}$$

.../...

Remarquons que les formules (5) 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).

$$R^2_{ij} = \sum_t (sit - sjt)^2 / r \text{ avec } sit, sjt \text{ les A.S.C.}$$

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

$$VHE = \text{Variance } (y_{ij} - \frac{y_{ii} + y_{jj}}{2}) = 2 \overline{V}_r + V_I + \frac{1}{2} V_p - 2 \overline{W}_r$$

sachant que $V_I = \overline{V}_r - V_r$, on retrouve bien les expressions de MATHER et JINKS (1971, p. 269) : $H_1 = \overline{R}^2 + 2 VHE = V_p + 4 \overline{V}_r - 4 \overline{W}_r$

2. Modèle du diallèle 9 x 9

Ce modèle tient compte de la structure particulière des lignées et des épistasies entre paires de locus pour exprimer le contenu génétique de l'hétérosis, de la distance ainsi que des paramètres génétiques de HAYMAN calculés dans les essais Triple et Diallèle.

Le modèle génétique

Le modèle génétique retenu est le modèle Foo de VAN DER VEEN (1959) limité à des épistasies entre paires de locus, applicable aux descendance d'un hybride entre deux lignées homozygotes.

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

$$y_m = C + \sum_a \theta_{am} d_a + \sum_a (1 - \theta_{am})^2 h_a + \sum_{a < a'} \theta_{am} \theta_{a'm} i_{aa'} + \sum_{a < a'} \theta_{am} (1 - \theta_{a'm}^2) j_{a/a'} + \sum_{a < a'} (1 - \theta_{am}^2) \theta_{a'm} j_{a'/a} + \sum_{a < a'} (1 - \theta_{am}^2) (1 - \theta_{a'm}^2) l_{aa'}$$

C désigne l'effet de la partie invariable du génome, commun à l'ensemble des lignées.

Les effets i désignent les interactions homozygote x homozygote, les effets j les interactions homozygote x hétérozygote, les effets l les interactions hétérozygote x hétérozygote.

.../...

Calculons la valeur $y_{m \times m'}$ du croisement $M \times M'$. Le modèle précédent a montré que $\theta_{a, m \times m'} = (\theta_{am} + \theta_{am'})/2$ et $(1 - \theta_a^2)_{m \times m'} = (1 - \theta_{am} \theta_{am'})/2$.

Si le taux de recombinaison des gènes à la formation des gamètes vaut $1/2$ et si les différents gamètes de deux unités M et M' s'unissent au hasard, les situations alléliques aux divers locus sont indépendantes et on a par exemple :

$$(\theta_a \theta_{a'})_{m \times m'} = \theta_{a, m \times m'} \theta_{a', m \times m'} = \left(\frac{\theta_{am} + \theta_{a'm}}{2} \right) \left(\frac{\theta_{a'm} + \theta_{a'm'}}{2} \right)$$

Il vient alors :

$$\begin{aligned} y_{m \times m'} &= C + \sum_a (\theta_{am} + \theta_{am'}) da/2 + \sum_a (1 - \theta_{am} \theta_{am'}) ha/2 \\ &+ \sum_{a < a'} (\theta_{am} + \theta_{am'}) (1 - \theta_{a'm} \theta_{a'm'}) ja/a'/4 \\ &+ \sum_{a < a'} (1 - \theta_{am} \theta_{am'}) (\theta_{a'm} + \theta_{a'm'}) ja'/a/4 \\ &+ \sum_{a < a'} (1 - \theta_{am} \theta_{am'}) (1 - \theta_{a'm} \theta_{a'm'}) laa'/4 \quad (6) \\ &+ \sum_{a < a'} (\theta_{am} + \theta_{am'}) (\theta_{a'm} + \theta_{a'm'}) iaa'/4 \end{aligned}$$

Remarquons que la formule (6) est toujours valable pour des individus quelconques mais aussi pour des populations panmictiques.

Structure génotypique des neuf lignées et de leurs parents

Les deux parents de la F1 source des 9 lignées sont appelés P et P' .

Les neuf lignées sont désignées par l'indice m lorsqu'elles sont testées pour l'hétérosis ou la distance et par l'indice f lorsqu'elles sont testeurs.

Les neuf lignées sont obtenues par des tirages indépendants de la F1. On admet, en outre, que les différents gènes d'une lignée sont obtenus par tirages indépendants dans le réservoir génique de la F1.

Cette hypothèse implique au moment de la formation des gamètes des diverses générations d'autofécondation une liaison absolue de certains locus et une ségrégation 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 dans l'échantillon particulier des neuf lignées ainsi obtenu, chaque paramètre θ_{am} peut être vu comme une

.../...

variable aléatoire indépendante d'espérance θ et variance 1.

Pour la même raison, on admettra que dans une lignée M particulière, $\sum a \theta_{am} \neq 0$. 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.

Statistique	Symbole	Espérance
$V_a = 4 \sum U_a V_a$	V	$(r - 1)/r$
$\frac{V_a^2}{V_a}$	V2	$(r - 1)[r(r - 1) + 2]/r^3$
$C_{aa'} = \text{Covariance } (\theta_{af}, \theta_{a'f})$		0
$\frac{C_{aa'}^2}{C_{aa'}}$	C2	$(r - 1)/r^2$
$\theta_{am} \theta_{a'm} C_{aa'}$		C2
Covariance $[(\theta_{am} - \theta_{am'})^2, (\theta_{a'm} - \theta_{a'm'})^2]$		4 C2
$V_n = \text{Variance } [\sum a (\theta_{am} - \theta_{am'})^2/4]$		$N[2V - V2 + (N - 1)C2]/4$

Tableau 4 : Espérance de diverses statistiques

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

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

Plus généralement, diverses statistiques calculées sur les neuf lignées, constituées de variables θ_{am} peuvent être vues comme des aléatoires dont on peut calculer l'espérance.

La suite des calculs fera intervenir les statistiques suivantes :

V_a = Variance de θ_{am} dans les neuf lignées ($V_a = 4 U_a V_a$)

$C_{aa'}$ = Covariance ($\theta_{am}, \theta_{a'm}$) dans les neuf lignées.

$C_{aa'}^2$ = la covariance au carré

V_a^2 = la variance au carré

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

V_n = Variance de $\sum_a \left(\frac{\theta_{am} - \theta_{am'}}{4} \right)^2$ sur l'ensemble des couples (m, m')

cv = Covariance [$(\theta_{am} - \theta_{am'})^2, (\theta_{a'm} - \theta_{a'm'})^2$]

Le tableau 4 donne les espérances de ces diverses expressions dans lesquelles r désigne le nombre de lignées et N le nombre de locus pour lesquels la variance V_a n'est pas nulle. Le détail des calculs est donné en annexe 3.

Observons que l'espérance de l'expression $\theta_{ap} \theta_{a'p} C_{aa'}$ est nulle parce que les parents ne figurent pas dans le calcul de $C_{aa'}$.

Définition et calcul de l'hétérosis

L'hétérosis est toujours défini par la relation :

$$H_e(m, m') = y_m \times m' - (y_m \times m + y_{m'} \times m')/2$$

L'application de la formule (6) dans le cas de lignées homozygotes ($\theta_{am}^2 = 1$) donne :

$$H_e(m, m') = \sum_a (\theta_{am} - \theta_{am'})^2 h_a/4 + \sum_{a < a'} (\theta_{am} - \theta_{am'}) (\theta_{a'm'} - \theta_{a'm}) i_{aa'}/4 + \sum_{a < a'} (\theta_{am} + \theta_{am'}) (1 - \theta_{a'm} \theta_{a'm'}) j_{a/a'}/4 + \sum_{a < a'} (1 - \theta_{am} \theta_{am'}) (\theta_{a'm} + \theta_{a'm'}) j_{a'/a}/4$$

.../...

$$+ \sum_{a < a'} [(\theta_{am} - \theta_{am'})^2 (\theta_{a'm} - \theta_{a'm'})^2] laa'/16$$

Posons les restrictions $iaa' = i$, $ja/a' = ja'/a = j$

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

$$\sum_{a' \neq a} (\theta_{a'm} - \theta_{a'm'}) = \theta_{am} - \theta_{am'}$$

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

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

$$i/2 \sum_a (\theta_{am} - \theta_{am'})^2 / 4$$

$$He(m, m') = \sum_a (\theta_{am} - \theta_{am'})^2 (ha + \frac{i}{2}) / 4$$

$$+ \sum_{a' \neq a} (\theta_{am} - \theta_{am'})^2 (\theta_{a'm} - \theta_{a'm'})^2 laa' / 32$$

Avec $X_a = (\theta_{am} - \theta_{am'})^2 / 4$, on a les expressions équivalentes plus simples :

$$He(m, m') = \sum_a X_a (ha + i/2) + \sum_{a, a'} X_a X_{a'} laa' / 2$$

$$He(m, m') = \sum_a X_a (ha + i/2 + \frac{N-1}{2} la) - \sum_{aa'} X_a (1 - X_{a'}) \frac{laa'}{2} \quad (7)$$

avec $(N-1) la = \sum_a laa'$

La formule (7) est valable pour les parents P et P' s'ils sont à l'état dispersé. Alors $X_a = 1$ et

$$He(p, p') \text{ en dispersion} = \sum_a (ha + i/2) + \sum_{a \neq a'} laa' / 2 \quad (7 \text{ bis})$$

Si les parents sont à l'état d'association totale, on a la formule exacte, sans aucune hypothèse restrictive :

$$He(p, p') \text{ en association} = \sum_a ha + \sum_{a \neq a'} (laa' - iaa') / 2 \quad (7 \text{ ter})$$

Calcul de la distance

$$R^2(m, m') = \text{Variance sur } f \text{ de } (ymf - ym'f)$$

Le calcul laborieux de l'espérance de $R^2(m, m')$ dans le cas de deux lignées M et M' ayant les valeurs X_a fixées est présenté en annexe 2.

Le résultat est le suivant :

$$\frac{E[R^2(m, m')] - C_2 He^2(m, m')}{V - C_2} = \sum_a X_a (ha + \frac{i}{2} + \frac{N-1}{2} la)^2$$

$$+ \sum_{aa'} X_a (1 - X_{a'}) \frac{laa'^2}{4} \quad (8)$$

.../...

On admettra que $R^2(m, m')$ observé est voisin de l'espérance, ce qui doit être d'autant plus exact que le nombre de gènes différents est élevé.

Appelons distance ajustée $R_a^2(m, m')$ l'expression :

$$R_a^2(m, m') = [R^2(m, m') - C_2 H_e^2(m, m')]/(V - C_2)$$

$R_a^2(m, m')$ correspond à $H_e(m, m')$ en portant au carré les termes $(ha + i/2 + \frac{N-1}{2} la)$ et $(- laa'/2)$ dans la formule de l'hétérosis.

Dans le cas des parents P et P', la distance $R^2(p, p')$ n'est autre que l'estimation de H1 dans le croisement triple.

Si les parents sont à l'état dispersé, on a :

$$R^2(p, p') = \text{Variance de } \sum_a \theta_{af} (\theta_{ap'} - \theta_{ap}) (ha + i/2 + \frac{N-1}{2} la)/2$$

Les termes de covariance disparaissent en espérance et

$$E(R^2(p, p')) \text{ en dispersion} = V \sum_a (ha + i/2 + \frac{N-1}{2} la)^2 \quad (8 \text{ bis})$$

Si les parents sont à l'état associé, on a y exactement pour les termes en h, i et l mais approximativement pour termes en j :

$$E[R^2(p, p') \text{ en association}] = V \sum_a (ha + \frac{N-1}{2}(la - ia))^2 \quad (8 \text{ ter})$$

avec la convention $ia = \sum_a i_a$, $iaa'/(N-1)$, $la = \sum_a l_a$, $laa'/(N-1)$

Ces formules d'hétérosis et de distance appellent quelques commentaires :

- . Les caractères strictement additifs (sans dominance, sans épistasie) engendrent des mesures d'hétérosis et de distances toujours nulles, donc inutilisables. De tels caractères, vu leur simplicité et leur rareté, ne relèvent guère de la génétique quantitative.
- . La limite supérieure du nombre de gènes décelables dans les lignées dépend du nombre de locus qui présentent chez les testeurs une variabilité. Les formules données ne sont plus valables si les testeurs sont moins variables que les unités à tester, dans le cas d'épistasie.
- . L'épistasie de type j apparait négligeable dans l'hétérosis et la distance. Chez les lignées, l'épistasie i est confondue avec la dominance mais peut se montrer chez les parents en association.
- . Les formules sont exactes pour les effets h et l. Pour des unités testées proches du maximum de diversité, l'épistasie devient invisible. Apparemment, le modèle sans épistasie convient à condition de considérer une dominance apparente $(ha + i/2 + \frac{N-1}{2} la)$ ou $(ha + \frac{N-1}{2}(la - ia))$.

.../...

Expression du test d'épistasie du croisement triple

Ce test calcule la variance de $[ypf + yp'f - yp] = T'$

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

$$T = \text{Variance de } [\sum_{a < a'} \theta_{af} \theta_{a'f} (\theta_{ap} \theta_{a'p} laa' - iaa')/2] \quad (9)$$

Pour les parents en association, on a exactement :

$$T = \text{Variance de } [\sum_{a < a'} \theta_{af} \theta_{a'f} (laa' - iaa')/2] \quad (9 \text{ bis})$$

si $iaa' = laa'$, l'épistasie n'est pas décelée.

Expression des estimations D dans les essais Triple et Diallèle

D diallèle = Variance (ym) = Variance $[\sum_a \theta_{af} da + \sum_{a < a'} \theta_{af} \theta_{a'f} iaa']$

Pour les parents dispersés, on a approximativement pour j :

D triple = Variance $[ypf + yp'f]$

D Triple = Variance $[\sum_a \theta_{af} da + \sum_{a < a'} \theta_{af} \theta_{a'f} (laa' \theta_{ap} \theta_{a'p} + iaa')/2]$

(10)

Pour les parents en association, on a exactement :

D Triple = Variance $[\sum_a \theta_{af} da + \sum_{a < a'} \theta_{af} \theta_{a'f} (laa' + iaa')/2]$ (10 bis)

Ici aussi, si $iaa' = laa'$, D Triple = D diallèle, sinon les estimations sont différentes.

Il apparaît donc que si les parents sont en phase d'association totale (lignées extrêmes) et si $iaa' = laa'$, l'épistasie ne sera pas vue ni dans le test d'épistasie Triple, ni dans l'hétérosis et la distance des parents, ni dans l'estimation de D.

Expression de H1 diallèle estimé dans le modèle simple

Le calcul exact est possible en l'absence d'épistasie.

$$H_1 \text{ diallèle} = \overline{R^2} + 2 \text{ VHE}$$

$$R^2 (m, m') = \text{Variance de } \sum_a (\theta_{am} - \theta_{am'}) ha/2 \theta_{af}$$

$$R^2 (m, m') = \sum_a \frac{(\theta_{am} - \theta_{am'})^2}{4} Va ha^2 + \sum_{a \neq a'} \frac{(\theta_{am} - \theta_{am'}) (\theta_{a'm} - \theta_{a'm'})}{4} C_{aa'} ha ha'$$

.../...

$$\overline{R^2} = \sum_a h_a^2 V a^2 / 2 + \sum_{a \neq a'} C a a' ha ha' / 2$$

$$E(\overline{R^2}) = \frac{V2}{2} \sum_a h_a^2 + \frac{C2}{2} \sum_{a \neq a'} h_a h_{a'} = \left(\frac{V2 - C2}{2} \right) \sum_a h_a^2 + \frac{C2}{2} (\sum_a h_a)^2$$

$$VHE = \text{variance de } \sum_a \frac{(\theta_{am} - \theta_{am'})^2}{4} h_a$$

$$E(VHE) = \frac{2V - V2}{4} \sum_a h_a^2 + \sum_{a \neq a'} \frac{C2}{4} h_a h_{a'} \quad (\text{voir calcul de } V_n \text{ en annexe})$$

$$\text{d'où } E(H1 \text{ Diallèle}) = V \sum_a h_a^2 + C2 \sum_{a \neq a'} h_a h_{a'}$$

donc H1 Diallèle en espérance représente la distance maximale possible entre 2 lignées du diallèle : $R^2 \text{ Max}$.

En cas d'épistasie, le calcul est approximativement :

$$\text{Sachant que } E(\overline{Xa}) = V/2 \quad E(\overline{XaXa'}) = E(\overline{Xa(1 - Xa')}) = V/4$$

$$\text{il vient } E(\overline{HE}) = \frac{V}{2} \left(h_o + \frac{100}{4} \right) \quad h_o = \sum_a (h_a + i/2) / 100 = \sum_a a' 1aa'$$

$$E(\overline{R^2}_a) \neq \frac{V}{2} \sum_a \left(h_a + i/2 + \frac{100}{2} \right)^2 + \frac{V}{16} \sum_a a' 1aa' \quad (11)$$

un calcul en annexe 4 montre que

$$E(\overline{He^2}) = \frac{V}{4} \left(h_o + \frac{100}{4} \right)^2 + \frac{V}{4} \sum_a \left(h_a + \frac{i}{2} + \frac{100}{2} \right)^2 - \frac{V}{32} \sum_a a' 1aa' \quad (12)$$

$$\text{On admet que } E(\overline{HE})^2 \sim [E(\overline{HE})]^2$$

c'est-à-dire que \overline{HE} a une valeur quasi constante dans tous les diallèles possibles.

Il vient alors :

$$E(H1) = (V - C2) \sum_a \left(h_a + \frac{i}{2} + \frac{100}{2} \right)^2 + C2 \left(h_o + \frac{100}{2} - \frac{100}{4} \right)^2 - \frac{V}{32} (2 + 3 C2 - 2V) \sum_a a' 1aa'$$

Le dernier terme en $\sum_a a' 1aa'$ a un coefficient de l'ordre de 0,01 et doit être négligeable devant les deux premiers.

$$\text{Or } R^2 \text{ MAX} = (V - C2) \sum_a \left(h_a + i/2 + \frac{100}{2} \right)^2 + C2 \left(h_o + 100/2 \right)^2$$

En définitive, pour une épistasie de type dupliqué (opposée à la dominance), en espérance H1 diallèle est supérieur à $R^2 \text{ MAX}$, la distance maximale possible entre deux lignées du diallèle. Pour une épistasie de type complémentaire (de même sens que la dominance), $E(H1 \text{ diallèle}) < R^2 \text{ MAX}$.

.../...

Les relations (11) appliquées dans (12) permettent d'écrire :

$$E(\overline{H\bar{E}}^2) = \frac{1}{V} [E(\overline{HE})]^2 + \frac{1}{2} E(\overline{R\bar{a}}^2) - \frac{V}{16} \Sigma aa' l_{aa'}^2 \quad (13)$$

En l'absence d'épistasie, la connaissance de la moyenne des hétéroses et de la moyenne des distances ajustées permet d'approcher avec (13) la moyenne des $\overline{H\bar{E}}^2$. Si la valeur $\overline{H\bar{E}}$ calculée est inférieure à la valeur observée, il y a présomption d'épistasie.

Analyse de la relation entre l'hétérosis et la distance ajustée

Rappelons les expressions (7) et (8) de l'hétérosis et de la distance ajustée.

$$H_e = \Sigma a X_a (h_a + i/2) + \Sigma aa' X_a X_a' l_{aa'}/2$$

$$R_{\bar{a}}^2 = \Sigma a X_a (h_a + i/2 + l_{ao}/2)^2 + \Sigma aa' X_a (1 - X_a') l_{aa'}^2/4$$

La distance est dite ajustée parce qu'elle élimine les effets de la covariance entre situations alléliques aux divers locus.

La relation entre H_e et $R_{\bar{a}}^2$ a été étudiée numériquement en construisant un programme pour ordinateur.

Soit un ensemble de 5 gènes, les effets h et l associés.

Toutes les valeurs possibles de H_e et $R_{\bar{a}}^2$ sont calculées en considérant toutes les combinaisons C_n^5 de n gènes tels que $X_a = 1$, soit au total $2^5 = 32$ situations, y compris le point (0,0).

Sur l'ensemble de ces 32 situations, les moyennes $\overline{X_a}$ et $\overline{X_a X_a'}$ sont pour les locus en cause $\overline{X_a} = \frac{1}{2}$, $\overline{X_a X_a'} = \frac{1}{4}$

(les effets i sont ignorés puisqu'ils sont confondus avec h) d'où, les relations sur les moyennes :

$$2 \overline{HE} = h_0 + \frac{100}{4} = H_e \text{ maximum} - 100/4$$

$$2 \overline{R_{\bar{a}}^2} = \Sigma a (h_a + l_{ao}/2)^2 + \Sigma aa' l_{aa'}^2/8 = R_{\text{MAX}}^2 + \Sigma aa' l_{aa'}^2/8$$

Cinq cas de figures ont été analysés correspondant aux 5 distributions suivantes des effets de dominance :

Cas A $h = 1, 1, 1, 1, 1$

Cas B $h = 1, 2, 3, 4, 5$

Cas C $h = 1, 1, 1, 4, 5$

Cas D $h = -3, -1, 1, 2, 5$

Cas E $h = -5, -3, 1, 3, 5$

.../...

