

Étude de la diversité génétique du blé tendre. I. Comparaison de distances biochimiques, agromorphologiques et généalogiques

B Picard *, G Branlard, FX Oury, M Rousset

INRA, station d'amélioration des plantes, domaine de Crouelle, 63000 Clermont-Ferrand, France

(Reçu le 17 janvier 1992; accepté le 5 juin 1992)

Résumé — La diversité génétique existant entre des lignées de blé tendre a été estimée à partir de données biochimiques (électrophorèse des protéines de réserve du grain), de caractères agronomiques et morphologiques et à partir de l'étude de leurs généalogies. Ces données ont permis le calcul de distances biochimiques et agromorphologiques, à l'aide des formules de distances du χ^2 et absolue, et des coefficients de parenté selon Malecot. Il est apparu que chacune de ces distances donne une image différente de la diversité au niveau génétique. De plus, la prudence quant à l'utilisation des généalogies déclarées s'est avérée indispensable.

***Triticum aestivum* / distance génétique / généalogie / diversité / protéine de réserve**

Summary — Investigation on genetic diversity of bread wheat. I. Comparison of biochemical agromorphological and pedigree distances. Genetic diversity between wheat lines was estimated from biochemical data (storage, protein electrophoresis), measures of agronomic and morphological traits and the study of pedigrees. These data enabled the calculation of biochemical and agromorphological distances with the χ^2 and absolute formulas, and of coefficients of parentage according to Malecot. It appears that every distance gives a different representation of genetic diversity, as these distances are in no way correlated. Pedigrees have to be considered with caution, as glutenin electrophoregrams of some lines do not agree with those of their alleged parents.

***Triticum aestivum* / genetic distance / pedigree / diversity / wheat storage protein**

INTRODUCTION

L'estimation de la variabilité génétique est un problème fondamental en amélioration des plantes, à tous les niveaux d'un schéma de sélection (Lefort-Buson *et al*, 1988). L'évaluation de la diversité génotypique ou phénotypique peut être effectuée à partir de l'étude des généalogies des variétés considérées (Rodgers *et al*, 1983; Branlard et Chevalet, 1984; Lefort-Buson, 1986), à partir de caractères agromorphologiques (Behl, 1981; Behl *et al*, 1985; Camussi *et al*, 1983; Martinez *et al*, 1983; Branlard et Chevalet, 1984; Price *et al*, 1984; Cox *et al*, 1985a; Shamsuddin, 1985; Singh *et al*, 1985; Prasad et Singh, 1986) ou encore à partir de données moléculaires (Lookhart *et al*, 1983; Price *et al*, 1984; Cox *et al*, 1985a,b). Des techniques de

plus en plus performantes ont été développées ces dernières années pour l'étude de ces marqueurs moléculaires (électrophorèse monodimensionnelle de protéines diverses, séquençage de protéines ou d'acides nucléiques, électrophorèse bidimensionnelle de protéines dénaturées, analyse du polymorphisme de la longueur de fragments de restriction de l'ADN-RFLP).

Les descriptions de la variabilité obtenues avec ces différents outils ne sont pas forcément corrélées entre elles. Par exemple, chez le blé, Branlard et Chevalet (1984) et Cox *et al* (1985b) n'ont observé aucune relation entre le coefficient de parenté et un coefficient de similarité calculé à partir de l'électrophorèse des gliadines; alors que chez le soja, Cox *et al* (1985a) ont mis en évidence une forte corrélation entre le coefficient de parenté et un indice de similarité calculé à partir

de 11 systèmes enzymatiques. Chez le maïs, Damerval *et al* (1987) ont montré que des distances obtenues à partir d'électrophorèses bidimensionnelles sont différentes selon que l'on considère une distance qualitative (calculée à partir de la présence ou absence des protéines) ou quantitative (calculée à partir de l'intensité de ces mêmes protéines). Enfin, si plusieurs auteurs ont mis en évidence une relation positive entre distances génétiques et agromorphologiques, d'autres, en revanche, n'ont trouvé aucune relation entre ces deux distances (pour revue, voir Picard, 1990). Il convient d'être vigilant sur les relations mises en évidence entre marqueurs et diversité, et de distinguer les cas d'une liaison causale (marqueur d'incompatibilité, par exemple) et ceux d'une liaison génétique. Dans ce cas, on peut encore distinguer deux possibilités : l'apparement entre les individus et l'existence de linkats entre gènes.

Dans cette étude, réalisée dans le cadre de la collaboration entre l'INRA et le GIE Hybrible*, la diversité existant entre des lignées de blé tendre a été estimée par l'étude de leurs généalogies, par l'observation de caractères agromorphologiques, et par l'électrophorèse des protéines de réserve du grain. L'objet de notre travail étant la prédiction de l'hétérosis chez le blé tendre, nous avons utilisé ces 3 niveaux d'approche de la variabilité génétique pour calculer différentes distances dans le but de disposer d'indicateurs synthétiques de l'éloignement génétique existant entre les parents potentiels des hybrides. Dans cet article, nous comparons entre elles ces différentes distances, l'analyse des relations entre l'hétérosis et les distances interparentales faisant l'objet d'une seconde publication (Picard *et al*, 1992).