Dans chaque cas, on a considéré l'absence d'épistasie ($l = 0$) l'épistasie complémentaire ($laa' = (ha + ha')/15$) et l'épistasie dupliquée ($laa' = - (ha + ha')/15$).

Pour chaque situation (h, l), on a calculé la corrélation ($R_{aa'}^2, He$) la valeur de $\Sigma aa' l^2 aa'/8$, une estimation N' du nombre de gènes fondée sur R^2_{MAX} , la distance maximale et $He_{max} - He_{min}$, l'étendue de variation de l'hétérosis.

En l'absence d'épistasie, $He_{max} - He_{min} = \Sigma |h|$ et $R^2_{MAX} = \Sigma h^2$. $A = (\Sigma |h|)^2 / \Sigma h^2$ est une estimation de $N = 5$ si les effets $|h|$ sont tous égaux. En supposant la distribution $h (1, 2, 3, 4, 5)$, on a $\Sigma |h| = N(N+1)/2$, $\Sigma h^2 = N(N+1)(2N+1)/6$ et l'estimation :

$$N' = \frac{2A}{3} - 0,5 + \sqrt{\frac{2A}{3} + \left(\frac{2A}{3} - 0,5\right)^2}$$

Les figures 3 contiennent toutes ces informations.

En l'absence d'épistasie, les nuages de points présentent trois formes caractéristiques :

h constant : une ligne droite

h variable mais de signe constant : un S symétrique par rapport au milieu de la droite joignant les deux points extrêmes. h de signe variable : un losange plus ou moins déformé mais symétrique par rapport au milieu de la droite joignant les deux points extrêmes.

L'épistasie est sans effet sur la forme du nuage pour les h de signe variable. Par contre pour les h de signe constant, l'épistasie complémentaire crée ou accentue une concavité vers le bas : la branche du S concave vers le haut tend à disparaître.

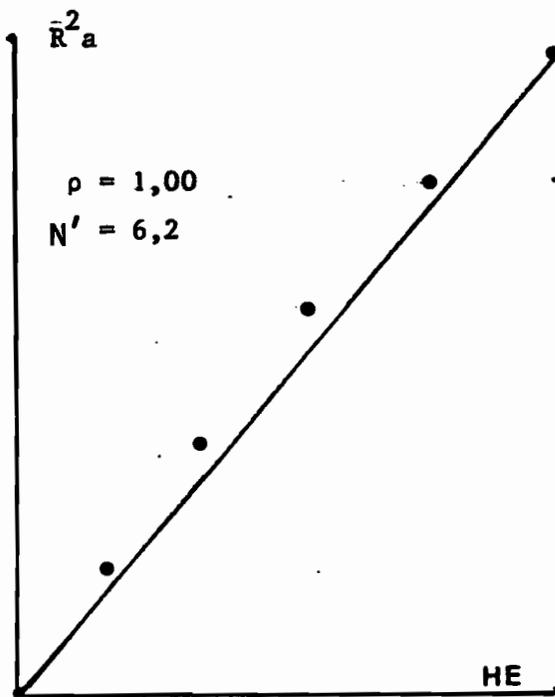
L'épistasie dupliquée agit en sens inverse : la branche du S concave vers le bas tend à disparaître. Cette dernière est plus visible que l'épistasie complémentaire : son nuage se distingue mieux d'une droite passant par l'origine.

La corrélation ($R_{aa'}^2, He$) est toujours très forte avec h de signe constant avec ou sans épistasie : elle vaut au moins 0,97 même avec la distribution aussi asymétrique : (1, 1, 1, 4, 5).

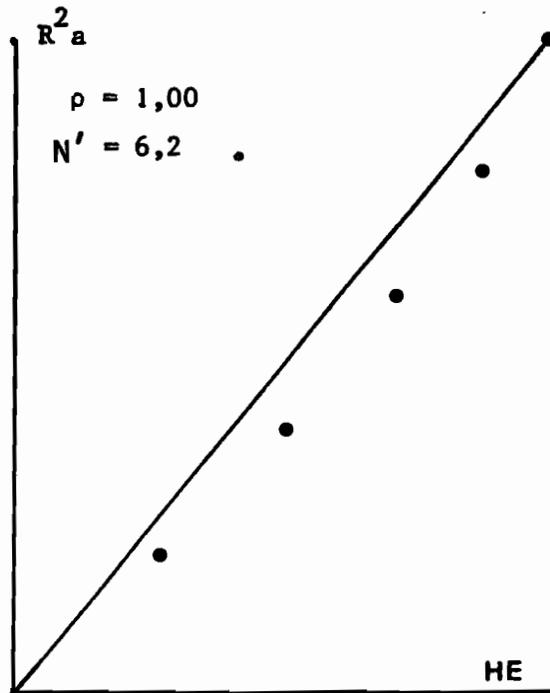
Par contre, elle baisse fortement dès que h varie en signe, et peut devenir nulle.

Le paramètre $\Sigma aa' l^2 aa'/8$ apparaît négligeable dans tous les cas. Il vaut au plus 1,1 pour un R^2_{MAX} de 67,2 (rapport = 0,016). Donc

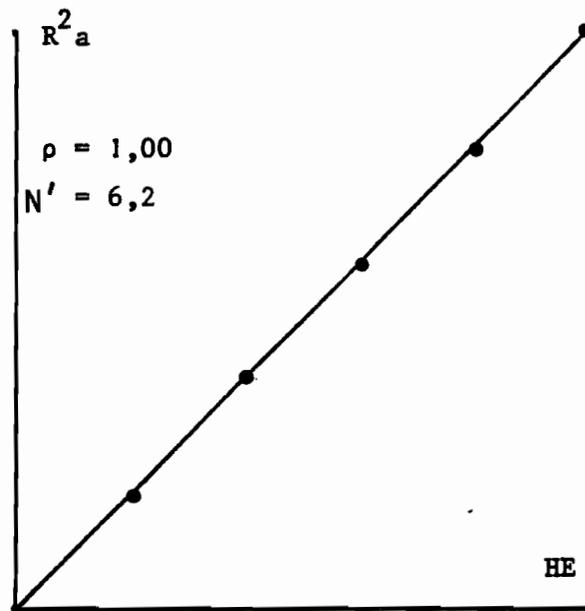
.../...



1 complémentaire



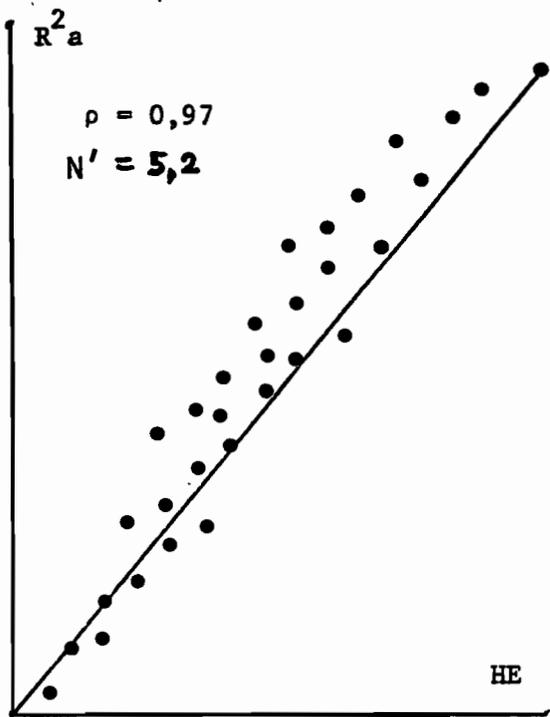
1 dupliqué



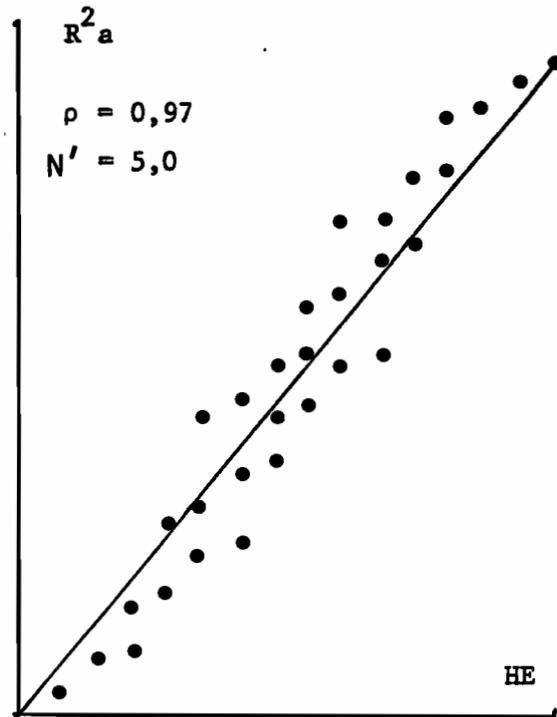
1 = 0

h = 1, 1, 1, 1, 1

FIG. 3 A Relations entre heterosis et distance ajustée sans épistasie, avec épistasie dupliquée ou complémentaire.

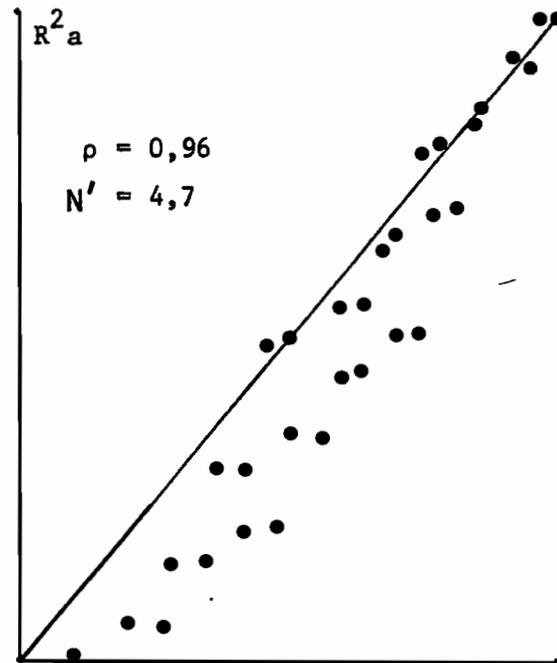


1 complémentaire



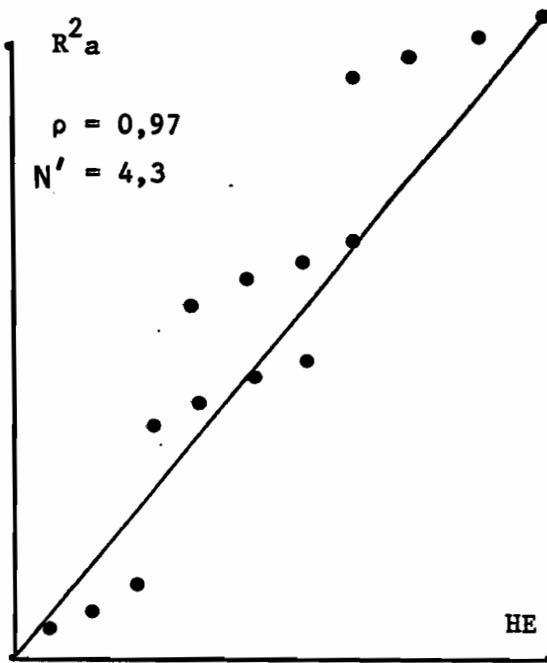
1 = 0

$h = 1, 2, 3, 4, 5$

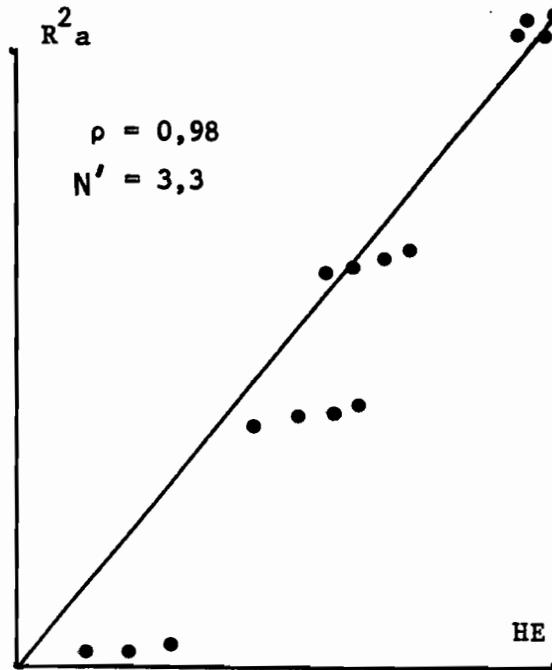


1 dupliqué

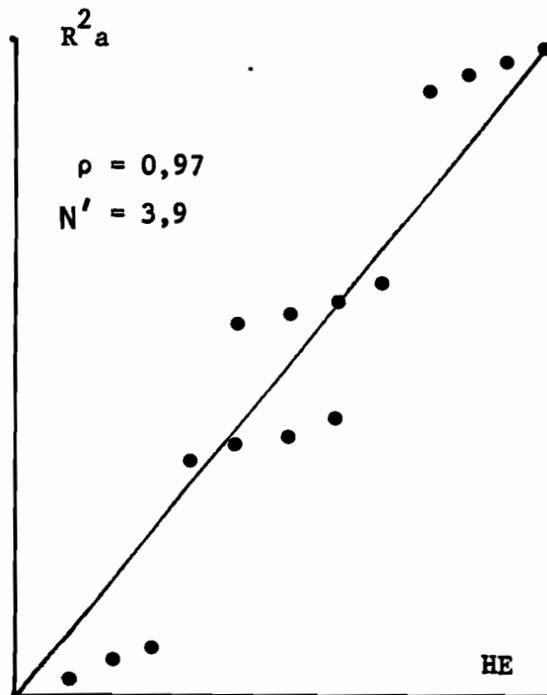
FIG. 3 B



1 complémentaire



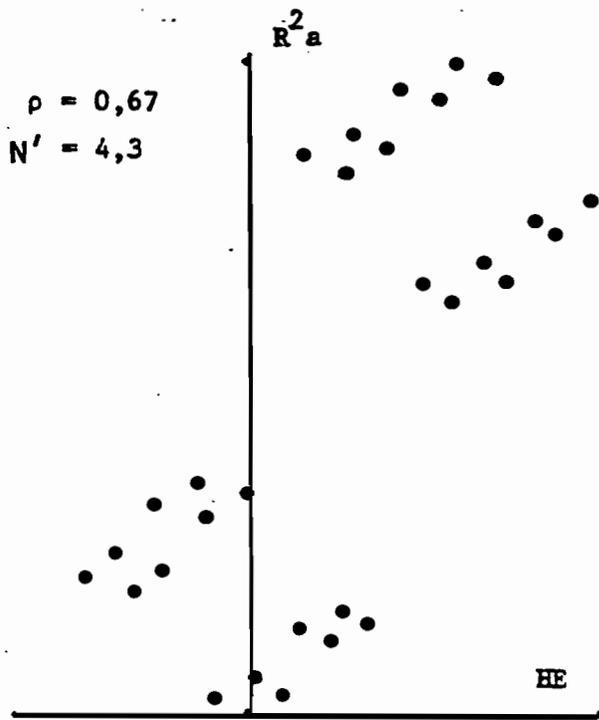
1 dupliqué



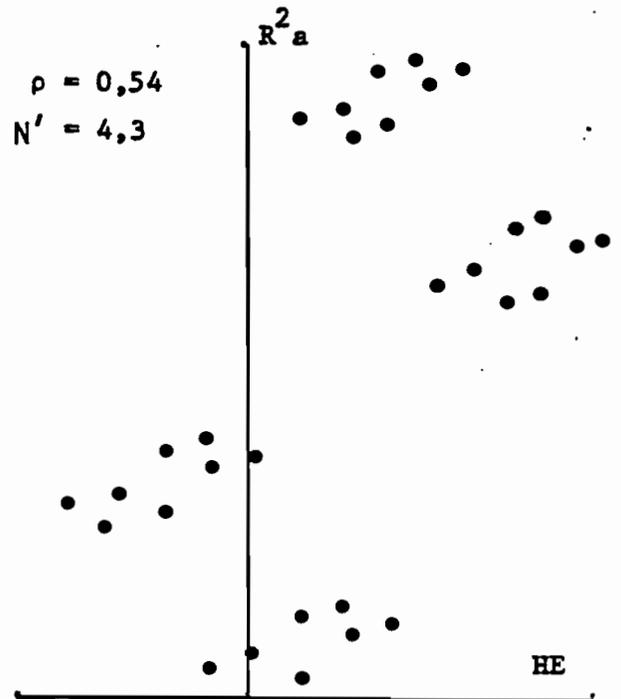
1 = 0

h = 1, 1, 1, 4, 5

FIG. 3 C



1 complémentaire



1 dupliqué

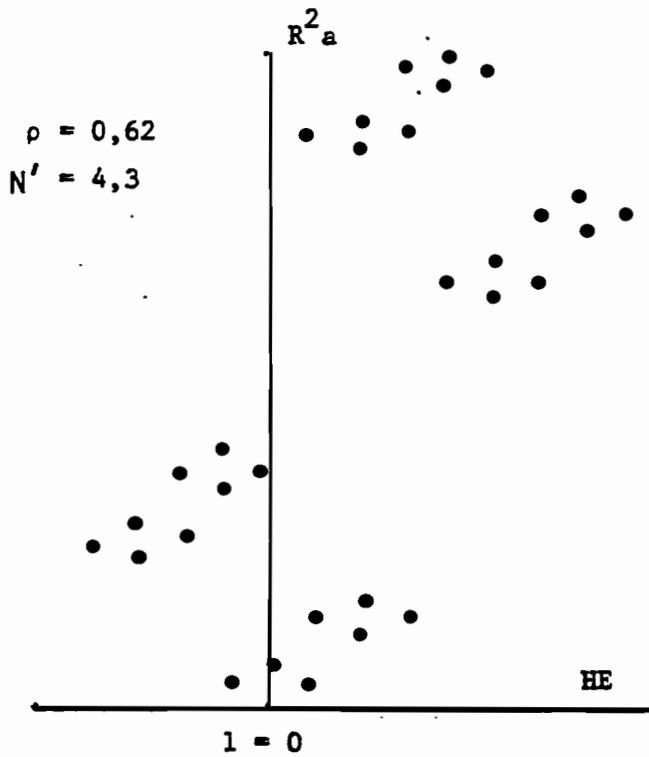
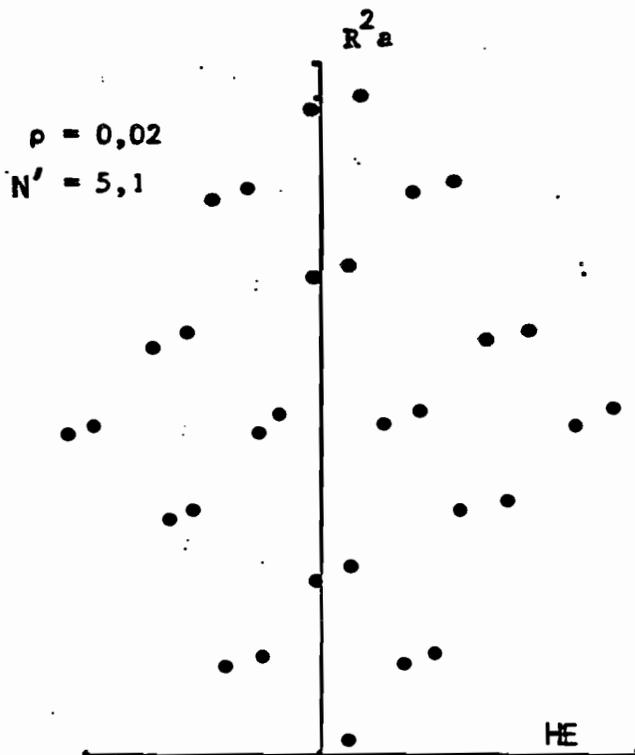
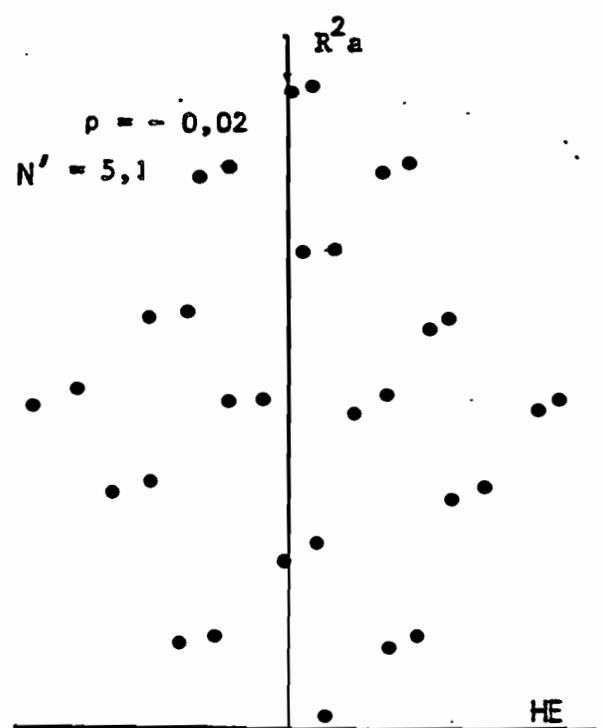
 $h = -3, -1, 1, 2, 5$

FIG. 3 D



1 complémentaire



1 dupliqué

$$h = - 5, - 3, 1, 3, 5$$

FIG. 3 E

pour R_a^2 , tout se passe comme si l'épistasie n'existait pas avec une dominance apparante ($h_a + 1a_0/2$).