MATÉRIEL ET MÉTHODES

Matériel végétal

Une première expérimentation (exp 1) a porté sur 306 lignées de blé tendre, d'origines génétique et géographique très variées, qui ont été fournies par les membres du GIE Hybrible et les partenaires INRA de ce GIE. Ce matériel a été évalué sur plusieurs années et en plusieurs lieux, aussi bien au champ (observations en pépinière, essais pour l'étude du rende-

ment) qu'au laboratoire (étude de la qualité technologique).

Une deuxième expérimentation (exp 2) portant sur 42 lignées est venue compléter la première.

Évaluation des distances biochimiques et agromorphologiques

Données biochimiques

Les données biochimiques ont été obtenues par électrophorèse monodimensionnelle de protéines de réserve du grain de blé : les gliadines et les sous-unités gluténines de haut poids moléculaire (que l'on notera gluténines HPM). Le déterminisme génétique de ces 2 groupes de protéines met en jeu des locus situés sur le bras court des chromosomes 1A, 1B, 1D, 6A, 6B, 6D pour les gliadines, et sur le bras long des chromosomes 1A, 1B, 1D pour les gluténines HPM. L'étude des gluténines HPM a été effectuée à l'aide de la technique d'électrophorèse en gel de polyacrylamide en présence de sodium dodécyl sulphate (SDS-Page) de Laemmli (1970) modifiée par Payne *et al* (1979). La lecture des diagrammes a été réalisée à l'aide de la nomenclature établie par Payne (1983) et complétée par Branlard *et al* (1990b).

Les gliadines ont été analysées par la technique d'électrophorèse en gel de polyacrylamide en milieu acide (Acid-Page) de Bushuk et Zillman (1978), modifiée par Courvoisier (1984) puis par Branlard *et al* (1990a). La lecture des diagrammes a consisté en un relevé des mobilités des bandes gliadines et une notation de leur intensité (Branlard *et al*, 1990a).

Caractères agronomiques et morphologiques

Pour exp 1, 11 caractères avaient été observés en plusieurs lieux et sur plusieurs années, dans un but de connaissance du matériel végétal. Ces 11 caractères, pour lesquels nous disposons d'une information fiable, ont été retenus pour le calcul de la distance agromorphologique. Il s'agit de la hauteur, de la précocité, de l'alternativité, de la résistance au froid, du poids de 1 000 grains et de la sensibilité à différentes maladies (oïdium, rouille brune, rouille jaune, rouille noire, septoriose et piétin-verse).

Les 11 lieux d'expérimentation ayant fourni, en 2 années d'observation des 306 lignées, 23 notations pour l'oïdium et la rouille jaune, 22 pour la rouille brune, 17 pour la hauteur, 17 pour la précocité, 15 pour le poids de 1 000 grains, 7 pour la résistance au froid, 7 pour la septoriose et 3 pour l'alternativité, il a fallu synthétiser les données correspondant à ces caractères. Pour la hauteur, la précocité, la résistance au froid, l'alternativité, le poids de 1 000 grains et la

* Le GIE Hybrible regroupe les établissements de sélection Blondeau, Momont, Orsem et UCASP, et la firme Orsan (pour la recherche sur les gamétocides).

septoriose, cela a été fait par un simple calcul de moyenne. Pour les sensibilités à l'oïdium et aux rouilles brune et jaune, un premier regroupement a été réalisé au moyen d'analyses en composantes principales, les individus étant les 306 génotypes, et les variables les différentes notations disponibles pour une même maladie. Les représentations des variables dans les plans factoriels définis par les 3 premiers axes principaux, et l'étude des corrélations entre ces variables et les composantes principales ont permis de distinguer 7 groupes de notations pour l'oïdium, et 6 groupes de notations pour les rouilles brune et jaune. Nous avons admis que, pour chacune de ces maladies, les différents groupes correspondaient à des associations particulières de gènes de résistance. La moyenne des notations correspondant à un même groupe nous a permis d'attribuer à chacun des individus la valeur 1 s'il possédait l'association de gènes de résistance considérée, et la valeur 0 s'il ne la possédait pas. Ce sont ces 19 variables (7 pour l'oïdium, 6 pour la rouille jaune et 6 pour la rouille brune) qui ont été finalement prises en compte, avec celles correspondant aux 8 autres caractères, pour le calcul des distances agromorphologiques.

Pour exp 2, les 42 lignées n'ayant pas fait l'objet d'observations aussi poussées que pour exp 1, il n'y a pas eu de calcul de distance agromorphologique.

Calcul de distances

Deux formules différentes ont été utilisées : la distance du χ^2 et la distance absolue. Ces deux formules ont été retenues pour leur facilité de calcul. La distance du χ^2 , contrairement à la distance absolue, a cependant l'inconvénient d'être biaisée, car il y a dans son cas une association entre moyenne et variance.

Le principe utilisé dans le calcul de la distance du χ^2 est celui d'une classification hiérarchique ascendante (Benzecri, 1973). Si l'on considère une matrice $C(n, p)$ à p colonnes correspondant à la présence ou l'absence de p bandes (ou allèles) ou à leur intensité relative, et n lignes pour les n individus étudiés, la distance du χ^2 entre 2 génotypes i et i' est donnée par la formule suivante :

$$d^2(i, i') = \sum_{j=1}^{j=p} 1/C.j ((C_{ij}/C_i) - (C_{i'j}/C_{i'}))^2 \quad [1]$$

C_i représente la moyenne marginale correspondant à l'individu i , et $C.j$ la moyenne marginale correspondant à la bande j . C'est la racine carrée de [1] qui fournit la valeur de la distance.

La distance absolue est donnée par la formule :

$$d(i, i') = \sum_{j=1}^{j=p} |C_{ij} - C_{i'j}|/2 \quad [2]$$

Les formules [1] et [2] ont servi au calcul des $n(n-1)/2$ distances caractérisant les $n(n-1)/2$ couples que l'on peut constituer à partir de n individus.

L'étude des gluténines HPM a conduit au calcul :

- d'une distance qualitative, où C_{ij} prend la valeur 1 (présence) ou 0 (absence) pour chacune des différentes bandes de gluténines HPM observées.
- d'une distance génétique, où C_{ij} prend les valeurs 1 ou 0 lorsqu'un allèle des gluténines HPM est présent ou absent. L'ensemble des allèles des gluténines HPM a été catalogué et chacun d'eux est formé soit d'une bande (allèle nul) soit de 1 ou 2 bandes (Payne *et al*, 1984).

L'étude des gliadines a également conduit au calcul de 2 distances :

- une distance qualitative, où C_{ij} prend la valeur 0 (absence) ou 1 (présence) pour chacune des différentes bandes gliadines observées;
- une distance quantitative, où C_{ij} prend la valeur correspondant à l'intensité de chaque bande. L'intensité des bandes a été notée visuellement de 0 à 4.

Pour exp 2, les distances correspondant aux gliadines ont été calculées dans un premier temps sur l'ensemble des gliadines, puis sur les ω -gliadines codées par le bras court des chromosomes 1A, 1B et 1D.

Pour la distance agromorphologique se posait le problème de la pondération des différents caractères pris en compte, ceux-ci étant exprimés dans des unités très diverses. Pour la distance absolue, nous avons résolu ce problème en effectuant le calcul sur les données centrées-réduites. L'utilisation du centrage-réduction n'étant pas possible dans le cas de la distance du χ^2 (le terme $1/C.j$ correspondrait à une division par zéro), nous avons, pour cette dernière, transformé nos variables quantitatives en variables qualitatives ayant toutes un nombre équivalent de classes.

Travaux sur les coefficients de parenté

Étude des généalogies

Pour chacune des variétés étudiées, nous avons établi l'arbre généalogique le plus complet possible. Cela a conduit à la création d'un fichier informatique de plus de 2 800 individus qui a servi de base pour le calcul du coefficient de parenté.

Calcul du coefficient de parenté

Le coefficient de parenté (Phi) a été calculé à partir de la formule suivante (Malécot, 1948) :

$$\text{Phi}(i, j) = \sum_{k=1}^{k=1} (1 + F_k) (1/2)n_k + p_k + 1$$

où l est le nombre d'ancêtres communs aux lignées i et j , le k^e ancêtre commun ayant un coefficient de consanguinité F_k et étant séparé de i par n_k générations et de j par p_k générations.

Comparaison des distances entre elles et avec le coefficient de parenté

La comparaison entre les différentes distances a été réalisée par calcul du coefficient de corrélation de rangs de Spearman. En effet, certaines distances présentant une distribution très éloignée d'une distribution Gaussienne, nous avons préféré utiliser une statistique non paramétrique.

Les différents traitements statistiques ont été réalisés à l'aide des logiciels Consistent System (avec son gestionnaire de bases de données Janus), SPAD et S (pour la partie graphique). Pour le calcul des différentes distances (et particulièrement le coefficient de parenté), plusieurs programmes ont été écrits en Fortran.

RÉSULTATS ET DISCUSSION

Description de la diversité mesurée à partir des données biochimiques

Diversité des gluténines HPM

Parmi les 306 lignées de exp 1, 52 diagrammes différents contenant 18 des 24 bandes recensées ont été identifiés. Quinze diagrammes ne se retrouvent que chez une seule variété, et 12 diagrammes sont présents chez 2 ou 3 variétés seulement. La population étudiée présente donc une diversité importante pour ces protéines.

De la même façon, parmi les 42 lignées de exp 2, 21 diagrammes différents ont été rencontrés, 9 ne se trouvant que chez une seule variété.

Le déterminisme génétique des gluténines HPM étant connu précisément, nous avons pu calculer la fréquence des allèles gluténines correspondant aux bandes observées (tableau I). Les diagrammes les plus fréquents dans nos 2 échantillons (2-7-12, 2-6-8-12, 2-7-8-12, 2-7-9-12) caractérisent assez bien les blés français. Si l'on effectue l'analyse d'une collection mondiale, on constate que ces diagrammes se retrouvent moins fréquemment parmi les génotypes d'origine étrangère (Branlard *et al*, 1990b).

Diversité des gliadines

L'étude de 306 variétés de exp 1 a permis de révéler 226 bandes de mobilités différentes, dont 96 bandes Ω -gliadines, 103 bandes γ - β et 27 bandes α .