$2 \overline{R^2 a}$ est une bonne estimation de R^2_{MAX} et il est vain de vouloir utiliser la relation (13) concernant l'espérance de He^2 pour déceler la présence d'épistasie.

L'estimation N' du nombre de gènes basée sur une distribution intermédiaire de $|h|$ (1, 2, 3, 4, 5) donne des résultats satisfaisants pour h de signe variable (la plus mauvaise estimation vaut 4,3).

Elle est assez bonne pour h de signe constant sauf dans le cas C avec épistasie dupliquée ($N' = 3,3$).

Si on ignore les points extrêmes R^2_{MAX} , He_{max} et He_{min} , et si la corrélation (R^2_a, He) est très forte, on peut tenter l'estimation N' avec $2 \overline{He}$ et $2 \overline{R^2 a}$ mais \overline{He} est très sensible à l'épistasie. Dans le cas B, N' vaut alors 4,1 en épistasie complémentaire, 5 évidemment sans épistasie, 6,7 en épistasie dupliquée.

En définitive, l'épistasie se décèle surtout à la forme du nuage dans les cas où la corrélation (R^2, He) est très forte. L'effet de l'épistasie est alors analogue à celui observé par Mather (1967) sur le graphique (V_r, W_r).

Construction d'une distance générale

Pour chaque caractère quantitatif, l'analyse des relations entre l'hétérosis et R^2_a la distance ajustée permet l'estimation de N , le nombre de gènes en cause et $\Sigma a (h_a + 1a_0/2)^2$ la distance ajustée maximale. Il est alors possible de calculer une distance standardisée qui estime le nombre r de gènes qui différencient deux unités.

$$n = \frac{R^2_a}{\Sigma a (h_a + 1a_0/2)^2} \times N$$

On a approximativement pour le caractère C_i :

$$n_i(m, m') = \Sigma a \left(\frac{\theta_{am} - \theta_{am'}}{4} \right)^2 S_{ai}$$

avec $S_{ai} = 0$ si le locus a n'a pas d'effet sur C_i

$S_{ai} = 1$ si le locus a a un effet sur C_i .

Il serait intéressant de construire une distance générale qui serait la somme de tous les locus différents pour l'ensemble des caractères, en évitant de compter plusieurs fois les mêmes locus.

.../...

Les covariances entre distances particulières standardisées permettent d'apprécier la proportion de locus communs à deux caractères. En effet, en espérance, (annexe 3) Covariance (ni nj) = $\frac{2V - V2 - C2}{4} \Sigma a_{sai} saj + \frac{C2}{4} (\Sigma a_{sai}) (\Sigma a_{saj})$. Or $\Sigma a_{sai} = Ni$ le nombre de gènes associés au caractère Ci.

$\Sigma a_{sai} saj = Nij$ le nombre de locus communs à Ci et Cj.
Donc covariance (ni, nj) = $\frac{2V - V2 - C2}{4} Nij + \frac{C2}{4} Ni Nj$ (14)
Cette formule (14) permet de calculer Nij.

Une distance générale R^2_g peut alors être construite en ajoutant successivement les distances standardisées les unes aux autres dans la proportion des locus non comptés par les distances précédentes suivant la formule suivante :

$$R^2_g = n1 + \left(1 - \frac{N12}{N2}\right) n2 + \left(1 - \frac{N13 + N23}{N3}\right) n3 + \dots$$

qui entraîne pour Ng le nombre total de locus associés à R^2_g :

$$Ng = N1 + \left(1 - \frac{N12}{N2}\right) N2 + \left(1 - \frac{N13 + N23}{N3}\right) N3 + \dots$$

L'assemblage est arrêté quand pour tout nouveau caractère

$$\Sigma_j Nij = Ni$$

Il est préférable de commencer l'assemblage avec des Ni le plus fort possible et des Nij le plus faible possible.

La vraisemblance de R^2_g peut être appréciée en calculant les covariances de R^2_g avec les distances ni. Il faut que Ngi soit aussi voisin que possible de Ni.

3. Estimation des hétérosis et des distances

La moyenne génotypique y_{fxm} n'est pas connue. On dispose de la moyenne phénotypique P_{fxm} telle que :

$$P_{fxm} = Y_f \times m + \epsilon_{fxm}$$

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

$$\Lambda_j R^2(m, m') = \frac{\Sigma_f (P_{mf} - P_{m'f})^2 / r - [\Sigma_f (P_{mf} - P_{m'f})]^2 / r^2}{s_e^2 / r^2} - 2(r^2 - r + 1)$$

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

.../...

$$VHE = \Sigma mm' H^2 e (m, m')/r^2 - [\Sigma mm' He (m, m')]^2/r^4$$

$$- \frac{r(r-1)}{r^4} [(r^2 - 2) s^2_e + \frac{r^2 - 2r + 2}{2} s^2_p]$$

s^2_e est l'estimation de σ^2_e sur les hybrides

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

4. Modèle n°2 de la Caroline du Nord

Le schéma du croisement n°2 de la Caroline du Nord (COCKERHAM 1963, COMSTOCK et coll. 1948) est particulièrement efficace pour mesurer les distances R^2 .

Dans ce schéma, les groupes M et F sont distincts et peuvent servir à la fois de testeurs et de testés pour la distance.

Il est permis d'espérer dans le cas le plus général que l'effet des covariances sur les mesures de distances sera minime.

Ceci peut être perçu dans le cas particulier du biallelisme. Des expressions comme $\Sigma a \neq a' \theta_{am} \theta_{am'}. Caa'$, ou $\Sigma a \neq a' Caa'$, m Caa', f ou $\Sigma a \neq a'$ Covariance ($\theta_{am} \theta_{af} \theta_{a'm} \theta_{a'f}, \theta_{a''m} \theta_{a''f} \theta_{a'''m} \theta_{a'''f}$) par le jeu d'une alternance des signes de chaque élément resteront au voisinage de zéro.

Avec un caractère sans épistasie, le modèle simple peut être envisagé. Pour des ensembles symétriques de lignées homozygotes telles que $Va, m = Va, f, Caa' \neq 0$ une analyse génétique peut être tentée à partir des statistiques $He (m, f)$, $R^2 (m, m')$ et $R^2 (f, f')$.

Les relations (5) restent applicables en considérant :

$$\overline{HE} (m, f) = \Sigma a \frac{Va}{2} ha$$

$$\overline{R^2} = (\overline{R^2} [m, m'] + \overline{R^2} (f, f'))/2 \neq \Sigma a \frac{Va^2}{2} \frac{h_a^2}{2}$$

$$VHE = \text{variance des hétérosis } (m, f) \neq \Sigma a (2 Va - Va^2) \frac{h_a^2}{4}$$

.../...

IV LES TECHNIQUES DE SELECTION

IV. A. Evaluation du rendement/ha sur les caractéristiques de pieds individuels.

Il est évident en considérant divers pieds d'un même génotype que plus la végétation est exubérante, plus la production de grain du pied augmente mais aussi plus l'occupation du sol augmente, la variation du rendement/ha étant imprévisible à priori. Il s'agit de corrélations internes du développement de la plante telles que les gènes réglant les caractères végétatifs ont des effets secondaires sur la production de grain.

Cet aspect des corrélations de développement intéressant la génétique des caractères quantitatifs a été en particulier souligné par THOMAS et coll. (1971).

L'intérêt de cette étude est d'une part de saisir l'importance des facteurs de rendement non liés aux caractères de végétation, d'autre part d'apprécier le rendement/ha d'un génotype représenté par un seul pied non gêné dans sa croissance par d'autres pieds.

La technique consiste à calculer la régression multiple du Poids de grain d'un pied individuel sur ses caractères de végétation en rapport évident avec l'occupation du sol, ce à l'intérieur d'un même génotype.

Si les régressions particulières à chaque génotype sont suffisamment proches, une régression commune aux divers génotypes est calculée puis un Poids de grain ajusté à un vecteur donné de mesures de végétation.

L'analyse génétique du Poids ajusté est conduite comme pour les autres caractères.

Plus précisément, sur chaque essai diallèle, on a calculé la régression du Poids de grain sur les caractères de végétation : LF, lf, LT, PT, NC, PRO à l'intérieur de chaque feuille.

Au préalable, les caractères de végétation ont été transformés en composantes principales relatives à la matrice moyenne des corrélations internes à chaque feuille.

Sur ces composantes principales, on a calculé la régression commune et les régressions particulières puis le Poids de grain ajusté au niveau de la végétation des pieds les plus vigoureux de l'essai.

La régression commune était facile à calculer puisque la matrice des sommes de carrés centrés des six composantes était égale à la matrice unité. Par contre, la régression particulière à chaque famille a nécessité

.../...

l'inversion de la-matrice susdite.

Appelons Z_i pour la famille i la matrice des variables de végétation placées en colonne et centrées. Soit Z_i ($r \times 6$)

r = nombre de pieds de la famille.

Soit la somme $\sum_i Z_i' Z_i$ transformée en matrice de corrélation R dont on calcule vecteurs et valeurs propres.

Les nouvelles coordonnées -mesures de végétations- s'écrivent :

C_i matrice des coordonnées de végétation des r pieds de la famille i

$C_i = \Lambda^{-1/2} V' S^{-1} Z_i$ avec V vecteurs propres en colonne

Λ matrice diagonale des valeurs propres

S matrice diagonale portant les racines carrées des numérateurs des variances des variables d'origine.

Il se vérifie que $\sum_i C_i' C_i = I$

L'application de la méthode générale d'analyse de variance conduit aux résultats suivants (LEBART et coll. 1971).

B vecteur des régressions communes sur les 6 composantes

$B = \sum_i C_i' Y_i$ où Y_i est le vecteur des Poids de grain de la famille i

B_i = vecteur des régressions particulières à la famille i

$B_i = (C_i' C_i)^{-1} C_i' Y_i$

La somme des carrés erreur du modèle général est :

$S \Omega = \sum_i Y_i' Y_i - \sum_i B_i' C_i' Y_i$

La somme des carrés erreurs du modèle restreint (régression commune) est :

$S \omega = \sum_i Y_i' Y_i - B' B$

Ces deux sommes servent au test de l'hétérogénéité des régressions.

Poids ajusté = $Y_{ij} - \sum_a C_{a, ij} b_a + A$

$A = \sum_a b_a C_a$ correspondant au pied le plus vigoureux.

IV. B. Maximisation du rendement d'un synthétique à la première génération

Soit un tableau diallèle des rendements d'un ensemble de parents et de leurs hybrides.

Cette étude se propose de calculer les fréquences de chaque parent telles que la première génération de synthétique composé avec ces parents ait un rendement maximum. Il est intéressant de voir si cette maximisation conduit à des fréquences inégales et si le maximum est différent du rendement obtenu avec des fréquences égales.

Si l'on admet que la sélection naturelle va faire évoluer le synthétique vers un équilibre stable et maximum, on peut penser que l'opération de maximisation ici décrite va rapprocher les fréquences géniques de cet équilibre. Ce serait le cas si les valeurs sélectives à chaque locus étaient constantes et sans épistasie (ARUNACHALAM et OWEN 1971).

D'une façon plus pratique et plus vague, disons que le sélectionneur a intérêt à contrôler la structure de son synthétique au cours des premières générations. Ultérieurement, après un brassage important de gènes présents, il peut espérer une évolution lente du synthétique laissé à lui-même.

Soit f_i la fréquence du parent L_i

W_{ij} le rendement du croisement $L_i \times L_j$.

Il s'agit de maximiser $\sum_{ij} f_i f_j W_{ij}$ avec $\sum_i f_i = 1$

Ce problème est traité dans ELANDT-JOHNSON (1971).

Soit W la matrice de terme général W_{ij} d'ordre s

soit W^{ij} le terme général de la matrice inverse

T la matrice de terme général $t_{kl} = W_{kl} + W_{ss} - W_{ks} - W_{sl}$ d'ordre $s - 1$.

Il existe un rendement maximum si

. Déterminant de $W \neq 0$

. $\sum_j W^{ij} > 0, \forall i$

. Le mineur principal d'ordre r de la matrice T a pour signe

$(-1)^r$, r variant de 1 à $s - 1$.

Alors $f_i = \sum_j W^{ij} / \sum_{ij} W^{ij}$

et le rendement maximum vaut :

$$W_{MAX} = 1 / \sum_{ij} W^{ij}$$

.../...

V. RESULTATS.

A. Analyse génétique

1. Qualité des essais et valeur du témoin.

Les essais dialleles ont été réussis. Le coefficient de variation erreur calculé pour les moyennes est particulièrement faible pour le diallele 6x6 (tableau 5) grâce au grand nombre de pieds de chaque moyenne.

La valeur relativement élevée du coefficient de variation erreur du diallele 9x9 à densité 45x15 ne s'explique pas par l'échec de la culture mais par un effet de compétition.

L'essai triple, par contre, présente un coefficient élevé à cause de mauvaises conditions de culture : de nombreux pieds ont manqué sur les lignées autofécondées et une sécheresse en fin de culture impossible à compenser faute d'irrigation a empêché de mesurer les caractères NC, PRO et PG. Les autres caractères ont été analysés sur les moyennes de chaque répétition.

Le niveau des moyennes des deux dialleles 9x9 et 6x6 à 1m x 1m peut être comparé grâce aux lignées L1 et L6 du diallele 9x9 et L'1, L'6 du diallele 6x6. Le tableau 5 montre que les F1 L1xL6 et L'1xL'6 ont des valeurs voisines pour tous les caractères sauf PT et PC. Les deux essais ont donc des moyennes hybrides directement comparables. Par contre, les moyennes de lignées ne sont pas comparables à priori parce que le diallele 9x9 se situe au stade de consanguinité F4 contre F6 et plus pour 6x6.

2. Variabilité génétique et hétérosis.

La valeur de l'hétérosis est appréhendée par la moyenne, v des lignées, l'hétérosis moyen l et le rapport entre l'hybride le plus fort et le parent le plus fort de l'ensemble du diallele.

CARACTERE - essai	Coefficient de Variation		Moyenne L1 x L6	v	1
	ERREUR	GENETIQUE			
CY TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	3,7				
	1,9	4	48,0	51,5	- 4,59
	2,8	5		51,8	- 5,16
	0,7	5,0	48,5	58,0	- 7,8
LF TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	3,2				
	1,6	3	68,5	61,1	5,9
	2,8	5		54,7	5,9
	0,8	4,4	66,8	61,3	8,7
1f TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	4,4				
	3,1	6	54,6	44,0	5,3
	4,1	9		35,5	4,4
	1,3	8,8	50,3	41,9	8,3
LT TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	4,1				
	2,8	7	79,9	71,0	12,6
	3,1	8		69,7	14,8
	1,3	10,3	81,2	70,2	9,0
PT TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	4,2				
	3,5	8	39,9	32,0	5,2
	5,0	10		25,6	2,5
	1,4	4,4	45,2	41,0	4,7
NC 9x9 1 ^m x1 ^m 6x6	9,8	32	6,4	3,8	1,1
	4,1	19	6,3	4,2	2,8
PRO 9x9 1 ^m x1 ^m 6x6	8,2	23	21,9	16,0	3,2
	4,7	19	21,8	19,5	9,4
LC TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	7,1				
	3,8	4	40,5	34,2	6,9
	5,4	10		30,1	5,0
	1,4	7,3	40,0	33,6	6,5
PC TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	3,5				
	2,8	6	76,1	69,3	3,2
	3,5	6		64,2	2,5
	1,0	6,0	70,6	62,6	6,3
NE TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	6,4				
	5,4	14	8,5	8,8	- 0,06
	6,3	16		10,0	- 0,46
	2,3	10,5	10,4	12,7	- 1,5
P500 TRIPLE 9x9 1 ^m x1 ^m 45x15 6x6	7,3				
	3,2	5	4,2	3,1	0,7
	5,1	9		2,8	0,4
	1,4	6,3	4,2	2,5	1,5
PG 9x9 1 ^m x1 ^m 45x15 6x6	12,6	31	156	28,4	67,4
	17,7	27		6,8	6,0
	4,4	14,8	191	22,1	175,8

TABLEAU 5

Le tableau 6 montre que dans l'ensemble, l'hétérosis moyen comme l'hétérosis extrême est élevé pour les caractères NC, P500, PG, modéré pour LT et faible pour les autres caractères. Pour ces derniers, la sélection en consanguinité apparaît efficace pour fixer la vigueur.

Le tableau 6 montre également dans l'ensemble un hétérosis plus fort chez 6x6 que chez 9x9. Pour tenir compte des différences de consanguinité, les hétérosis moyens des deux essais ont été comparés après pondération par les hétérosis respectifs du témoin L1xL6 (voir colonne $(\bar{H}_e/HeL1xL6)_{6x6}/(\bar{H}_e/HeL1xL6)_{9x9}$). Certaines valeurs n'ont pu être calculées parce que l'hétérosis du témoin dans l'un des deux essais était voisin de zéro. Il apparaît alors que l'hétérosis 6x6 n'est supérieur à l'hétérosis 9x9 que pour LF, LC, P500 et PG.

C'est là un indice d'une diversité génétique supérieure dans le diallele 6x6 mais modérée dans la mesure où elle n'affecte pas tous les caractères.

La comparaison des coefficients de variation génétique portés dans le tableau 5 ne montre pas de tendance marquée. Dans l'absolu, ils sont assez faibles sauf pour NC et PG. Ce coefficient d'ailleurs n'exprime pas toujours la diversité génétique d'un ensemble dans la mesure où toutes les lignées quoique diverses présentent toutes deux à deux une vigueur hybride voisine comme dans 6x6.

Dans le diallele 9x9 au contraire, la ségrégation des gènes jouant a permis l'expression de la diversité : le coefficient de variation génétique conduit à penser que ce diallele ne contient pas une forte variabilité sauf pour NC et PG.

L'examen des valeurs l , l_i , l_{ij} montre la structure de l'hétérosis : les effets l sont du même ordre dans les deux densités du diallele 9x9 mais les effets l_i et l_{ij} sont nettement plus forts avec la densité forte. C'est l'inverse dans la comparaison 9x9 et 6x6 : les effets l 6x6 sont dans l'ensemble les plus forts mais les effets l_i et l_{ij} sont du même ordre. Il s'agit sans doute de la manifestation de la consanguinité qui élève globalement l'hétérosis.

CARACTERE - ESSAI	1/v	F1 MAX P MAX	$\frac{1+1i+1ij}{vi+1+1i+1ij}$	$\frac{\overline{HE}/He \text{ L1XL6) } 6x6}{(\overline{HE}/He \text{ L1xL6) } 9x9}$
CY 9x9 1 ^m x1 ^m	0,09	1,10	0,60	0,6
	45x15	0,10	0,66	
	6x6	0,13	1,12	
LF 9x9 1 ^m x1 ^m	0,09	1,05	0,74	4,1
	45x15	0,11	0,87	
	6x6	0,14	1,08	
1f 9x9 1 ^m x1 ^m	0,12	1,04	0,41	1,0
	45x15	0,12	0,68	
	6x6	0,20	1,14	
LT 9x9 1 ^m x1 ^m	0,18	1,15	0,45	1,0
	45x15	0,21	0,59	
	6x6	0,13	1,26	
PT 9x9 1 ^m x1 ^m	0,16	1,04	0,38	1,1
	45x15	0,10	0,59	
	6x6	0,11	1,11	
NC 9x9 1 ^m x1 ^m	0,29	1,08	0,24	-
	6x6	0,66	1,66	
PRO 9x9 1 ^m x1 ^m	0,20	1,23	0,30	-
	6x6	0,48	1,44	
LG 9x9 1 ^m x1 ^m	0,20	1,13	0,64	1,6
	45x15	0,17	0,70	
	6x6	0,19	1,15	
PC 9x9 1 ^m x1 ^m	0,05	1,05	0,18	1,0
	45x15	0,04	0,50	
	6x6	0,10	1,08	
NE 9x9 1 ^m x1 ^m	0,00	1,10	0,22	-
	45x15	0,04	0,31	
	6x6	0,12	< 1	
P500 9x9 1 ^m x1 ^m	0,22	1,36	0,42	1,6
	45x15	0,14	0,40	
	6x6	0,60	1,59	
PG 9x9 1 ^m x1 ^m	2,37	3,90	0,86	2,2
	45x15	0,88	0,72	
	6x6	7,95	7,6	

TABLEAU 6 : Anova des essais (suite)

3. Test du modèle additif.

Dans tous les essais, pour tous les caractères, le modèle additif (ni dominance, ni épistasie) est à rejeter : les 3 tests F ne sont jamais simultanément non significatifs.