Tableau I. Effectif des allèles gluténines HPM de expl (1) et de expll (2).

Chromosome	Allèles	(1)	(2)
1A	1	56	26
	2*	55	6
	nul	195	10
1B	7	79	10
	7-8	80	12
	7-9	75	10
	6-8	46	8
	20	4	1
	13-16	3	1
	17-18	15	0
21	4	0	
1D	2-12	144	21
	3-12	18	1
	4-12	24	1
	5-10	119	19

Parmi les 42 lignées de expll, 182 bandes ont été observées, 84 ω , 71 γ - β et 27 α .

Distributions des différentes distances

Les figures 1 à 3 représentent les distributions des différentes distances calculées pour exp 1 et exp 2.

D'une manière générale, les distances du χ^2 présentent des distributions plus dissymétriques que les distances absolues. Cela s'explique par le fait que la formule de χ^2 introduit des pondérations par les moyennes marginales. Ces pondérations confèrent une importance relative plus grande aux individus rares, d'où une assymétrie droite plus marquée qu'avec la distance absolue.

La dissymétrie est moins importante dans le cas des gliadines (fig 1 et 3), en raison de la prise en compte d'un nombre de bandes beaucoup plus élevé. Par ailleurs, les distributions correspondant aux gliadines présentent un indice d'aplatissement plus élevé que celui d'une distribution Gaussienne, ce qui est dû au fait que beaucoup de génotypes ne diffèrent que par un faible nombre de bandes.

La répartition des coefficients de parenté apparaît sur les figures 2 et 3. Seize lignées seulement, parmi les 306 de exp 1, ont un coefficient

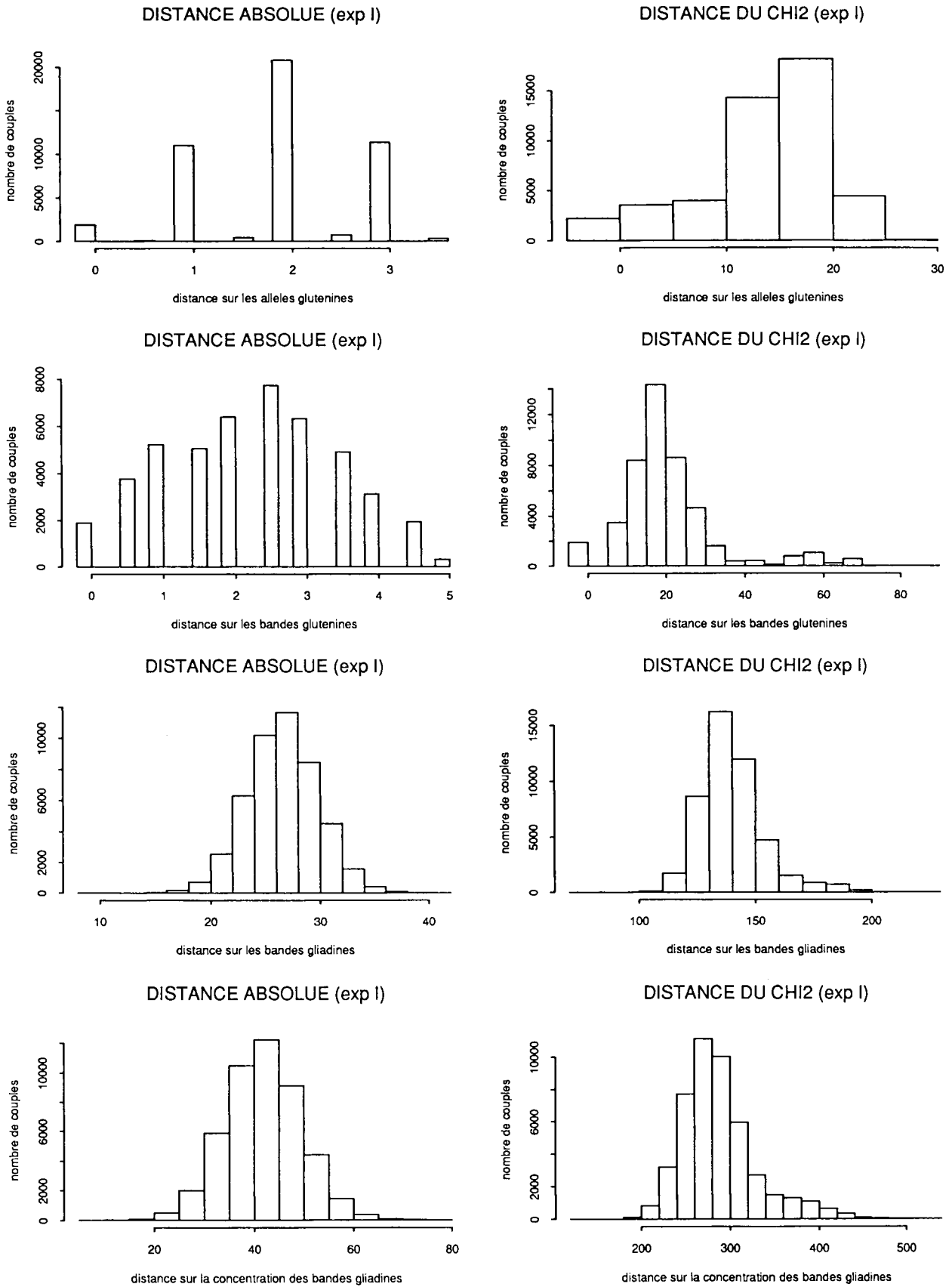


Fig 1. Distributions des distances biochimiques calculées pour les 46 665 couples de exp 1.

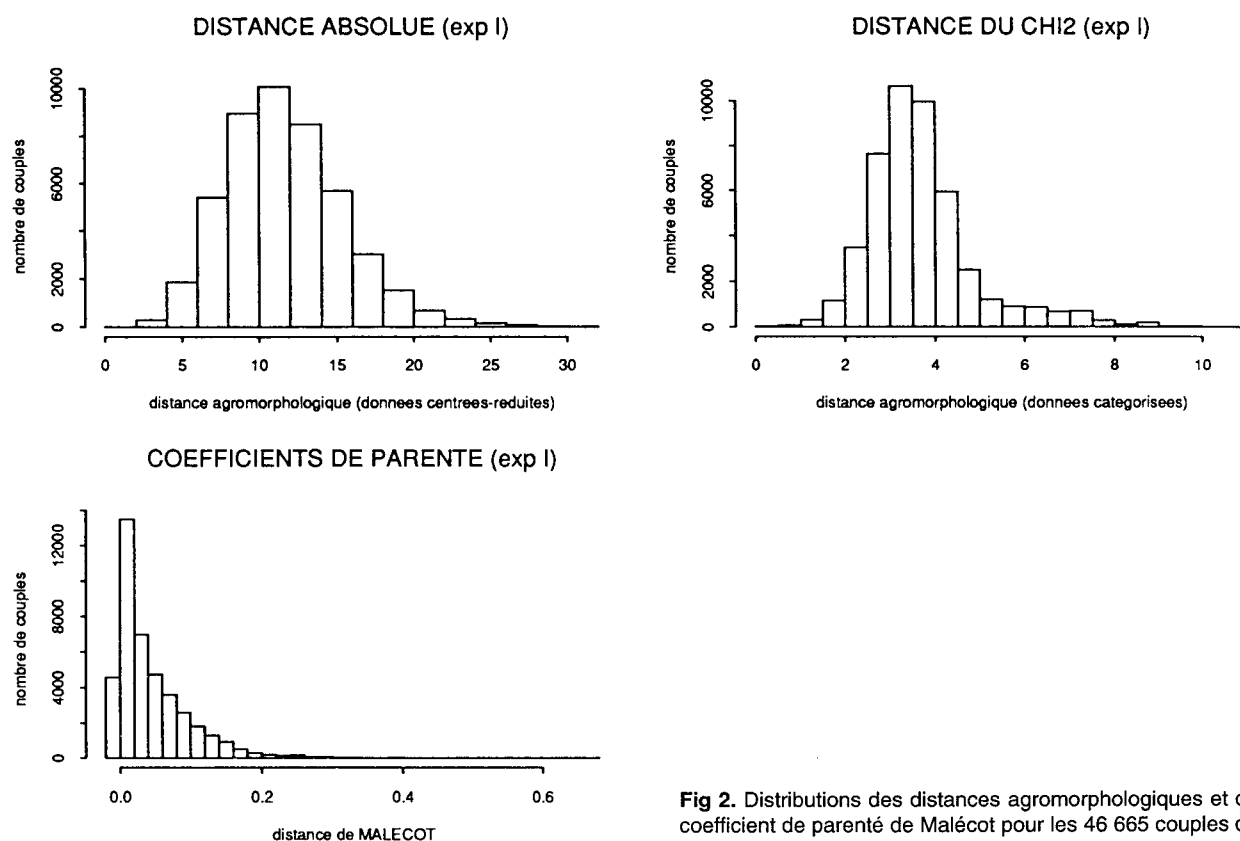


Fig 2. Distributions des distances agromorphologiques et du coefficient de parenté de Malécot pour les 46 665 couples de exp 1.

de parenté supérieur à 0,5. Cela reflète bien l'origine très variée de ces lignées.

Comparaison des distances

Dans les tableaux II et III sont donnés les résultats des différentes études de corrélation effectuées entre matrices de distances.

Compte tenu du nombre très élevé de degrés de liberté (46 664 pour exp 1; 860 pour exp 2), les seuils de signification indiqués dans les légendes des tableaux II et III correspondent à des liaisons extrêmement faibles. Nous discutons donc les résultats non pas en fonction du caractère significatif de la corrélation, mais en fonction de la part de variance expliquée par la relation. Nous considérons qu'une liaison n'est d'aucune utilité lorsque le coefficient de détermination est inférieur à 20%.

Distances du χ^2 et absolue

Les distances du χ^2 et absolue sont bien corrélées dans le cas des caractères agromorphologiques ($\rho = 0,774$ pour exp 1). Elles sont assez

fortement corrélées dans le cas des gluténines HPM (pour la distance génétique, $\rho = 0,696$ pour exp 1 et $\rho = 0,817$ pour exp 2; pour la distance qualitative, $\rho = 0,700$ pour exp 1). La liaison est moins forte dans le cas des gliadines (pour la distance quantitative, $\rho = 0,486$ pour exp 1 et $\rho = 0,614$ pour exp 2), et devient même très faible pour la distance qualitative de exp 1 ($\rho = 0,054$).