L'importance des effets non-additifs est appréciée par l'indice $(1+li+lij/vi+li+l+lij) = \frac{S_{Cl}+S_{Cli}+S_{Clj}}{SC(v+1)+SC(vi+li)+S_{Clj}-SCVA}$ soit le rapport des sommes de carrés dues aux effets non-additifs sur la somme de carrés génotypique.

Le tableau 6 montre que la part des effets non-additifs augmente avec la densité pour tous les caractères sauf P500 et PG.

Dans l'absolu, chez 9x9 la part des effets non-additifs est supérieure à 50% pour CY, LF, LC, PG. Pour tous les autres caractères, la part de l'additivité est prépondérante. Chez 6x6, les caractères à prédominance additive sont : CY, lf, LC, PC, NE. Donc la majorité des caractères est à prédominance non-additive. D'une façon relative, 8 caractères sur 12 montrent chez 6x6 une part non-additive supérieure à celle de 9x9. C'est là un indice de plus grande diversité génétique chez 6x6, surtout visible pour NC, P500 et PG.

4. Influence de la densité de culture.

Rappelons les observations précédentes sur l'effet densité :

- . La densité augmente la variance erreur des moyennes génotypiques.
- . La densité modifie la structure de l'hétérosis en exaltant les interactions lij.
- . La densité augmente la part des effets non-additifs dans la variance génétique, sauf pour P500 et PG.

Reste à voir l'effet sur les moyennes des hybrides (v+1) :

- . La densité ne modifie pas les moyennes de CY, LT, NE.
- . La densité diminue les moyennes des 9 autres caractères.

5. Corrélations distances (R^2) et hétérosis (He).

Les corrélations ont été calculées dans les dialleles 9x9 1mx1m et 6x6 entre les distances et les hétérosis d'une part, entre les distances et les valeurs F1 des hybrides d'autre part. Le tableau 7 montre que les corrélations (R^2 , He) sont très fortes pour tous les caractères sauf NE 9x9.

Elles ne dépendent pas de l'intensité de l'hétérosis. Ainsi PC 9x9 et PC 6x6 avec des hétérosis de l'ordre de 5 à 10% ont des corrélations supérieures à 0,90. NE 6x6 a une corrélation de - 0,97 bien que les hybrides extrêmes restent encadrés par les parents extrêmes.

Caractère	CORRELATIONS			
	9x9		6x6	
	R^2 , He	R^2 , F1	R^2 , He	R^2 , F1
CY	- 0,94	- 0,43	- 0,99	- 0,25
LF	0,91	0,47	0,97	0,84
lf	0,80	0,44	0,99	0,56
LT	0,91	0,87	0,89	0,77
PT	0,93	0,61	0,93	0,51
PRO	0,64	0,57	0,88	0,53
NC	0,85	0,68	0,90	0,76
LC	0,95	0,53	0,98	0,38
PC	0,90	0,50	0,96	0,70
NE	- 0,17	0,00	- 0,97	- 0,17
P500	0,97	0,89	0,99	0,77
PG	0,95	0,94	0,99	0,97

Tableau 7.

Les corrélations (R^2 , F1) présentent une chute spectaculaire par rapport aux précédentes, pour les caractères qui ne présentent pas une dépression consanguine accentuée. Ce résultat tend à prouver

que la bonne mesure de divergence entre deux unités génotypiques est l'écart entre la F1 et le parent moyen et non la valeur de la F1 (HEIDRICH-SOBRINHO, 1975) ou l'écart entre la F1 et le parent supérieur (ZALI et ALLARD, 1976).

6. Estimation des paramètres de HAYMAN.

Indépendamment du contenu génétique des estimations obtenues, il a été vérifié numériquement que les valeurs D, H1, H2, h2 calculées avec les statistiques (Vr, Wr) de HAYMAN (1954b) ou avec le tableau de l'Anova Eberhart ou avec les statistiques (R^2 , He) étaient bien exactement les mêmes.

7. Les tests d'épistasie.

La discussion porte sur le diallele 9x9 confronté au croisement triple pour lequel la théorie de la distance R^2 a des chances de s'appliquer.

Le tableau 8 compare les conclusions du test F8,27 du croisement triple, du test de régression Wr/Vr et de l'analyse des graphiques R^2 , He.

La comparaison des graphiques observés N° 4 avec les graphiques théoriques 3 conduit aux constatations suivantes :

- 7 caractères ne montrent pas d'hétérosis négatif et la corrélation (R^2 , He) est supérieure à 0,90 avec un alignement accusé des points.

- 5 caractères montrent plusieurs valeurs négatives de l'hétérosis, une corrélation (R^2 , He) au-dessous de 0,90, une grande dispersion des points. Pour NE 9x9, la forme losange apparaît nettement. Pour ces 5 caractères, la dominance ambidirectionnelle camoufle la présence éventuelle d'épistasie.

Parmi les 7 autres caractères, P500 9x9 et CY 9x9 présentent la forme en S caractéristique d'effets de dominance d'intensité variable en l'absence d'épistasie. Toutefois, pour CY 9x9, la branche

concave vers le bas semble moins développée que l'autre et on ne peut exclure une légère épistasie dupliquée. Pour les caractères, LF, LT, PT, LC et PG seule la branche concave vers le haut apparaît : il y a présomption d'épistasie dupliquée.

L'excellente courbe en S de P500 permet de penser que l'effet des covariances est bien pris en compte par le modèle théorique et que l'absence du S pour les autres caractères à très forte corrélation est bien due à une épistasie dupliquée. Ce résultat témoigne aussi de l'invalidité des conclusions du test (Wr/Vr) selon lequel CY, LF, lf, PT, PRO et PC ne présenteraient ni épistasie ni covariances.

CARACTERE	Valeur du test F8,27 de TRIPLE	Valeur de la pente b(Wr/Vr)	Analyse des graphiques R ² a, He
CY	2,48*	0,85	dupliquée légère
LF	1,18	0,88	dupliquée
lf	0,83	1,25	?
LT	2,60*	0,29*	dupliquée
PT	0,96	0,92	dupliquée
NC	-	0,59*	?
PRO	-	0,65	?
LC	1,20	0,41*	dupliquée
PC	1,89	0,92	?
NE	3,86**	0,58*	?
P500	0,74	0,52*	nulle
PG	-	0,10*	dupliquée

Tableau 8.

Le croisement triple confirme la conclusion (R²a, He) pour P500, n'infirme pas les autres conclusions sauf pour LF, PT et LC.

Dans le cas de LF, ce désaccord peut s'expliquer par des parents P et P' en phase d'association totale et ia = la, comme on

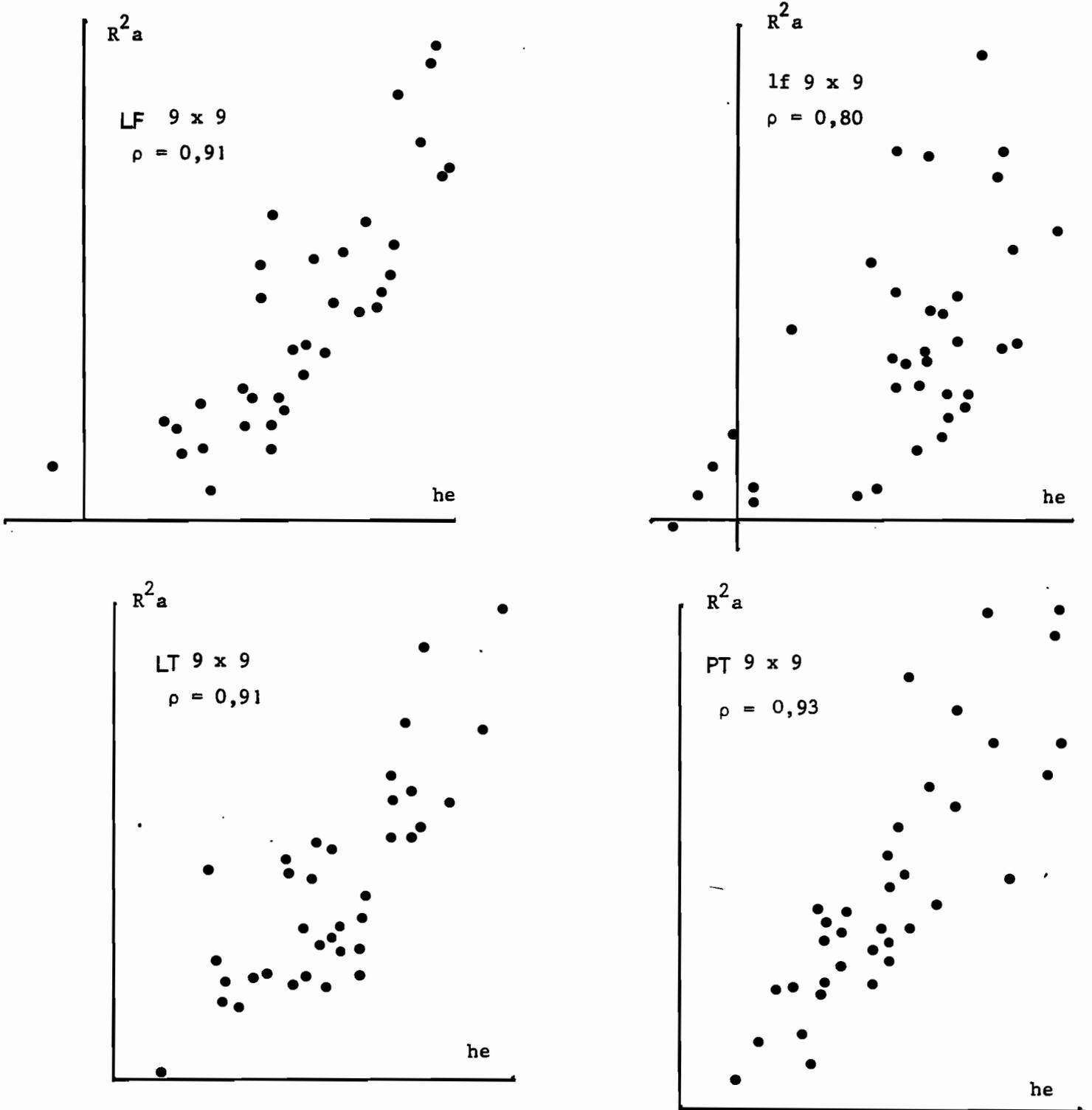


FIG. 4 : Relations entre les hétérosis et distances ajustées observées, dans le diallèle 9 x 9
 ρ = Corrélation (He , R^2 non ajusté)

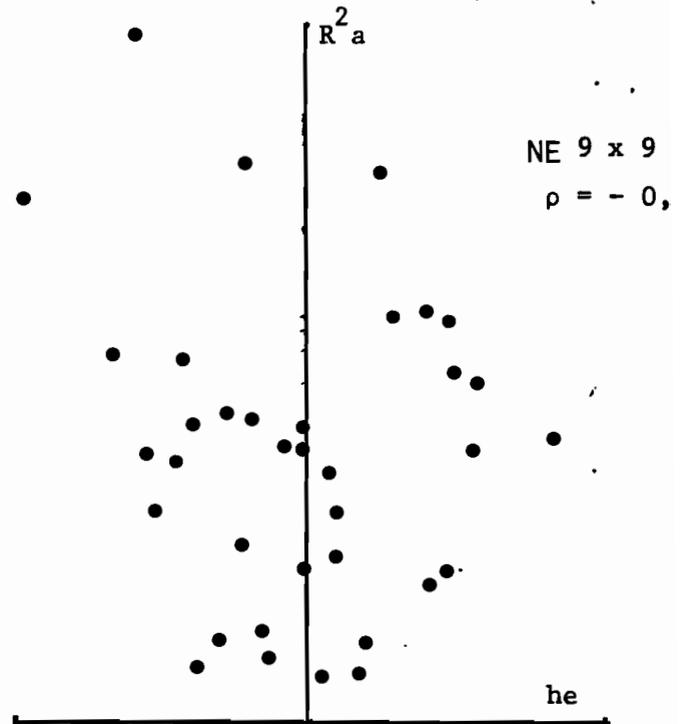
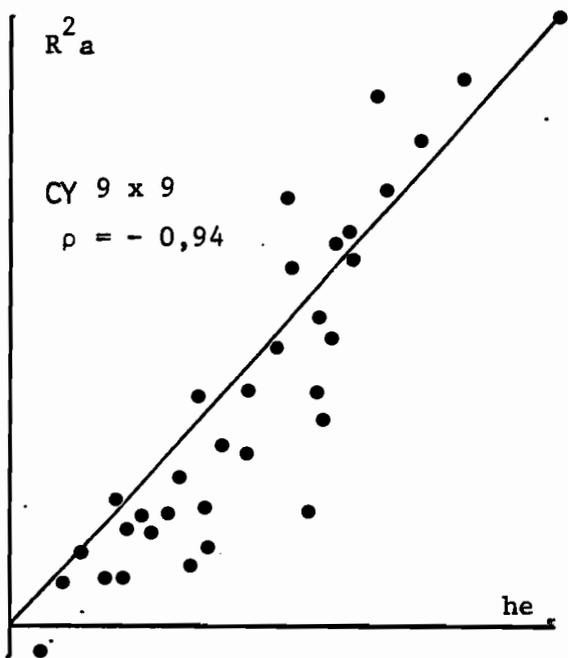
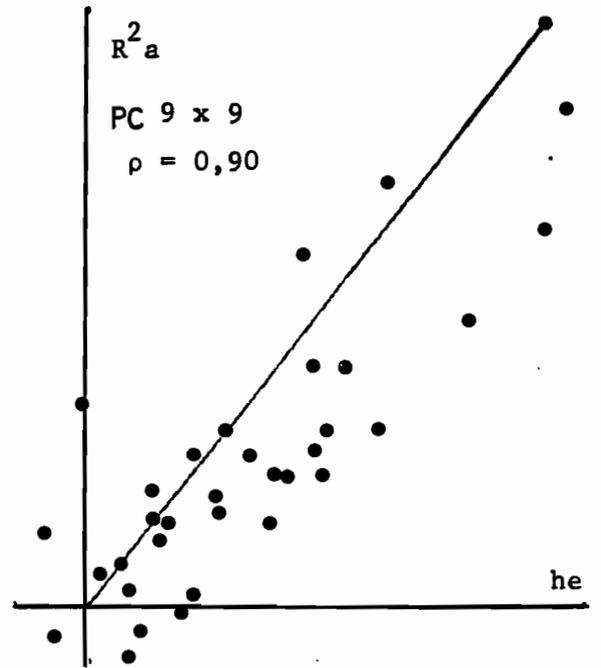
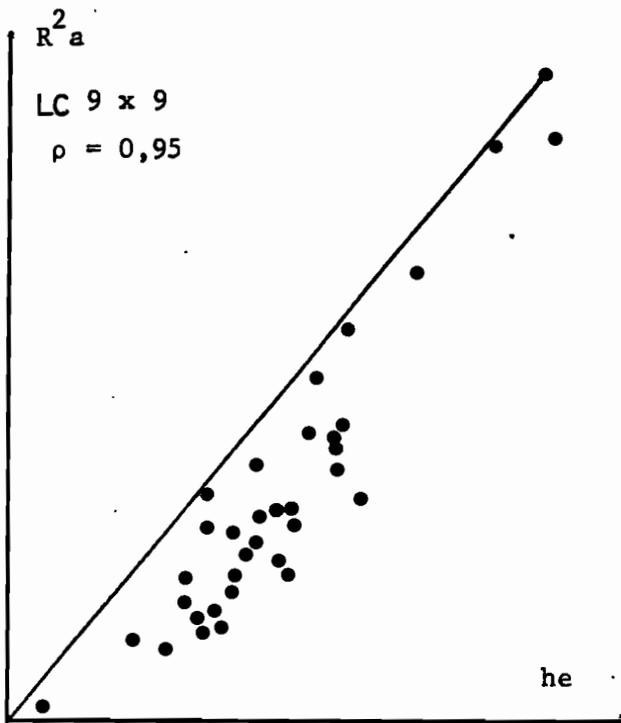


FIG. 4 : suite

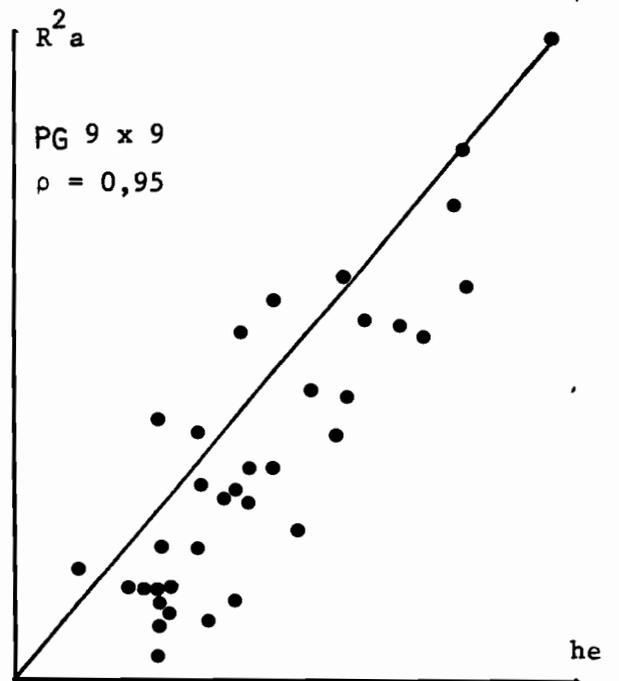
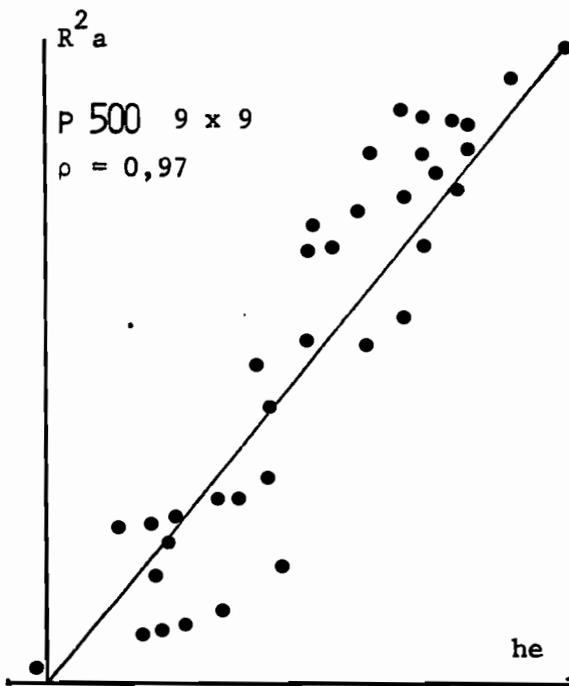
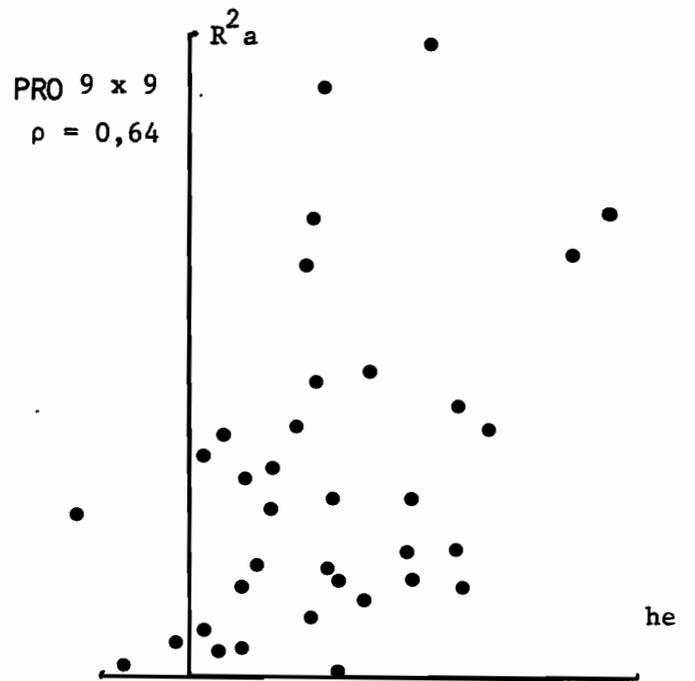
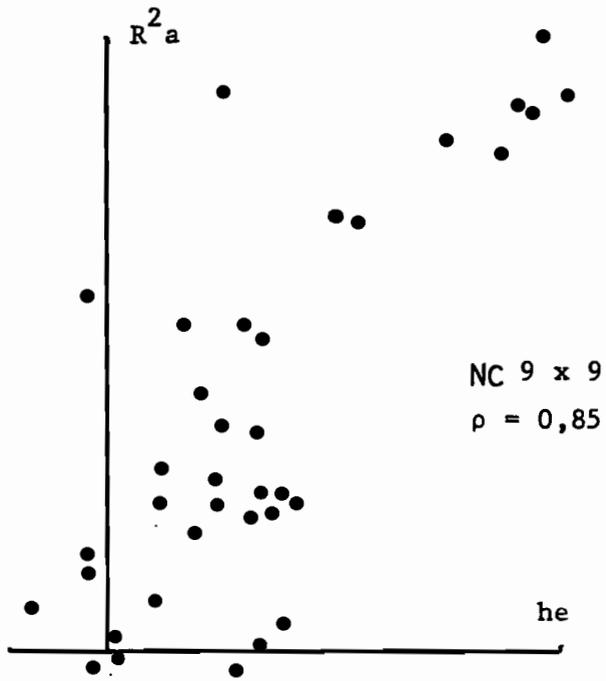


FIG. 4 : suite

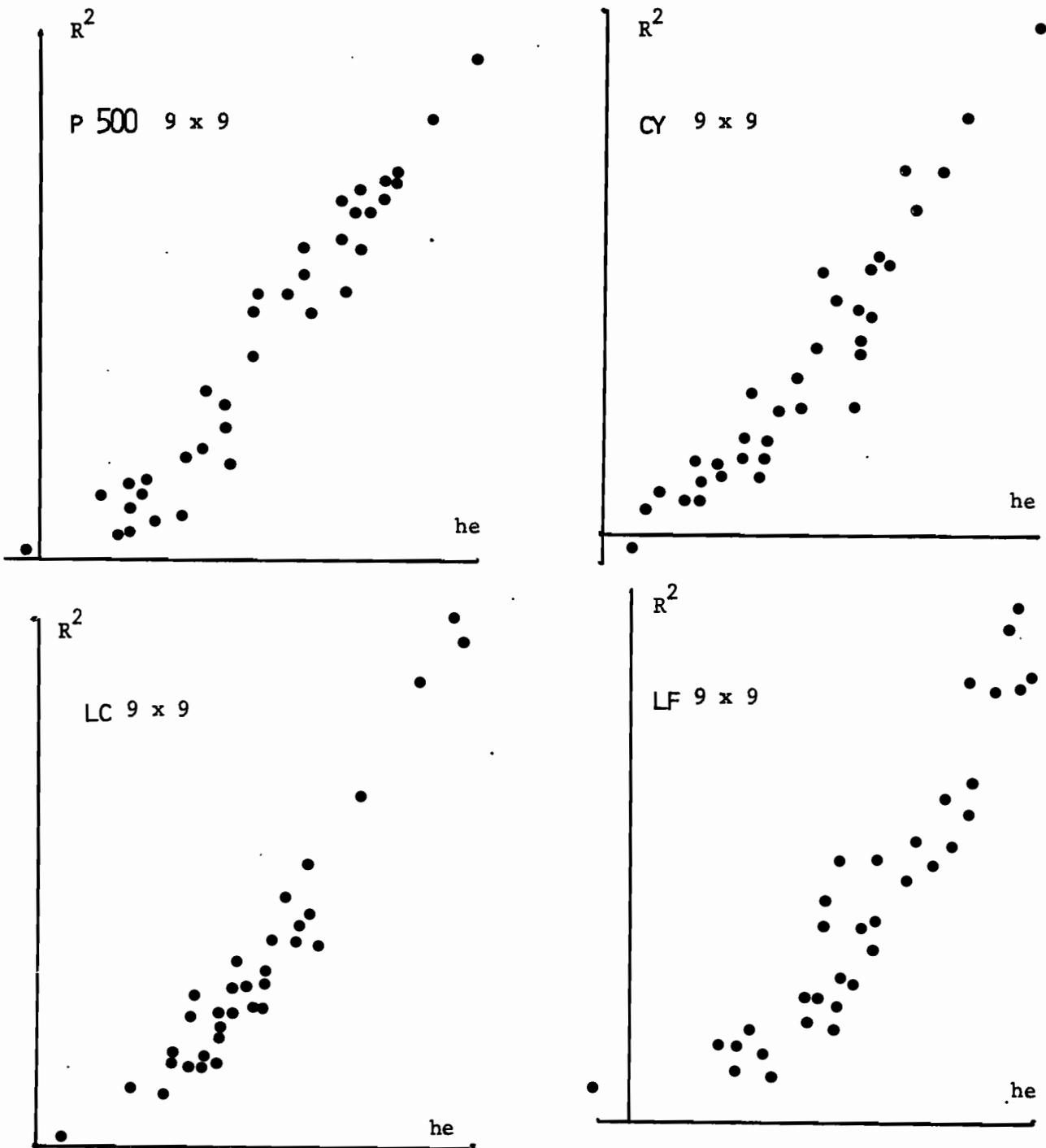


FIG. 4 BIS : Relation entre l'hétérosis et les distances R^2 brutes. La courbe en S très nette sur les figures précédentes pour P 500 et CY avec les R^2 ajustés disparaît ici. La concavité est ici plus forte pour LC et LF.

verra plus loin.

La figure 4bis montre les graphiques des relations entre l'hétérosis et les distances R^2 brutes, non ajustées, pour illustrer l'effet des covariances par comparaison avec les figures 4, ce pour les caractères P500, CY, LC, LF.

8. Comparaison H1 Triple, H1 Diallele, R^2 a Max et Triple, D Diallele

Rappelons que suivant les formules (8) :

$$\begin{aligned} \text{H1 Triple} &= V \sum a (ha+i/2+\frac{N-1}{2} la)^2 && \text{parents en association.} \\ &\text{ou } V \sum a (ha+\frac{N-1}{2}(la-ia))^2 && \text{parents en dispersion.} \end{aligned}$$

$$R^2_{a, \text{Max}} = \sum a (ha+i/2+\frac{N-1}{2} la)^2$$

On attend, grossièrement, $VR^2_{a\text{Max}} = \text{H1 Triple}$

et $\text{H1 Diallele} = R^2_{\text{Max}}$ non ajusté.

Caractère	H1 Triple	$VR^2_{a \text{ Max}}$	H1 Diallele	R^2_{Max}	Triple D	Diallele
CY	1060	452	429	665	740	208
LF	909	428	542	607	582	334
lf	520	585	675	685	919	953
LT	1445	1993	2241	2985	4550	522
PT	168	441	372	554	335	398
NC	-	68	61	86	-	90
PRO	-	779	519	787	-	488
LC	426	854	627	1118	1471	246
PC	390	622	234	711	2290	982
NE	41	35	28	32	105	28
P500	6,3	7,0	11,6	11,5	3,3	1,3
PG	-	86777	93134	129147	-	2014

Tableau 9.

L'accord entre H1 Triple et $VR^2_{a \text{ Max}}$ est assez bon pour lf, LT, NE, P500 sur la base de parents P, P' en dispersion (voir tableau 9)

L'accord pour CY et LF est aussi bon mais si on suppose $ia\#$ la et des parents P, P' en association. Cela expliquerait un test d'épistasie Triple non significatif. Par contre, il y a désaccord pour PT, LC, sans explication évidente.

H1 Diallele est proche de R^2_{Max} non ajusté pour LF, lf, NE, P500. Pour les autres caractères, H1 Diallele est trop faible par rapport à ce que prévoit la théorie. Ce hiatus pourrait s'expliquer par un excès de couples voisins et une moyenne $\overline{R^2}$ trop basse : les lignées ne seraient pas tirées au hasard de la F2 mais auraient subi une sélection surtout pour LC et PC.

Les estimations D Triple et D Diallele sont dans l'ensemble très divergentes sans explication évidente. Remarquons que D Triple utilise uniquement des observations sur des croisements (variance de $Y_{pm} + Y_{p'm}$) alors que D Diallele n'utilise que les lignées autofécondées (variance Y_m). Il y aurait peut être un effet de la consanguinité non pris en compte par notre modèle.

9. Estimation du nombre de gènes N.

Comme il a été expliqué dans la théorie, N est ici estimé avec les valeurs $R^2_{a Max}$, $H_e Max - H_e Min$ et une distribution des effets de dominance en valeur absolue $|h|$ du type (1,2,3...). Compte tenu de l'inégale répartition des points ($R^2_{a He}$) sur les graphiques 4, les valeurs moyennes $\overline{R^2}_{a}$ et $\overline{H_e}$ ne sont pas utilisées.

Les estimations obtenues sont les suivantes :

Caractère :	CY	LF	lf	LT	PT	NC	PRO	LC	PC	NE	P500	PG
N	6,3	7,6	7,6	6,7	5,3	6,5	6,7	5,9	5,6	12,5	8,1	6,5

A part NE, tous les caractères semblent régis par un nombre de "gènes" voisin de 7 à savoir le nombre haploïde des chromosomes du mil .

10. Estimation du degré de dominance apparente.

D'une façon pragmatique, un degré de dominance apparente est calculée parallèlement avec $\sqrt{H1 \text{ Triple}/D \text{ Triple}}$ et $\sqrt{VR^2 a \text{ Max}/D \text{ Diallele}}$

Caractère	$\sqrt{\frac{H1 \text{ Triple}}{D \text{ Triple}}}$	$\sqrt{\frac{VR^2 a \text{ Max}}{D \text{ Diallele}}}$
CY	1,2	1,5
LF	1,2	1,1
lf	0,7	0,8
LT	0,6	1,9
PT	0,7	1,0
MC	-	0,9
PRO	-	1,3
LC	0,5	1,8
PC	0,4	0,8
NE	0,6	1,1
P500	1,4	2,3
PG	-	6,6

Pour LC, LT et NE les deux estimations sont trop discordantes pour qu'une conclusion soit possible.

Pour P500 et PG, la superdominance est nette.

CY et LF manifestent une légère superdominance.

Les autres caractères manifestent une dominance intermédiaire.

11. Construction d'une distance générale R^2_g .

La formule (14) s'est avérée pratiquement inutilisable pour estimer N_{ij} , le nombre de locus communs à deux caractères i et j : les valeurs numériques sont souvent négatives.

Ce résultat s'explique par l'invalidité du modèle d'action génique envisagé : il n'y a pas deux classes de locus ceux où la dominance est nulle et ceux où la dominance a une valeur constante non nulle, mais des locus avec des degrés de dominance variables et dans un ordre variable suivant le caractère considéré.

Une approche plus vague et plus réaliste a été tentée : la considération des corrélations entre distances R^2_{ai} et R^2_{aj} . Estimant avec quelque raison que les différents caractères sont régis en gros par le même nombre de facteurs génétiques dont les effets varient d'un facteur à l'autre, la corrélation indique pour un couple de caractères si les facteurs à forte action sont les mêmes ou non pour les deux caractères.

Appelons h_{ai} la dominance apparente du facteur A sur le caractère i. La corrélation (R^2_{ai}, R^2_{aj}) sera forte si pour tout couple A, A', on a :

$$\frac{h^2_{ai} - h^2_{a'i}}{h^2_{aj} - h^2_{a'j}} > 0 \text{ ou si corrélation } (h^2_{ai}, h^2_{aj}) \neq 1$$

Ce résultat a été démontré numériquement dans l'analyse des corrélations ($\Sigma h, \Sigma h^2$).

Les corrélations entre distances ajustées du tableau 10 montrent que 4 caractères suffisent à exprimer la variabilité de tous les facteurs génétiques : P500, NC, CY, NE. CY et NE s'expriment surtout par des facteurs qui leur sont spécifiques. Ils ont une corrélation faible avec les autres caractères.

P500 au contraire a les mêmes facteurs prépondérants que LT, lf, LC, PC, PT, LF, PG.

Une distance générale a donc été construite en additionnant les distances standardisées de P500, NC, CY, NE.

Caractère	LT	lf	NE	CY	LC	PC	P500	NC	PRO	PT	LF	PG
LT	1,00	0,18	0,26	0,52	0,22	0,04	0,40	0,36	0,37	0,23	0,24	0,51
lf	0,18	1,00	0,17	0,08	0,56	0,62	0,63	0,34	0,34	0,55	0,57	0,57
NE	0,27	0,17	1,00	-0,01	0,22	0,08	0,28	0,05	0,40	0,12	0,13	0,39
CY	0,52	0,08	-0,01	1,00	0,19	-0,02	0,07	0,27	0,30	0,08	0,08	0,22
LC	0,22	0,56	0,22	0,19	1,00	0,74	0,66	-0,03	0,21	0,43	0,46	0,35
PC	0,04	0,62	0,08	-0,02	0,74	1,00	0,55	-0,08	0,13	0,34	0,26	0,17
P500	0,40	0,63	0,28	0,07	0,66	0,55	1,00	0,08	0,18	0,61	0,51	0,60
NC	0,36	0,34	0,05	0,27	-0,03	-0,08	0,08	1,00	0,42	-0,03	0,15	0,58
PRO	0,38	0,34	0,41	0,30	0,20	0,13	0,18	0,42	1,00	-0,01	0,32	0,22
PT	0,23	0,55	0,12	0,08	0,43	0,34	0,61	-0,03	-0,01	1,00	0,59	0,38
LF	0,24	0,58	0,13	0,08	0,46	0,26	0,51	0,15	0,32	0,59	1,00	0,30
PG	0,51	0,57	0,39	0,22	0,35	0,16	0,59	0,58	0,22	0,38	0,30	1,00
R^2_g	0,62	0,54	0,67	0,44	0,47	0,48	0,67	0,52	0,55	0,37	0,39	0,77

Tableau 10 : Corrélations (R^2_i, R^2_j) 9 x 9

Rappelons la définition d'une distance standardisée n :

$$n = R^2_a \times N/R^2_a \text{ Max}$$

$$R^2_g = n (P500) + n (NC) + n (CY) + n (NE)$$

La standardisation tient compte des différences d'échelle dans la mesure des effets génétiques.

La distance générale ainsi calculée montre de fortes corrélations avec tous les caractères : elle est ainsi justifiée dans son rôle de mesure de tous les facteurs différents, indépendante des variations d'intensité des effets génétiques.

Cette distance générale présente en outre l'intérêt pratique d'une forte corrélation (0,77) avec PG.

12. Analyse des distances dans le diallèle 6 x 6.

Les formules données dans la partie théorique pour la distance et l'hétérosis ne s'appliquent plus ici à cause d'une distribution à priori quelconque des allèles aux divers locus et à cause du multiallélisme probable.

Néanmoins, faute de mieux, l'analyse de HAYMAN et l'analyse des distances ont été conduites comme dans le diallèle 9 x 9 pour voir si des conclusions pouvaient être dégagées.

Les graphiques N° 5 des hétérosis et des distances ajustées montrent des distributions nettement plus irrégulières que dans le diallèle 9 x 9. Ainsi pour CY et PG, il manque des points intermédiaires entre l'origine et le nuage observé. Chez LC et PC, il manque des points au milieu du nuage. Ces irrégularités traduisent sans doute un nombre d'hybrides inférieur à celui du diallèle 9 x 9, mais aussi une plus grande asymétrie dans les distributions alléliques. Le groupement des points assez loin de l'origine est aussi un indice de diversité génétique supérieure à celle de 9 x 9.

Les nuages dispersés de PRO et LT et quelques hétérosis négatifs sont un indice de dominance apparente légèrement ambidirectionnelle.

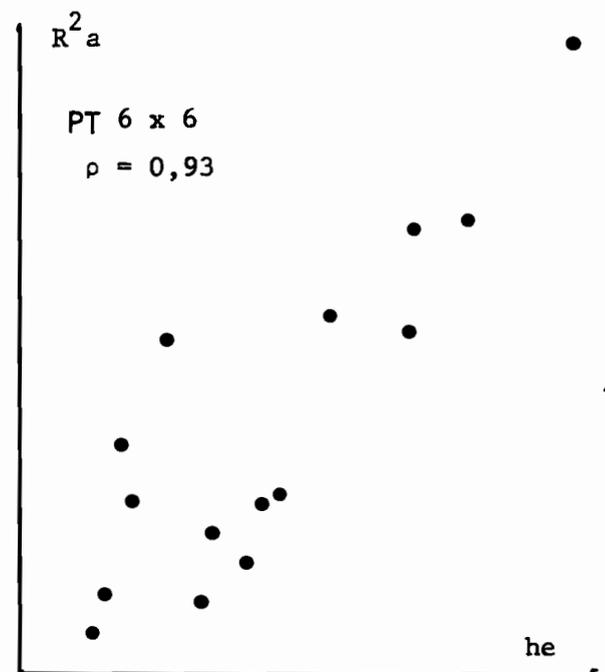
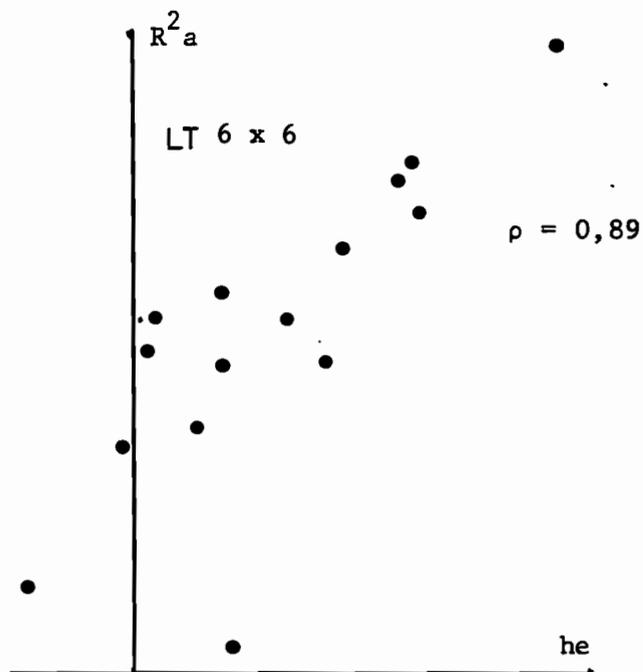
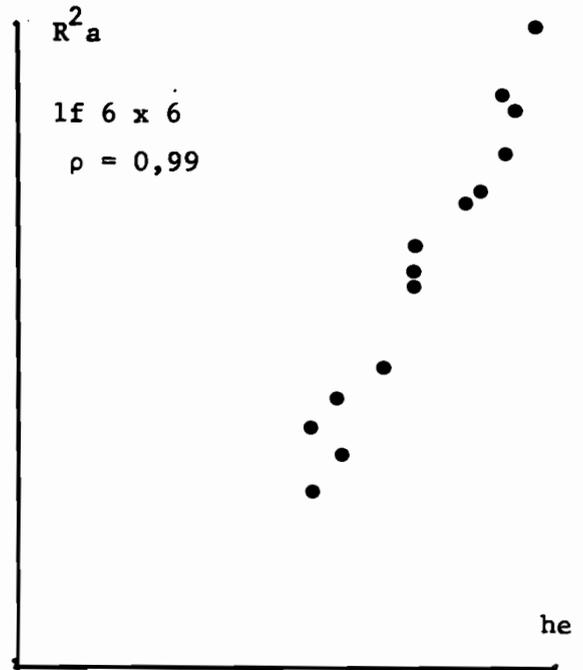
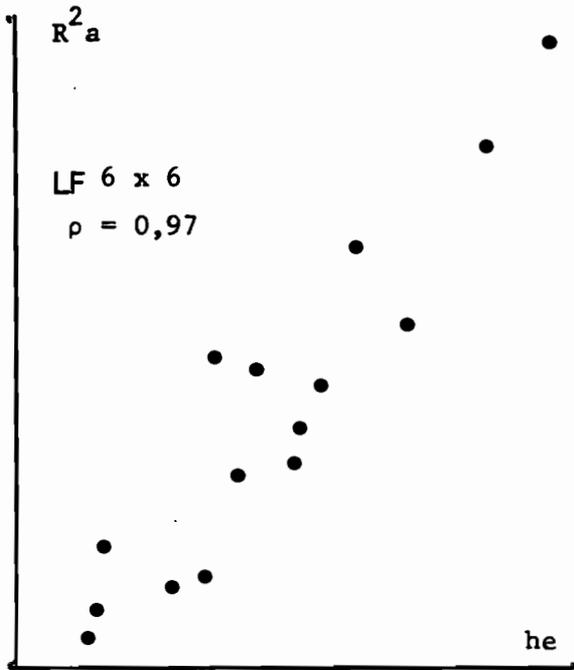