La plus grande indépendance entre les 2 distances dans le cas des gliadines est à relier au nombre de variables considérées (226 bandes pour les gliadines, contre 18 bandes pour les gluténines et 11 caractères agromorphologiques) : la distance du χ^2 ne prenant pas en compte les individus rares de la même manière que la distance absolue, la distorsion des résultats fournis par les 2 formules de distance est plus importante lorsque les phénomènes de rareté peuvent se cumuler sur 226 variables, que lorsqu'ils ne peuvent concerner que 18 ou 11 variables.

Le niveau des corrélations observées indique que les distances calculées avec les 2 formules ne sont pas totalement redondantes. D'un point de vue théorique, la distance absolue présente l'avantage de ne pas introduire de biais statistique, car il y a dans son cas indépendance entre moyenne et variance.

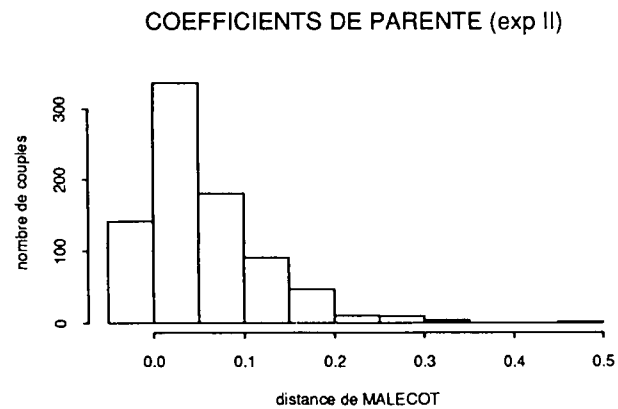
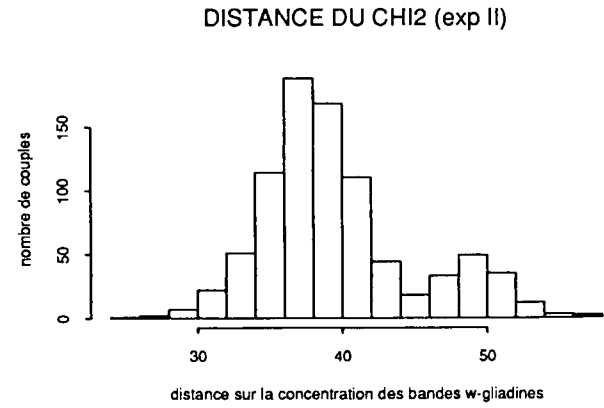
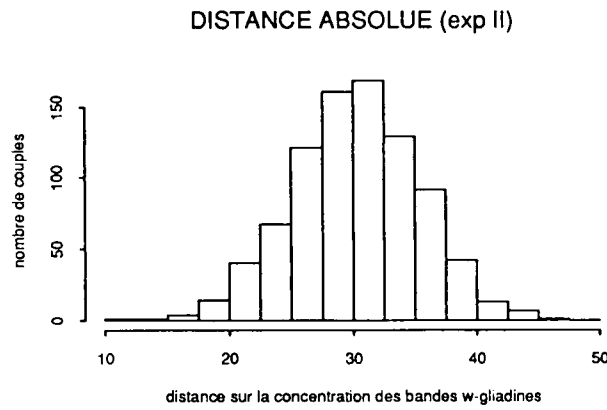
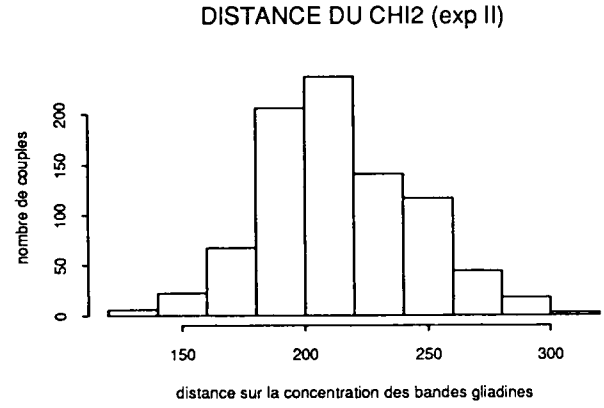
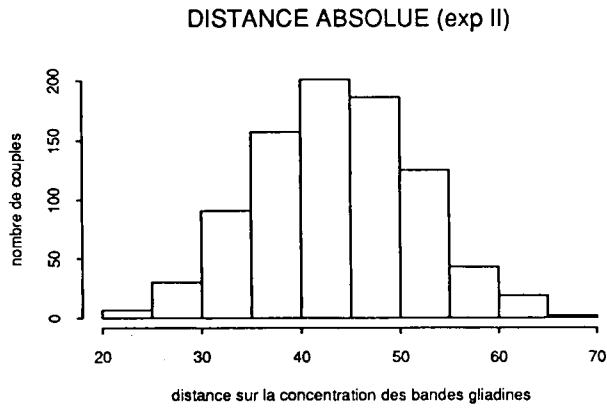
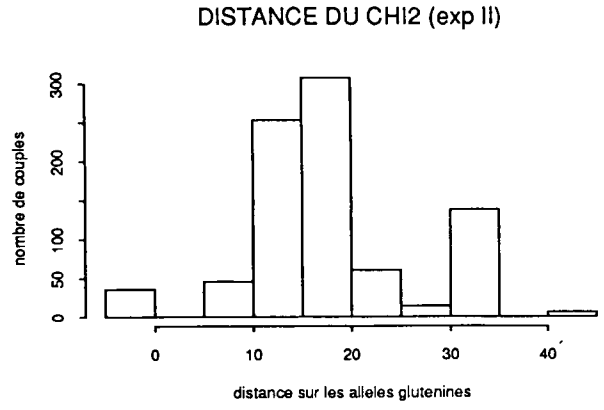
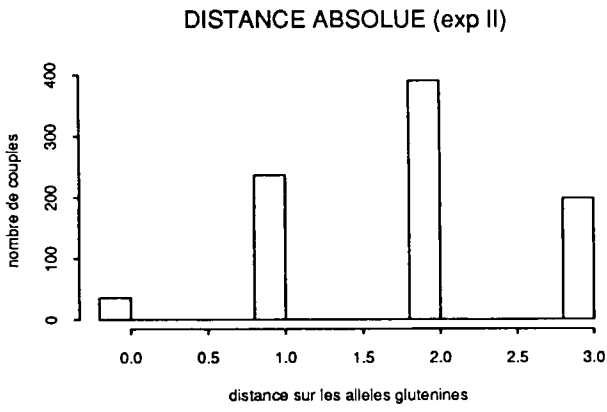


Fig 3. Distributions des différentes distances calculées pour les 861 couples de exp 2.

Tableau II. Coefficients de corrélation de rangs de Spearman entre les différentes distances de exp 1 (pour 46 664 ddl, les valeurs critiques sont de 0,008 au seuil de 5%, 0,011 au seuil de 1% et 0,014 au seuil de 1‰).

	<i>aglug</i>	<i>cglug</i>	<i>aglub</i>	<i>cglub</i>	<i>aglip</i>	<i>cglip</i>	<i>aglic</i>	<i>cglic</i>	<i>amor</i>	<i>cmor</i>	<i>copar</i>
<i>aglug</i>	1,000	0,696	0,796	0,592	0,042	0,090	0,055	0,095	0,094	0,099	-0,015
<i>cglug</i>	0,696	1,000	0,425	0,170	-0,008	0,068	0,036	0,092	0,045	0,029	-0,111
<i>aglub</i>	0,796	0,425	1,000	0,700	-0,002	0,023	0,005	0,029	0,032	0,033	-0,101
<i>cglub</i>	0,592	0,170	0,700	1,000	-0,700	-0,074	-0,019	-0,051	0,006	0,030	-0,089
<i>aglip</i>	0,042	-0,008	-0,002	-0,007	1,000	-0,054	0,505	-0,131	0,031	-0,000	-0,025
<i>cglip</i>	0,090	0,068	0,023	-0,074	-0,054	1,000	0,074	0,721	0,040	0,068	-0,039
<i>aglic</i>	0,055	0,036	0,005	-0,019	0,505	0,074	1,000	0,486	-0,001	-0,020	-0,067
<i>cglic</i>	0,095	0,092	0,029	-0,051	-0,131	0,721	0,486	1,000	0,004	0,033	-0,071
<i>amor</i>	0,094	0,045	0,032	0,006	0,031	0,040	-0,001	0,004	1,000	0,774	-0,071
<i>cmor</i>	0,099	0,029	0,033	0,030	-0,000	0,068	-0,020	0,033	0,774	1,000	-0,056
<i>copar</i>	-0,015	-0,111	-0,101	-0,089	-0,025	-0,039	-0,067	-0,071	-0,071	-0,056	1,000

Aglug et *cglug* : distances absolue et du χ^2 calculées à partir des allèles gluténines HPM; *aglub* et *cglub* : distances absolue et du χ^2 calculées à partir des bandes gluténines HPM; *aglip* et *cglip* : distances absolue et du χ^2 calculées à partir des bandes gliadines; *aglic* et *cglic* : distances absolue et du χ^2 calculées à partir de l'intensité des bandes gliadines; *amor* et *cmor* : distances absolue et du χ^2 calculées à partir des données agromorphologiques; *copar* : distance correspondant au coefficient de parenté de Malécot.

Tableau III. Coefficients de corrélation de rangs de Spearman entre les différentes distances de exp 2 (pour 860 ddl, les valeurs critiques sont de 0,056 au seuil de 5%, 0,079 au seuil de 1% et 0,105 au seuil de 1‰).