FIG. 5 ; Relations He, R^2_a dans le diallèle 6 x 6
 ρ = Corrélations (He, R^2_a non ajusté)

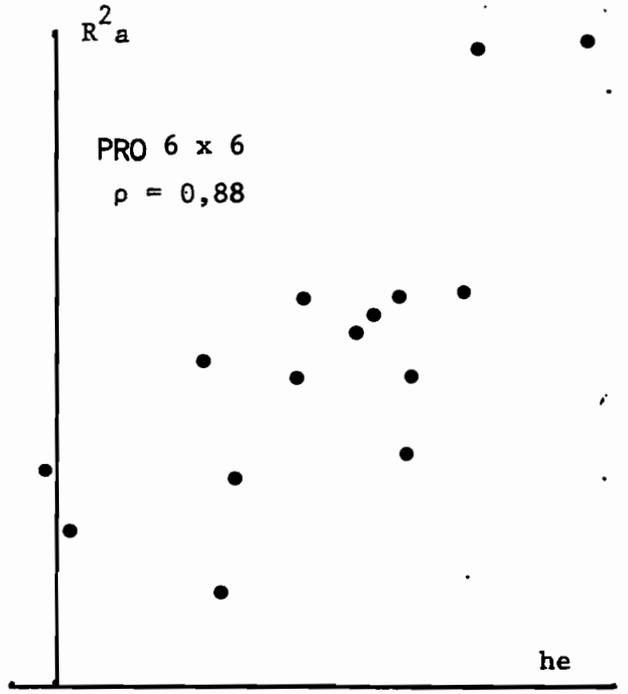
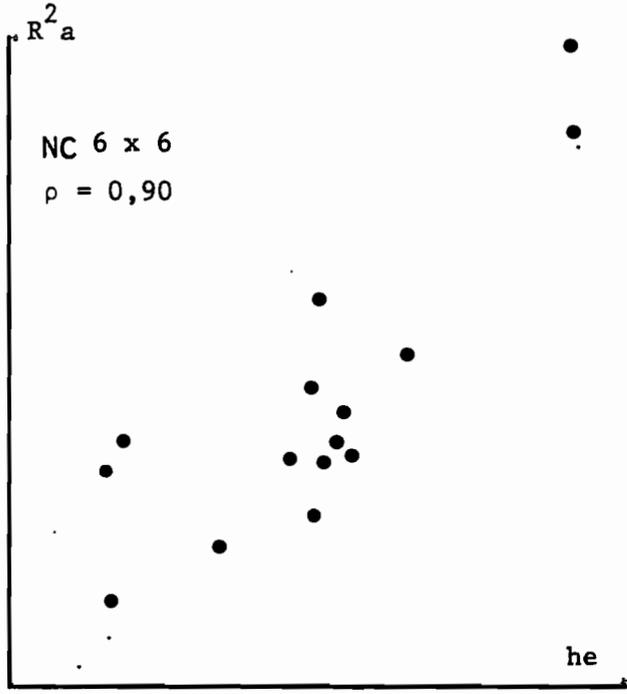
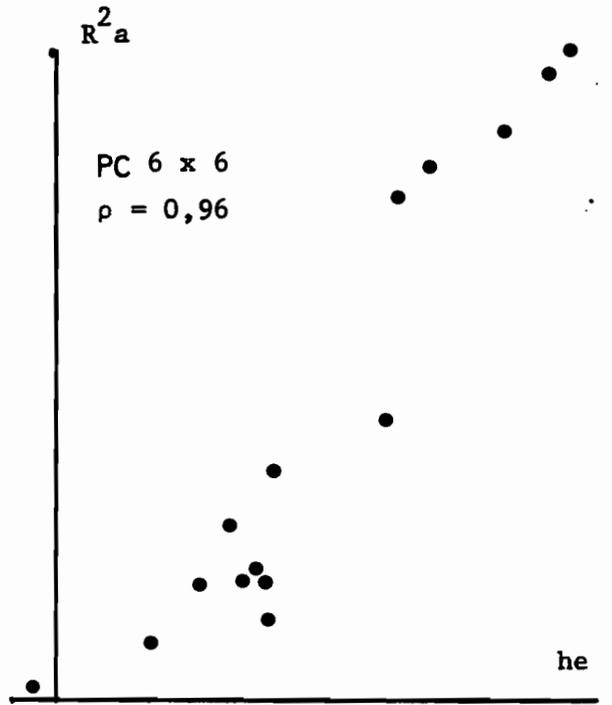
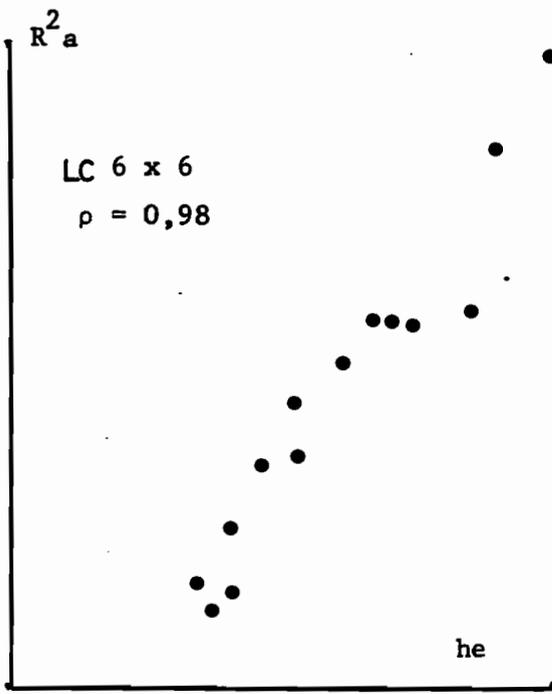


FIG. 5 : suite

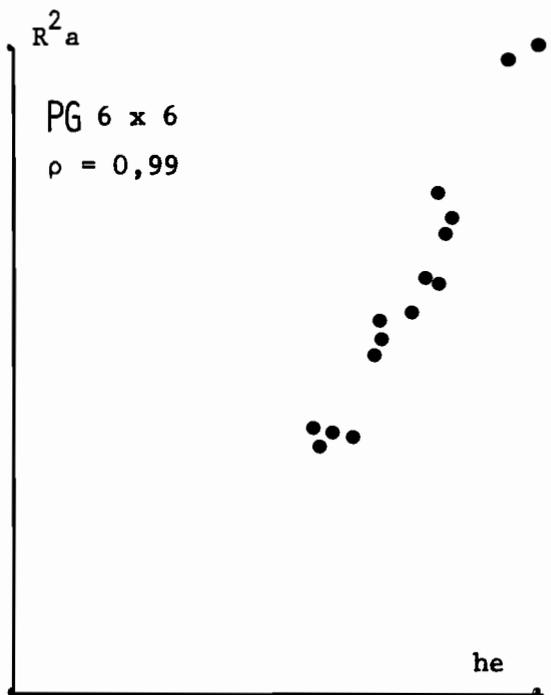
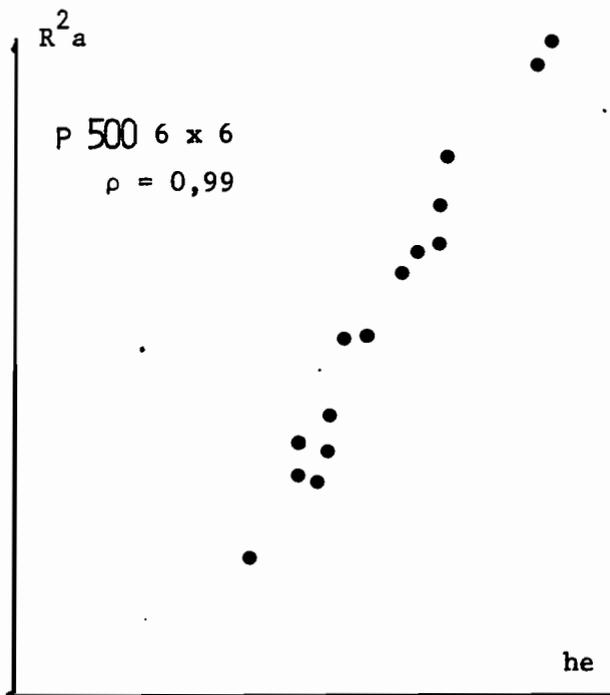
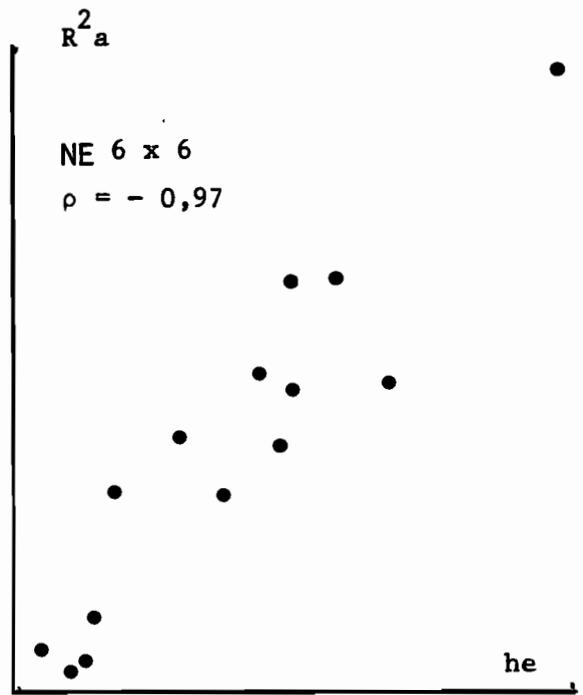
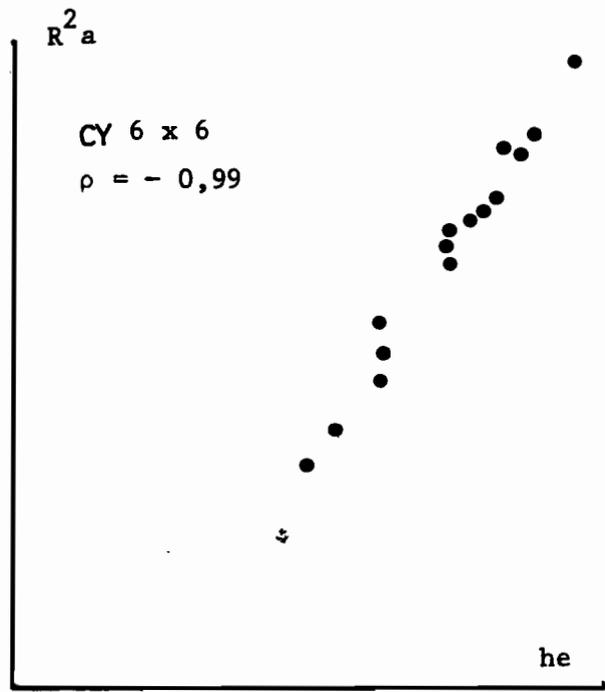


FIG. 5 : suite

Les autres caractères montrent une dominance toujours de même signe.

Dans l'ensemble, les courbes correspondent à une épistasie dupliquée sauf pour LF et NE pouvant relever d'effets de dominance d'intensité constante, sans épistasie.

Les pentes de régression (W_r/V_r) portées dans le tableau 11 avec la signification statistique révèlent des extravagances :

Des caractères comme LT ont une pente de 0,33 qui n'est pas significativement différente de 1 tellement les points sont dispersés.

Caractère	b(W_r/V_r)	D		$\frac{VR^2 a_{Max}}{D}$	N
		9x9	6x6		
CY	0,93	208	431	1,4	8,1
LF	0,13*	334	505	2,0	4,5
lf	0,98	953	950	1,0	7,7
LT	0,33	522	1920	1,6	6,7
PT	0,39*	398	295	1,5	4,4
PRO	0,67*	488	962	1,6	4,0
NC	0,49	90	49	1,9	4,3
LC	0,62	246	590	1,2	5,4
PC	0,75	982	462	1,5	4,0
NE	0,51	28	132	0,8	3,9
P500	0,62*	1,3	2,5	3,6	6,8
PG	0,06*	2014	2950	11,8	7,8

Tableau 11 : Paramètres génétiques 6 x 6

La variance D des lignées 6 x 6 dépasse celle des lignées 9x9 pour CY, LF, LT, PRO, LC, NE, P500 et PG. Pour lf, les deux variances sont égales. Pour PT, NC et PC les variances 6x6 sont inférieures. Dans l'ensemble, les lignées 6x6 apparaissent ici plus diverses que les lignées 9x9.

Le degré de dominance apparente estimé par $R^2_{a \text{ Max/D}}$ apparaît dans l'ensemble plus fort chez 6x6. A part lf et NE, tous les caractères montrent de la superdominance.

Cette différence est à imputer partie à la consanguinité, partie à la diversité génétique, plus fortes toutes deux dans les lignées 6x6.

Le nombre N de facteurs génétiques varie de 4 à 8 avec en moyenne une valeur inférieure à celles du diallèle 9x9. Apparemment, ici aussi, tous les caractères quantitatifs mettent en jeu un nombre à peu près égal de facteurs génétiques.

Les corrélations entre distances ajustées du tableau 12 indiquent le degré de pléiotropie apparente.

Contrairement à CY 9x9, ici CY 6x6 montre une pléiotropie avec tous les autres caractères sauf NE et PC.

P500 et PG restent des caractères ayant les attributs d'une distance générale : ils expriment bien les distances particulières à chaque caractère.

LC et NE montrent une association plus nette que dans le diallèle 9x9, et qui paraît logique : les facteurs qui règlent la longueur de chandelle doivent agir aussi sur la densité des épillets.

PC et NE apparaissent des caractères pour lesquels les facteurs génétiques prépondérants sont spécifiques (sans pléiotropie).

Une distance générale a donc été construite avec la formule $R^2_g = n \text{ CY} + n \text{ PC} + n \text{ NE}$. Cette formule n'est sans doute pas la meilleure mais elle se montre capable d'après ses corrélations d'assez bien exprimer les distances particulières. Elle apparaît également capable de prédire le rendement PG.

Vb. EVALUATION DU RENDEMENT SUR LES CARACTERISTIQUES DE PIEDS INDIVIDUELS

La régression multiple du Poids de grain PG sur les caractères de végétation LF, lf, LT, PT, PRO, NC a été calculée dans les 45 familles du diallèle 9x9 et les 15 familles F1 du diallèle 6x6. Le test de l'hétérogénéité des régressions est non significatif au seuil 1% dans le diallèle 9x9 ($F_{264,135} = 1,37$) mais significatif dans le diallèle 6x6 ($F_{84,735} = 2,45$)

La reprise du calcul de la régression multiple de PG sur les caractères de végétation après transformation de ceux-ci en composantes principales a d'une part révélé la structure de la variabilité de la végétation à l'intérieur d'un même génotype et d'autre part révélé dans le diallèle 6x6 l'existence d'une régression commune sur la première composante de végétation C1.

Le tableau 13 donne pour chaque composante la combinaison linéaire des caractères initiaux dans laquelle les coefficients sont les coefficients de corrélation entre la composante et le caractère.

Composante	NC	PRO	PT	LT	lf	LF	Valeur propre %	
Diallèle 9x9								
C1	0,57	0,42	0,61	0,57	0,69	0,72	2,18	36
C2	-0,60	-0,59	0,39	-0,27	0,36	0,34	1,34	22
C3	0,01	-0,51	-0,20	0,67	-0,20	0,14	0,86	14
C4	0,06	-0,01	-0,62	-0,17	0,35	0,29	0,62	10
C5	-0,36	0,33	-0,05	0,07	-0,30	0,38	0,51	8
C6	0,36	-0,19	0,09	-0,31	-0,33	0,31	0,47	8
Diallèle 6x6								
C1	0,66	0,36	0,66	0,47	0,77	0,66	2,28	38

Tableau 13 : Composantes de végétation.

Dans les deux diallèles, la première composante traduit de 36 à 38% de la variabilité intragénotypique de la végétation. Sa structure est la même : elle traduit la variation quantitative de la végétation. Tous les coefficients sont forts et de même signe. Cette composante C1 qui représente la masse de végétation est celle qui nous intéresse le plus comme indice d'occupation du sol.

Les autres composantes traduisent des variations qualitatives de la végétation, sans doute moins liées à l'occupation du sol. Le test de l'hétérogénéité des régressions PG 6x6 sur C1 6x6 donne $F_{14,810} = 1,86$ significatif au seuil 1%.

Etant donné le grand nombre de degrés de liberté de l'erreur, on a un test très puissant : on peut considérer ici que les hétérogénéités ne sont pas très fortes puisque non significatives au seuil 5%.

La part des différentes composantes dans la régression commune apparaît dans le tableau 14.

Composante	Coefficient de régression		% de la régression commune	% de la variation PG
	original	standardisé		
Diallèle 9x9				
C1	315**	0,43	59	19
Ensemble Diallèle 6x6			100	32
C1	1419**	0,71	100	50

Tableau 14 : Régressions communes.

C1 9x9 qui n'emportait que 36% de la variation de végétation emporte 59% de la régression commune.

La seule régression sur C1 6x6 représente 50% de la variation PG 6x6 intragénotypique alors que la régression sur la végétation totale 9x9 ne représente que 32% de la variation PG 9x9.

Ce résultat signifie que les accidents survenus à partir du stade floraison (sécheresse, attaques s'insectes, stérilités génétiques) sont plus rares dans le diallèle 6x6 que dans le diallèle 9x9.

C'est un indice de meilleure vigueur d'ordre génétique quant aux caractères de production directe du grain.

Ce résultat signifie aussi que la variance erreur du Poids ajusté à un volume donné de végétation vaut respectivement 68% et 50% de la variance erreur de PG 9x9 et PG 6x6.

Les nouveaux caractères "Poids ajusté" des deux diallèles ont été soumis aux mêmes analyses que les autres caractères.

Il apparaît ce résultat essentiel :

Dans le diallèle 9x9, la variance génétique de Poids Ajusté représente 19% de la variance génétique de PG alors que dans le diallèle 6x6 cette proportion est de 99%.

Ce résultat signifie que les lignées 9x9 contiennent une diversité génétique surtout pour les caractères de végétation mais peu pour les caractères de production de grain non liés aux précédents : fertilité, nombre et grosseur des grains, taille des chandelles etc.. Au contraire, les lignées 6x6 à volume donné de végétation contiennent une forte diversité génétique pour la production de grain. La sélection pour un bon rapport grain sur paille, un bon rendement/ha apparaît plus prometteuse avec les lignées 6x6 qu'avec les lignées 9x9.

Le second résultat important de l'analyse génétique est la réduction des interactions géniques dans le mécanisme de l'hérédité du Poids ajusté par rapport à PG. Ceci apparaît au niveau de l'hétérosis :

. L'hybride moyen dépasse le parent moyen de 43% et 49% dans les diallèles 9x9 et 6x6 au lieu de 240% et 790% pour les poids de grain PG.

. La F1 maximale dépasse le parent maximal de 58% et 27% au lieu de 290% et 660%.

. Dans l'Anova 6x6, la part des effets (1, li, lij) dans la variation génotypique vaut 63% au lieu de 97% pour PG. Mais chez le diallèle 9x9, la part des effets non-additifs reste du même ordre que celle de PG.

Ces niveaux d'hétérosis sont plus conformes à ceux observés en vrai grandeur dans la pratique concernant le rendement/ha.

Reste à contrôler la valeur du Poids ajusté comme prédicteur du rendement/ha. L'expérimentation agronomique nécessaire n'a pu être conduite.

Vc. MAXIMISATION D'UN SYNTHÉTIQUE.

Cette procédure n'a été conduite que sur le diallèle 9x9 et le synthétique maximisé a été testé en essai agronomique.

La maximisation conduite sur les Poids de grain par pied du tableau diallèle complet n'a pas eu de solution : la matrice T de la théorie (4ème chapitre) n'est pas définie négative.

La maximisation a réussi sur un sous-ensemble du tableau diallèle ne comprenant que les lignées L1 - L2 - L6 - L8, choisies comme prometteuses.

Les fréquences obtenues sont respectivement : 0,39-0,15-0,22 et 0,24, pour un rendement de 116 grammes par pied.

Le synthétique (L1 - L2 - L6 - L8) constitué avec les fréquences (0,25-0,25-0,25-0,25) donne un rendement théorique de 112 grammes par pied.

Il est remarquable de voir que la maximisation n'a pas beaucoup modifié le rendement du synthétique (L1 - L2 - L6 - L8) mais par contre la composition du synthétique maximal est très différente de celle du synthétique à fréquences égales. Cette différence de composition est susceptible de se manifester fortement dans les générations ultérieures

de multiplication.

Il n'a pas été possible de suivre expérimentalement l'évolution de ces synthétiques.

Toutefois, le synthétique de 1ère génération maximale a été fabriqué, dans les proportions panmictiques correspondant aux fréquences des parents (0,39-0,15-0,22-0,24), par culture en fécondation libre dans un compartiment isolé des 4 lignées et de leurs 6 hybrides. Pour empêcher toute sélection naturelle prématurée à ce stade, la même quantité de semences a été récoltée sur chaque pied.

Le synthétique de deuxième génération ainsi obtenue a été testé en essai agronomique par rapport à un synthétique témoin composé à fréquences égales avec les 9 lignées.

Les rendements à l'hectare obtenus sont confrontés dans le tableau 15 aux rendements par pied théoriques calculés sur le diallèle de 1ère génération.

Variété	Rendement			
	Essai diallèle		Essai agronomique	
	g/pied	%	Kg/ha	%
Témoin : Diallèle 9x9	88	100	2100	100
Synth.Max (L1-L2-L6-L8)	116	131	2600	123
F1 maximale	178	201		

Tableau 15.

Il apparaît que la supériorité du synthétique maximum sur le témoin en 1ère génération (mesurée en g/pied) se maintient en 2ème génération (mesurée en kg/ha) sans qu'il soit possible de préciser quantitativement si cette supériorité tend à augmenter ou à diminuer.

VI. CONCLUSIONS.

Les résultats accumulés dans les travaux ici exposés permettent de tirer quelques conclusions générales concernant les méthodes d'analyse génétique, les techniques de sélection, la génétique quantitative du petit mil et la valeur des lignées étudiées pour le sélectionneur.

A. Méthodes et Techniques.

1. Simplification et amélioration de l'analyse classique d'un croisement dialléle

L'analyse classique est simplifiée par l'utilisation de l'Anova d'Eberhart qui, outre une décomposition statistique de l'hétérosis d'un grand intérêt pratique fournit immédiatement les estimations de D, H1, H2, h2.

L'analyse classique est améliorée par le calcul des distances qui apporte des informations sur la constitution génétique relative des lignées.

Ceci dit, ces avantages sont plutôt formels dans la mesure où l'analyse classique de Hayman-Jinks est actuellement refusée par tous les généticiens modernes à cause de ses hypothèses simplistes.

2. Possibilité d'une sélection plus efficace.

L'existence d'une forte corrélation entre la distance R^2 et l'hétérosis dans un ensemble donné de génotypes devrait permettre une bonne prédiction de l'hétérosis entre génotypes sans avoir à effectuer leur croisement.

Le sélectionneur devrait au minimum doubler ses capacités d'observation. Considérons, en effet, le schéma de croisement N° 2 de la Caroline du Nord (COMSTOCK et coll. 1948) dans lequel N unités A sont croisées avec N unités B. La valeur des N^2 hybrides AxB est directement observée et permet en outre de calculer les distances donc d'estimer les hétérosis entre AxA et BxB. Au total N^2 croisements permettent d'évaluer N (2N-1) hybrides. L'économie est d'environ 50%. Elle serait encore plus forte dans un schéma de croisements asymétriques N A x n B (N supérieur à n).

Cette méthode implique le contrôle préalable de la corrélation hétérosis-distance et une variabilité égale du groupe A et du groupe B à chaque locus. Ces conditions signifient que les groupes A et B doivent être issus d'un même ensemble de génotypes "compatibles" entre eux, dans lequel n'apparaissent pas de phénomènes de stérilité ou de chute des relations de dominance.

Cette méthode de sélection est donc inapplicable à des populations très divergentes. Un travail préalable de panmixie sur ces populations est nécessaire pour produire la superpopulation qui elle sera soumise à la sélection.

Un diallèle récent entre populations de mil très divergentes a effectivement produit des corrélations nulles pour tous les caractères.

3. Possibilité d'une taxonomie plus fine.

Les diverses méthodes de la taxonomie numérique utilisent des informations abondantes mais limitées aux phénotypes des unités à classer (SOKAL et coll. 1963 ; RAO 1952 ; GOWER 1971, etc...).

Ces méthodes permettent de classer des effectifs importants avec efficacité pour la majorité des unités à classer. Il n'est pas pensable de remplacer ces méthodes par la distance générale R^2g pour deux raisons :

- la distance R^2g ne permet de classer qu'une vingtaine d'unités à la fois à cause du laborieux effort de croisements.
- la distance R^2g ne peut que donner les mêmes résultats que la taxonomie numérique classique pour des unités clairement divergentes.

Ce fait a pu être vérifié expérimentalement et récemment avec des échantillons de mil extrêmement différents par la morphologie et l'origine géographique.

Une corrélation de 0,90 (pour 15 couples) a été observée entre R^2_g et la distance phénotypique.

Par contre, pour analyser plus en détail des unités peu nombreuses et phénotypiquement très proches, la distance R^2_g devrait être utile et mieux renseigner sur les différences génétiques que les méthodes classiques.

Les conclusions opposées apportées par CERVANTES (1978) ne sont pas convaincantes. Il a classé 25 populations de maïs mexicains par R^2_g , par les aptitudes générales à la combinaison pour 21 caractères et par les interactions génotype x milieu. La classification par les aptitudes générales lui paraît la meilleure et la classification sur les aptitudes spécifiques la moins bonne. Ce résultat peut fort bien s'expliquer par une erreur méthodologique : 16 caractères sur 21 ont été mesurés sur des hybrides provenant de localités et d'années différentes.

Les différences Sit-Sjt deviennent inconsistantes et les R^2 quelconques. Il n'est pas étonnant dans ce cas que les 6 premières valeurs propres de la matrice des corrélations (Sit-Sit') soient toutes égales : les échantillons de maïs n'ont pas de structure, ils se répartissent dans un nuage sphérique.

Les aptitudes générales au contraire sont moins sensibles à l'hétérogénéité des lieux d'observation parce qu'elles comportent toutes approximativement le même effet moyen du milieu.

Il semble en définitive que pour la taxonomie la distance totale (distance sur aptitudes générales + distance sur aptitude spécifique) utilise au mieux l'information disponible.

Les méthodes taxonomiques les meilleures sont sans doute celles utilisant les fréquences alléliques de gènes connus (NEI 1975). Malheureusement, pour une espèce donnée, le nombre de ces gènes identifiés est généralement restreint et ces caractères à sens génétique précis ne peuvent que servir d'appoint aux caractères polygéniques des méthodes précédentes.

4. Possibilité d'analyse génétique de l'hétérosis.

Les observations du diallèle 9x9 conjuguées aux relations théoriques établies entre l'hétérosis et la distance ont permis de préciser le mode d'action des différents gènes actifs.

L'existence d'effets épistatiques opposés aux effets de dominance s'accorde bien avec les observations de POWERS sur la tomate et de JINKS sur *Nicotiana rustica* (MATHER et JINKS, 1971), de LAW sur blé (LAW et coll. 1978), de GAMBLE sur maïs (1962).

De même, GALLAIS (1977) a pu observer sur luzerne une exaltation par la consanguinité des effets de dominance apparente qui semble assez bien expliquée par la relation (7) de l'hétérosis :

Si deux lignées ne diffèrent que par un seul locus, la dominance vaut h . Si deux lignées diffèrent par N locus, la dominance apparente vaut $h + (N-1) 1/2 < h$.

Cela s'accorde aussi avec la théorie physiologique de la dominance de WRIGHT (1977) et le modèle génique de FORKMANN et SEYFFERT (1977). Ces théories prévoient la relation entre la valeur génotypique et le nombre de facteurs génétiques sous forme d'une courbe de saturation.

L'utilisation des statistiques d'hétérosis et de distances est surtout intéressante dans une situation précise : diallélisme, fréquences égales, distributions géniques proches de l'indépendance et surtout corrélation forte entre hétérosis et distance.

Un schéma N° 2 de la Caroline du Nord est recommandé pour mesurer les distances parce qu'il diminue les effets de covariance.

L'utilisation des statistiques d'hétérosis recueillies dans deux diallèles $A \times A$ et $B \times B$ et des statistiques de distances recueillies dans un croisement N° 2 $A \times B$ devrait permettre de suivre l'évolution de l'hétérosis avec des distances croissantes dans un intervalle étendu et d'observer éventuellement une chute des relations de dominance.

5. Valeur du Poids de grain ajusté à un volume donné de végétation

Il est apparu que la première composante principale des caractères de végétation correspondait à la masse de végétation, à la paille.

Il est apparu que la régression du Poids de grain sur cette composante était commune à tous les génotypes, en gros.

Le Poids de grain ajusté a montré une variance génétique très différente entre les deux essais, des valeurs d'hétérosis plus conformes aux rendements/ha observés, sans doute une héritabilité plus forte que le Poids brut.

Ces propriétés font du Poids ajusté un critère de sélection pour le rendement/ha très prometteur pour une sélection massale dans des populations d'architecture variées, préalable à une sélection plus fine et plus laborieuse.

Ce critère devrait permettre d'évaluer l'intérêt de diverses architectures pour le rendement en évitant des essais agronomiques lourds et nombreux, vu l'extrême diversité des morphologies de pieds de mil actuellement à la disposition du sélectionneur.

Ce critère constitue une sorte de synthèse des deux critères de sélection proposés par DONALD et HAMBLIN (1976) : le rendement biologique (grain + paille) et le taux de récolte (harvest index) compte-tenu de la signification de la principale composante de végétation, il faudrait utiliser une seule mesure de la végétation : la paille.

6. Valeur du synthétique maximisé

Le synthétique maximisé a été construit avec des fréquences très inégales. Il a montré une nette supériorité sur le diallèle complet pris comme témoin. Reste à savoir si ce synthétique et un synthétique fabriqué avec les mêmes constituants à fréquences égales auraient divergé profondément dans les générations ultérieures. Il n'est, cependant, pas évident que les formules habituelles à fréquences égales soient les meilleures.

B. Génétique quantitative du petit mil

1. Héritabilité des caractères

Il semble que les phénomènes d'épistasie dupliquée soient fréquents dans le mécanisme héréditaire des divers caractères quantitatifs mais

que leur mise en évidence ne soit pas toujours claire dans les divers tests existants à cause d'une confusion avec les effets d'additivité ou de dominance.

Apparemment, le modèle sans épistasie semble bien approcher les phénomènes. Il faut alors parler de dominance et d'additivité apparentes.

Dans les deux essais, les caractères d'architecture ont présenté un mécanisme assez simple, en l'absence d'effets de compétition entre plantes : un hétérosis modéré (moins de 15%), une part des effets non-additifs modérée dans la variance génétique (moins de 70%), une dominance apparente modérée.

Ces caractères sont par conséquent à forte héritabilité et se prêtent à une sélection pédigrée.

Seul le Poids de grain s'oppose franchement à ce schéma : il présente un hétérosis extrême, de fortes interactions géniques, une forte dominance apparente.

Les caractères P500 et NC occupent une position intermédiaire.

La validité d'un modèle génétique simple (additivité + dominance complète) est confirmée par le bon accord entre les prévisions fondées sur l'analyse des lignées 6 x 6 dans le graphique n° 2 et les résultats des croisements 6 x 6.

Les lignées 1 - 3 - 5 sont effectivement les meilleures en croisement et la lignée 6 la plus mauvaise.

Les lignées extrêmes pour l'aptitude générale à la combinaison correspondent bien aux phénotypes parentaux extrêmes (voir graphique 2).

AGC	LF	lf	LT	PT	NC	LC
Lignée maximale	6	5	3	5	3	2
Lignée minimale	1	3	5	3	2	1

2. Effets de consanguinité

Le mauvais accord d'ensemble entre les estimations de D dans un contexte de croisements et dans un contexte de lignées autofécondées n'a pas été expliqué par notre modèle génétique.

Chez le mil allogame, la consanguinité produit des plantes malades surtout pour les caractères liés à la valeur sélective : Poids de grain, hauteur de plante, chez lesquelles la variabilité génétique

.../...

s'exprime mal.

Cette observation demanderait confirmation mais rejoint déjà celles de GALLAIS (1977) sur une autre plante allogame, la luzerne.

Il faut sans doute se méfier dans ce cas des tests d'épistasie et des mesures de paramètres génétiques utilisant des lignées fortement autofécondées.

3. Nombre de facteurs génétiques

La présente analyse ne prétend pas appréhender les gènes particuliers agissant sur un caractère mais des groupes de gènes à ségrégation plus ou moins indépendante appelés facteurs génétiques (effective factor de MATHER, 1949).

Les estimations conduisent à un nombre de facteurs voisin du nombre haploïde de chromosomes : 7. Par ailleurs, les effets apparents de pleiotropie semblent très forts : beaucoup de distances sont fortement corrélées.

Tout se passe comme si chaque chromosome contenait un facteur génétique. De plus, ces segments doivent être assez longs pour que les pleiotropies soient fréquentes ce qui suppose en dernier ressort une fréquence assez faible des chiasmats à chaque cycle d'autofécondation.

Sinon comme TOWEY et JINKS (1977) l'ont observé sur *Nicotiana rustica* autogame, le nombre de facteurs serait bien plus élevé et les pleiotropies moins prononcées.

Curieusement, notre observation sur une plante allogame va dans le sens de la théorie de PANDEY (1972) suivant laquelle l'état allogame contribuerait à limiter le taux de recombinaison.

4. Influence de la densité sur les paramètres génétiques

Les conclusions de C. CHRISTOPHE (1973) sont confirmées :

La densité modifie les valeurs relatives des divers paramètres génétiques.

Pour les deux caractères qu'elle avait étudiés, nos résultats concordent avec les siens. La densité n'agit pas sur la valeur moyenne du cycle et de la hauteur mais augmente la part des effets non additifs.

Pour les autres caractères sauf PG, la densité augmente également la part des interactions.

.../...

Enfin, pour tous les caractères, le coefficient de variation erreur augmente avec la densité.

C. Valeur des lignées pour la sélection

L'observation des hétérosis moyens, de la dominance apparente et de la fraction non-additive de la variance génétique montre une diversité génétique légèrement plus forte dans le diallèle 6 x 6 que dans le diallèle 9 x 9 pour les caractères d'architecture.

Seuls PG et P500 montrent une diversité 6 x 6 beaucoup plus forte que la diversité 9 x 9. L'analyse du Poids ajusté à la végétation renforce cette conclusion.

Ces résultats tendent à prouver d'une part que la sélection des lignées pour les caractères d'architecture a été efficace et a restreint la variabilité génétique de ces caractères et d'autre part que ces caractères sont moins polygéniques que PG ou P500.

La forte diversité génétique restée disponible dans les lignées 6 x 6 d'origine plus variée que les lignées 9 x 9 à base génétique restreinte laisse espérer des hybrides à rendement/ha élevé.

Le rendement/ha du témoin L1 x L6 a pu être apprécié en essai agronomique au niveau de 3800 Kg/ha. Sachant que le Poids de grain du meilleur hybride 6 x 6 dépasse dans le diallèle PG L1 x L6 de 40%, il est permis d'espérer pour le meilleur hybride 6 x 6 un rendement de 5300 Kg/ha à une densité de 40.000 pieds/ha avec une irrigation d'appoint et plus de 100 Kg d'azote à l'ha en fumure.

REFERENCES BIBLIOGRAPHIQUES

- ARUNACHALAN V., OWEN A.R.G., 1971 - Polymorphisms with linked loci
Chapman and Hall - London.
- BAINES K.S., 1971 - Genetic analyses for certain plant and ear characters
in Pearl Millet top crosses ; T.A.G. 41 : 302-305.
- BALAKRISHNAN V., SANGHVI L.D., 1968 - Distance between populations on the
basis of attribute data ; BIOMETRICS 24 : 859-65.
- BILQUEZ A., LECOMTE J., 1969 - Relations entre mils sauvages et mils cul-
tivés : étude de l'hybride *Pennisetum typhoides* Stapf et Hubb.
x *Pennisetum violaceum* L., (Rich.) ; L'Agr., Trop., 24 : 249-57
- BRUNKEN J., 1977 - A systematic study of *Pennisetum* sect. *Pennisetum* (*gram-
minae*) ; Am., J., Bot., 64 : 161-76.
- BURTON G., POWELL J., 1968 - Pearl Millet breeding and cytogenetics Advan-
ces in Agronomy 20 : 49-89.
- CASAS E., HANSON W.D., WELLHAUSEN E.J., 1968 - Genetic relationships among
collections representing three Mexican race composites of *Zea*
Mays L. ; Genetics 59 : 299-310.
- CERVANTES S., GOODMAN M., CASAS E., RAWLINGS J., 1978 - Use of genetic
effects and genotype by environmental interactions for the
classification of Mexican races of maize ; Genetics 90 : 339-348
- CHRISTOPHE CATHERINE, 1973 - Action de la densité de plantation sur des
paramètres génétiques de la luzerne ; Ann., Amél., Plantes 23 :
67-76.
- COCKERMAN C.C., 1963 - Estimation of genetic variances ; Statistical gene-
tics and plant breeding ; Publication 982 ; NAS ; NRC ;
Washington D.C.
- COMSTOCK R., ROBINSON H., 1948 - The components of genetic variances in
populations of biparental progenics and their use in estimating
the average degree of dominance ; Biometrics 4 : 254-66.
- COMSTOCK R., ROBINSON H., 1952 - Estimation of average dominance of genes ;
Heterosis ; Iowa state college press., ;
- CRESS C.E., 1966 - Heterosis of the hybrid related to gene frequency dif-
ferences between two populations ; Genetics 53 : 269-74.
- DONALD C., HAMBLIN S., 1976 - The biological yield and harvest index of
cereals as agronomic and plant breeding criteria ; Advances
in Agronomy 28 : 361-404.

.../...

- EBERHAT S., GARDNER C., 1966 - A general model for genetic effects ;
Biometrics 22 : 864-81.
- EDWARDS A.W.F., 1971 - Distances between populations on the basis of gene
frequencies ; Biometrics 27 : 873-82.
- ELANDT-JOHNSON R., 1971 - Probability models and statistical methods in
genetics ; John Wiley & Sons ; New York ; 592 p.
- FERRARIS R., 1973 - Pearl Millet (*Pennisetum typhoides*) ; Review series
n°1/1973 ; C.A.B.
- FORKMANN G., SEYFFERT W., 1977 - Simulation of quantitative characters by
genes with biochemically definable action ; VI Modifications
of a simple model ; Genetics 85 : 557-72.
- GALLAIS A., 1970 - Modèle pour l'analyse des relations d'associations bi-
naires ; Biométrie Praximétrie 11 : 51-80.
- GALLAIS A., 1977 - Contribution à l'étude théorique et expérimentale de
l'hétérosis chez une plante allogame autotétraploïde ; Thèse de
Doctorat ; Faculté des Sciences ; Université Paris-Sud.
- GOWER J.C., 1971 - A general coefficient of similarity and some of its
properties ; Biometrics 27 : 857-71.
- GRIFFING B., 1956 - Concept of general and specific combining ability in
relation to diallel crossing systems ; Austr., J., Biol., Sc.,
9 : 463-93.
- HANSON W., CASAS E., 1968 - Spatial relationship among eight populations
of Zea Mays L. utilizing information from a diallel mating
design., ; Biometrics 24 : 867-80.
- HANSON W., MOLL R., 1973 - Experimental evaluation of relationships among
populations resulting from intergradation among cultivars of Zea
Mays ; Genetics 74 : 133-38.
- HARLAN J., 1971 - Agricultural origins : centers and non-centers ; Science
174 : 468-74.
- HAYMAN B., 1954 a - The analysis of variance of diallel tables ; Biometrics
10 : 235-44.
- HAYMAN B., 1954 b - The theory and analysis of diallel crosses ; Genetics
39 : 789-809.
- HEIDRICH-SOBRINHO E., CORDEIRO A., 1975 - Codominant isoenzymic alleles as
markers of genetic diversity correlated with heterosis in maize ;
T.A.G. 46 : 197-200.
- HUNTER R., KANNENBERG L., 1971 - Isozyme characterization of corn (Zea
mays) inbreds and its relationship to single cross hybrid perfor-
mance ; Can., J., Gen., and Cytol., XIII : 649-55.

.../...

- JINKS J., PERKINS J., BREES E., 1969 - A general method of detecting additive, dominance and epistatic variation for metrical traits ; II. Application to inbred lines ; *Heredity* 24 : 45-57.
- JINKS J., TOWEY P., 1976 - Estimating the number of genes in a polygenic system by genotype assay ; *Heredity* 37 : 69-82.
- KEARNEY P., 1975 - Linear models for least squares analysis of combination differential, and reflexive effects in simple diallels plans ; *Irish J., Agr., Res.*, 14 : 321-36
- KEARSEY M., JINKS J., 1968 - A general method of detecting additive, dominance and epistatic variation for metrical traits ; I. Theory *Heredity* 23 : 403-9.
- LAW C., SNAPE J., WORLAND A., 1978 - The genetical relationships between height and yield in wheat ; *Heredity* 40 : 133-151.
- LEBART L., FENELON J., 1971 - *Statistique et Informatique Appliquées* ; DUNOD ; Paris : 426 p.
- MATHER K., 1949 - *Biometrical genetics* (1 st edition) ; Methuen ; London.
- MATHER K., 1967 - Complementary and duplicate gene interactions in biometrical genetics ; *Heredity* 22 : 97-103.
- MATHER K., JINKS J., 1971 - *Biometrical genetics* ; CHAPMAN and HALL ; London ; 382 p.
- MOLL R., LONNQUIST J., VELEZ FORTUNO J., JOHNSON E., 1965 - The relationships of heterosis and genetic divergence in maize ; *Genetics* 52 : 139-44.
- NEI M., 1975 - *Molecular population genetics and evolution* ; NORTH-HOLLAND Amsterdam ; Oxford ; 288 p.
- PANDEY K., 1972 - Origin of genetic variation : regulation and genetic recombination in the higher organisms ; A theory ; *T.A.G.* 42 : 250-261.
- PORTERES R., 1951 - *Geographie alimentaire ; Berceaux agricoles et migrations des plantes cultivées en Afrique intertropicale* ; C.R., Soc., *Biogéographie* 239 : 16-21.
- PÓUZET D., 1971 - *Rapport d'activité ; Amélioration des Mils* ; ORSTOM ; IRAT ; BAMBEY ; Sénégal.
- RAO C., 1952 - *Advanced statistical methods in biometric research* ; John WILEY ; New York.
- SCHEFFE H., 1959 - *The analysis of variance* ; John WILEY ; New York ; 477 p.
- SOKAL R., SNEATH P., 1963 - *Principles of numerical taxonomy* ; Freeman ; San Francisco.

.../...

- THOMAS T., GRAFIUS J., HAHN S., 1971 - Genetic analysis of correlated sequential characters ; *Heredity* 26 : 177-88.
- TOWEY P., JINKS J., 1977 - Alternative ways of estimating the number of genes in a polygenic system by genotype assay ; *Heredity* 39 : 339-410.
- VAN DER VEEN J., 1959 - Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines ; *Genetica* 30 : 201-32.
- WEHRMAHN C., ALLARD R., 1965 - The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat ; *Genetics* 51 : 109-119.
- WRIGHT S., 1977 - *Evolution and the Genetics of Populations* ; vol. 3 ; Experimental results and evolutionary deductions ; The University of Chicago Press ; 613 p.
- ZALI A., ALLARD R., 1976 - The effect of level of heterozygosity on the performance of hybrids between isogenic lines of barley ; *Genetics* 84 : 765-75.

ANNEXE 1 : Analyse de variance du croisement diallèle

Soit un croisement diallèle comprenant p parents et les $p(p - 1)/2$ hybrides, cultivé en blocs randomisés complets de k blocs et r pieds par bloc.

$$y_{iikr} = v + 2 v_i + b_k + t_{iik} + \rho_{iikr}$$

$$y_{ijk r} = v + v_i + v_j + 1 + l_i + l_j + l_{ij} + b_k + t_{ijk} + \rho_{ijk r}$$

Les variances erreur des parents et des hybrides sont évaluées séparément. Le rapport $w = \sigma^2 F_1 / \sigma^2$ parents est utilisé pour une analyse de variance pondérée :

les différents effets sont estimés en minimisant

$$\sum_{i, k, r} W (y_{iikr} - v - 2 v_i - b_k - t_{iik})^2 + \sum_{i < j, k, r} (y_{ijk r} - v - v_i - v_j - 1 - l_i - l_j - l_{ij} - b_k - t_{ijk})^2$$

Dans les formules suivantes, le signe 0 désigne somme pour l'indice concerné. Le signe o/o signifie $\sum_{i < j}$. Il vient :

$$v = \sum_i y_{iioo} / pkr \quad 1 = 2 y_{o/o oo} / p(p - 1) kr - v$$

$$v_i = y_{iioo} / 2kr - \sum_i y_{iioo} / 2pkr$$

$$l_i = y_{i/o oo} / (p - 2) kr - 2 y_{o/o oo} / p(p - 2) kr - v_i$$

$$l_{ij} = y_{ijoo} / kr - (y_{i/ooo} + y_{j/ooo}) / (p - 2) kr + 2 y_{o/ooo} / (p - 1) (p - 2) kr$$

$$b_k = [2 w \sum_i (y_{iiko} - y_{iio}) + 2 (y_{o/oko} - y_{o/o})] / pr (p + 2 w - 1)$$

$$t_{ijk} = (y_{iiko} - y_{ijio}) / r - b_k$$

Le modèle général engendre les sommes de carrés :

$$SC (v + 1) = W (\sum_i y_{iioo})^2 / pkr + 2 y_{o/o oo}^2 / p (p - 1) kr$$

$$SC (v_i + l_i) = W [\sum_i y_{iioo}^2 - (\sum_i y_{iioo})^2 / p] / kr$$

$$+ [\sum_i y_{i/ooo}^2 - 4 y_{o/ooo}^2 / p] / (p - 2) kr$$

$$SC l_{ij} = \sum_{i < j} y_{ijoo}^2 / kr - \sum_i y_{i/ooo}^2 / (p - 2) kr + 2 y_{o/ooo}^2 / (p - 1) (p - 2) kr$$

$$SC_{blocs} = 2 \sum_k [W (\sum_i y_{iiko} - \sum_i y_{iioo} / k + y_{o/oko} - y_{o/ooo} / k)]^2 / rp (p + 2 w - 1)$$

$$SC T_{ijk} = W \sum_{i, k} y_{iiko}^2 / r + \sum_{i < j} y_{ijko}^2 / r - W \sum_i y_{iioo}^2 / kr$$

$$- \sum_{i < j} y_{ijoo}^2 / kr - SC_{blocs}$$

$$SCE = W \sum_{i, k, r} y_{iikr}^2 + \sum_{i < j, k, r} y_{ijk r}^2 - \text{sommes précédentes}$$

Le modèle restreint ($l = 0, l_i = 0$) engendre les sommes

$$SCVA = 2 (W \sum_i y_{iioo} + y_{o/o oo})^2 / pkr (p + 2 w - 1)$$

$$SCVIA = [\sum_i (2 w y_{iioo} + y_{i/ooo})^2 - (2 w \sum_i y_{iioo} + 2 y_{o/ooo})^2 / p] / kr (p + 4 w - 2)$$

Ces deux groupes de sommes de carrés permettent de calculer les sommes de carrés ajustés suivantes nécessaires aux tests des effets concernés :

$$SC\ l\ ajusté = SC\ (V + 1) - SCVA$$

$$SC\ li\ ajusté = SC\ (vi + li) - SCVIA$$

Enfin le modèle restreint ($vi = 0$) permet de la même manière de calculer :

$$SCVi\ ajustés : W[\sum_i y_{iio}^2 - (\sum_i y_{iio})^2/p]/kr$$

ANNEXE 2 : Calcul de l'espérance de $R^2(m, m')$

La distance est mesurée dans un croisement diallèle $r \times r$ entre lignées homozygotes tirées indépendamment les unes des autres d'une même population F2. On admet, en outre, que les différents gènes d'une même lignée résultent de tirages indépendants.

Dans ces conditions, pour deux gènes quelconques de deux lignées quelconques, on a les espérances : $E(\theta_{ai} \theta_{aj}) = 0$ et $E(\theta_{ai} \theta_{a'i}) = 0$ soit M et M' deux lignées, $R^2(m, m')$ leur distance. On veut calculer l'espérance de $R^2(m, m')$ dans l'ensemble des diallèles $r \times r$ pouvant être tirés de la F2, sachant que ces deux lignées ont une distribution donnée de locus différents et de locus identiques.

$$R^2(m, m') = \text{variance sur indice } f \text{ de } (ymf - ym'f)$$

$$\begin{aligned} ymf - ym'f &= \sum_a (\theta_{am} - \theta_{am'}) da/2 + \sum_a (\theta_{am'} - \theta_{am}) \theta_{af} ha/2 \\ &+ \sum_{a < a'} [\theta_{am} \theta_{a'm} - \theta_{am'} \theta_{a'm'} + (\theta_{am} - \theta_{am'}) \theta_{a'f} + (\theta_{a'm} - \theta_{a'm'}) \theta_{af}] iaa'/4 \\ &+ \sum_{a < a'} [\theta_{af} \theta_{a'f} (\theta_{a'm'} - \theta_{a'm}) - \theta_{am'} + \theta_{am} + \theta_{a'f} (\theta_{am'} \theta_{a'm'} - \theta_{am} \theta_{a'm})] ja/a'/4 \\ &+ \sum_{a < a'} [\theta_{af} \theta_{a'f} (\theta_{am'} - \theta_{am}) - \theta_{a'm'} + \theta_{a'm} + \theta_{af} (\theta_{a'm'} \theta_{am'} - \theta_{a'm} \theta_{am})] ja'/a \\ &+ \sum_{a < a'} [(\theta_{am'} - \theta_{am}) \theta_{af} + (\theta_{a'm'} - \theta_{a'm}) \theta_{a'f} + (\theta_{am} \theta_{a'm} - \theta_{am'} \theta_{a'm'}) \theta_{af} \theta_{a'f}] laa'/4 \end{aligned}$$

Les termes sans indice f sont des constantes qui n'interviennent pas dans le calcul de $R^2(m, m')$ on suppose $iaa' = i$, $ja/a' = ja'/a = j$

$$\sum_a \theta_{af} = \sum_a \theta_{am} = 0$$

les termes variables en i deviennent $i/4 \sum_a \theta_{af} \sum_{a' \neq a} (\theta_{a'm} - \theta_{a'm'}) = \sum_a \theta_{af} (\theta_{am'} - \theta_{am}) i/4$

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

$$\text{Posons } A = \sum_a \theta_{af} \frac{(\theta_{am'} - \theta_{am})}{2} \left(ha + \frac{i}{2} + \frac{N-1}{2} la \right)$$

avec $(N-1) la = \sum_a laa'$

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

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

$$\text{Variance (A)} = \sum_a \frac{(\theta_{am'} - \theta_{am})^2}{4} V_a \left(ha + \frac{i}{2} + \frac{N-1}{2} la \right)^2 + \sum_{a \neq a'} \frac{(\theta_{am'} - \theta_{am})}{2}$$

$$\frac{(\theta_{a'm'} - \theta_{a'm})}{2} Caa' \left(ha + \frac{i}{2} + \frac{N-1}{2} la \right) \left(ha' + \frac{i}{2} + \frac{N-1}{2} la' \right)$$

Pour $N - n$ locus, $\theta_{am'} - \theta_{am} = 0$, pour n locus $(\theta_{am'} - \theta_{am})/2 = \theta_{am}$
 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 l'espérance de ces termes parmi tous les diallèles $r \times r$ éventuels.

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, $E(V_a = 1 - \sum_{ij} \theta_{ai} \theta_{aj}/r^2)$
 $= V = \frac{r-1}{r}$

Posons $X_a = \frac{(\theta_{am'} - \theta_{am})^2}{4}$, il vient :

$$E[\text{Variance (A)}] = (V - C_2) \sum_a X_a \left(h_a + \frac{i}{2} + \frac{N-1}{2} l_a \right)^2 + C_2 \left[\sum_a X_a \left(h_a + \frac{i}{2} + \frac{N-1}{2} l_a \right) \right]^2 \quad (1)$$

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 $2 \theta_{am} \theta_{a'm} \theta_{af} \theta_{a'f} l_{aa'}/4$; donc $B = \sum_{a,a'} X_a (1 - X_{a'}) \theta_{am} \theta_{af} \theta_{a'm} \theta_{a'f} l_{aa'}/2$

La variance de B contient $n(N-n)$ termes de la forme variance $(\theta_{af} \theta_{a'f}) = 1 - \sum_{ij} \theta_{ai} \theta_{aj} \theta_{a'i} \theta_{a'j}/r^2$

$E(\theta_{ai} \theta_{aj} \theta_{a'i} \theta_{a'j}) = 1$ si $i = j$, elle est nulle autrement. Donc $E[\text{Variance } \theta_{af} \theta_{a'f}] = 1 - \frac{1}{r} = V$

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

Alors, $E[\text{Variance (B)}] = V \sum_{a,a'} X_a (1 - X_{a'}) \frac{l_{aa'}^2}{4} + C_2 \sum_{a,a',a'',a'''} X_a (1 - X_{a'}) X_{a''} (1 - X_{a'''}) \frac{l_{aa'}}{2} \frac{l_{a''a'''}}{2}$

$$E(\text{Variance B}) = (V - C_2) \sum_{a,a'} X_a (1 - X_{a'}) \frac{l_{aa'}^2}{4} + C_2 \left[\sum_{a,a'} X_a (1 - X_{a'}) l_{aa'}/2 \right]^2 \quad (2)$$

La Covariance (A,B) contient $n \times n(N-n)$ termes non nuls de la forme $(h_a + \frac{i}{2} + \frac{N-1}{2} l_a) \frac{l_{a'a''}}{2}$ 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 $-C_2$

$$E[\text{Covariance (A, B)}] = -C_2 \sum_{a,a',a''} X_a X_{a'} (1 - X_{a''}) \left(h_a + \frac{i}{2} + \frac{N-1}{2} l_a \right) l_{a'a''}/2 \quad (3)$$

Après assemblage des relations (1), (2) et (3), on trouve :

$$E [R^2 (m, m')] = (V - C_2) \sum_a X_a (ha + i/2 + \frac{N-1}{2} la)^2 +$$

$$\sum_{a, a'} (V - C_2) X_a (1 - X_{a'}) la^2 a'/4 + C_2 [\sum_a X_a (ha + \frac{i}{2} + \frac{N-1}{2} la)$$

$$- \sum_{a, a'} X_a (1 - X_{a'}) \frac{laa'}{2}]^2$$

mais $\sum_a X_a (ha + \frac{i}{2} + \frac{N-1}{2} la) - \sum_{a, a'} X_a (1 - X_{a'}) \frac{laa'}{2} = \sum_a X_a (ha + i/2)$

$$+ \sum_{a, a'} X_a X_{a'} laa'/2 = He (m, m')$$

c'est l'expression de l'hétérosis entre M et M'.

Donc $E [R^2 (m, m')] = (V - C_2) \sum_a X_a (ha + \frac{i}{2} + \frac{N-1}{2} la)^2 + (V - C_2) \sum_{a, a'} X_a$

$$(1 - X_{a'}) \frac{laa'}{4} + C_2 He^2 (m, m')$$



ANNEXE 3 : Divers calculs d'espérance

On utilise les relations $E(\theta_{am}) = E(\theta_{am} \theta_{a'm}) = E(\theta_{am} \theta_{a'm'}) = 0$

$$\theta_{am}^2 = 1$$

r = nombre de lignes.

E (Caa')

$$Caa' = \text{Covariance}(\theta_{am}, \theta_{a'm}) = \frac{1}{r} \sum_m \theta_{am} \theta_{a'm} - \sum_{ij} \theta_{ai} \theta_{a'j} / r^2$$

Tous les termes des deux sommes ont une espérance nulle

$$E(Caa') = 0$$

E (C²aa')

$$C^2_{aa'} = \frac{1}{r^2} \underbrace{\sum_{ij} \theta_{ai} \theta_{a'i} \theta_{aj} \theta_{a'j}}_A + \underbrace{\sum_{ij} \theta_{ai} \theta_{aj} \times \sum_{kl} \theta_{a'k} \theta_{a'l}}_{B} / r^4 - 2 \underbrace{\sum_{ijk} \theta_{ai} \theta_{a'j} \theta_{ak} \theta_{a'k}}_C / r^3$$

L'expression A contient r termes de valeur $1/r^2$ quant $i = j$.

quant $i \neq j$, les espérances sont nulles. Donc $E(A) = 1/r$

De même, B contient r^2 termes d'espérance $1/r^4$, les autres d'espérance zéro.

Donc $E(B) = 1/r^2$.

Enfin, C contient r termes ($i = j = k$) d'espérance $2/r^3$, les autres d'espérance zéro. Donc $E(C) = 2/r^2$.

$$\text{Finalement, } E(C^2_{aa'}) = \frac{1}{r} + \frac{1}{r^2} - \frac{2}{r^2} = \frac{r-1}{r^2}$$

E (Vn)

$$\begin{aligned} Vn &= \text{Variance de } \sum_a \left(\frac{\theta_{am} - \theta_{a'm'}}{4} \right)^2 \\ &= \sum_a \underbrace{\text{Variance de } (1 - \theta_{am} \theta_{a'm'})}_{A} / 2 \\ &+ \sum_{a \neq a'} \underbrace{\frac{1}{4} \text{Covariance} [(1 - \theta_{am} \theta_{a'm'}) (1 - \theta_{a'm} \theta_{a'm'})]}_B \end{aligned}$$

$$A = \sum a (2 Va - Va^2)/4$$

Le lecteur pourra démontrer que $E (Va) = V = \frac{r-1}{r}$

$$E (Va^2) = V_2 = (r-1) [r(r-1) + 2]/r^3$$

$$E (A) = N (2 V - V_2)/4$$

$$\begin{aligned} \text{Covariance } (\theta_{am} \theta_{am'}, \theta_{a'm} \theta_{a'm'}) &= \sum_{ij} \theta_{ai} \theta_{aj} \theta_{a'i} \theta_{a'j}/r^2 \\ &- \sum_{ij} \theta_{ai} \theta_{aj} \times \sum_{kl} \theta_{a'k} \theta_{a'l}/r^4 \end{aligned}$$

L'espérance de cette covariance vaut $\frac{1}{r} - \frac{1}{r^2} = C_2$ et $E (B) = N (N-1) C_2/4$

$$E (V_n) = \frac{N}{4} [2 V - V_2 + (N-1) C_2].$$

ANNEXE 4 : Calcul de l'espérance de la moyenne de He^2 (m, m')

$$He(m, m') = \sum_a X_a h_a + \sum_{aa'} X_a X_{a'} l_{aa'}/2$$

$$He^2 = \underbrace{(\sum_a X_a h_a)^2}_A + \underbrace{\sum_{aa'} X_a X_{a'} l_{aa'}^2/4}_B + \underbrace{\sum_{aa', a''a'''} X_a X_{a'} X_{a''} X_{a'''}}_C \frac{l_{aa'}}{2} \frac{l_{a''a'''}}{2} + 2 \underbrace{\sum_{a a' a''} X_a X_{a'} X_{a''} h_a \frac{l_{aa'}}{2}}_D$$

$$E(\bar{A}) = E(\sum_a h_a^2 \bar{X}_a) + E(\sum_{aa'} h_a h_{a'} \overline{X_a X_{a'}}) = \frac{V}{2} \sum_a h_a^2 + \sum_{aa'} \frac{V}{4} h_a h_{a'}$$

$$E(\bar{B}) = V/4 \sum_{aa'} l_{aa'}^2$$

$$E(\bar{C}) = \sum_{a a' a''} 2 \frac{V}{8} \frac{l_{aa'}}{2} \frac{l_{aa''}}{2} + \sum_{aa' a''} 2 \frac{V}{8} \frac{l_{aa'}}{2} \frac{l_{a'a''}}{2} + \sum_{aa' a'' a'''} \frac{V}{16} \frac{l_{aa'}}{2} \frac{l_{a''a'''}}{2}$$

$$E(\bar{D}) = 2 \sum_{aa'} \frac{2V}{4} h_a \frac{l_{aa'}}{2} + 2 \sum_{aa' a''} \frac{V}{8} h_a \frac{l_{a'a''}}{2}$$

En utilisant les relations suivantes :

$$\sum_{a'' \neq a, a'} l_{aa''} = \sum_{a''} l_{aa''} - l_{aa'} = l_{ao} - l_{aa'}$$

$$\sum_a l_{ao} = l_{oo}$$

$$\sum_{a \neq a' \neq a''} h_a l_{a'a''} = \sum_{a \neq a'} h_a (l_{a'o} - l_{aa'}) = \sum_a h_a (l_{oo} - l_{ao}) - \sum_a h_a l_{ao}$$

Il vient, après simplification :

$$E[\overline{He^2}] = \frac{V}{4} (h_o + \frac{l_{oo}}{4})^2 + \frac{V}{4} \sum_a (h_a + i/2 + \frac{l_{ao}}{2})^2 - \frac{V}{32} \sum_{aa'} l_{aa'}^2$$