	<i>aglug</i>	<i>cglug</i>	<i>aglic</i>	<i>cglic</i>	<i>awglic</i>	<i>cwglic</i>	<i>copar</i>
<i>aglug</i>	1,000	0,817	0,093	0,103	0,040	0,017	0,037
<i>cglug</i>	0,817	1,000	0,086	0,079	0,011	-0,047	0,001
<i>aglic</i>	0,093	0,086	1,000	0,614	0,798	0,292	-0,005
<i>cglic</i>	0,103	0,079	0,614	1,000	0,463	0,547	0,035
<i>awglic</i>	0,040	0,011	0,798	0,463	1,000	0,443	-0,006
<i>cwglic</i>	0,017	-0,047	0,292	0,547	0,443	1,000	0,133
<i>copar</i>	0,037	0,001	-0,005	0,035	-0,006	0,133	1,000

Aglug et *cglug* : distances absolue et du χ^2 calculées à partir des allèles gluténines HPM; *aglic* et *cglic* : distances absolue et du χ^2 calculées à partir de l'intensité des bandes gliadines; *awglic* et *cwglic* : distances absolue et du χ^2 calculées à partir de l'intensité des bandes ω -gliadines; *copar* : distance correspondant au coefficient de parenté de Malécot.

Distances gluténines qualitative et génétique

Les distances gluténines qualitative et génétique sont assez fortement corrélées dans le cas des distances absolues ($\rho = 0,796$ pour exp 1). Par contre, la corrélation devient très faible dans le cas des distances du χ^2 ($\rho = 0,170$ pour exp 1).

Cette différence peut être expliquée par le déterminisme génétique des bandes gluténines : la plupart des allèles codent pour 2 bandes, ce qui confère aux individus rares une importance relative plus grande lorsqu'on considère la pré-

sence-absence des bandes que lorsqu'on considère les allèles.

Distances gliadines qualitative et quantitative

Les distances gliadines qualitative et quantitative sont assez fortement corrélées (pour exp 1, $\rho = 0,505$ pour la distance absolue et $\rho = 0,721$ pour la distance du χ^2).

Cette liaison positive peut s'expliquer par le fait que la distance quantitative tient compte à la fois de la présence de la bande et de son intensité.

Nos résultats confirment ceux obtenus par Seguin (1990) qui, par électrophorèse bidimensionnelle (E2D) de protéines foliaires de coléoptiles de blé, constate que variations quantitatives et qualitatives donnent la même mesure de la divergence génétique. Par contre chez le maïs, Damerval *et al* (1987) sur 5 lignées à un seul stade organe, puis Léonardi (1989) sur 8 lignées (dont 4 communes aux 5 précédentes) à plusieurs stades organes, n'ont trouvé aucune corrélation entre distances qualitative et quantitative obtenues par E2D de protéines foliaires. La variation génétique des régulateurs (révélée spécifiquement par les variations quantitatives en E2D) serait non neutre sélectivement, contrairement à la variation des gènes de structure (Klose, 1982; De Vienne et Damerval, 1985), ce qui expliquerait l'absence de corrélation entre ces deux distances chez le maïs.

Les résultats différents obtenus pour le blé et le maïs peuvent être expliqués par le fait que les variations quantitatives des protéines chez ces 2 espèces ont des origines différentes. Dans le cas du maïs, la variation quantitative des protéines reflète le polymorphisme des gènes de régulation qui contrôlent l'expression des gènes de structure. Pour le blé tendre, l'hexaploïdie complique l'interprétation des variations quantitatives (Zivy *et al*, 1984; Colas des Francs et Thiellement, 1985). En effet, les 3 génomes s'expriment simultanément chez l'hexaploïde et les phénomènes de répression intergénomique sont très rares (Gallili et Feldman, 1984; Bahrman et Thiellement, 1987). Des différences d'intensité peuvent alors refléter des différences de dosage génique des gènes de structure et non une variation des gènes de régulation.

Distances gluténines HPM et gliadines

Les liaisons entre les distances biochimiques – gliadines et gluténines HPM – sont extrêmement faibles (les valeurs du coefficient de corrélation de rangs de Spearman sont comprises entre – 0,074 et 0,095 pour exp 1, et 0,079 et 0,103 pour exp 2). De plus, pour exp 2, l'étude des corrélations entre les distances correspondant aux gluténines et des distances calculées à partir des oméga-gliadines seulement, donne un résultat identique ($\rho = 0,040$ pour la distance absolue et $\rho = -0,047$ pour la distance du χ^2).

Cette absence de relation n'est pas surprenante, compte tenu de l'absence de liaison génétique entre les gènes «gluténines HPM», situés sur le bras long des chromosomes du groupe 1,

et les gènes «gliadines» (ceux codant pour les gliadines α , β et γ sont situés sur le bras court des chromosomes du groupe 6; et ceux codant pour les Ω -gliadines sont situés sur le bras court des chromosomes du groupe 1, mais à une distance d'environ 50 centimorgans des gènes «gluténines»).

Distances biochimiques et agromorphologiques

La liaison entre distances agromorphologiques et distances biochimiques apparaît également extrêmement faible (les valeurs du coefficient de corrélation de rangs de Spearman sont comprises entre –0,020 et 0,099 pour exp 1).

Les résultats bibliographiques sont en fait variables selon les espèces et ne semblent pas suivre de loi générale. Chez le blé, Branlard et Chevalet (1984) n'ont trouvé aucune relation entre distance biochimique (obtenue par électrophorèse des gliadines) et distance agromorphologique. De même, Seguin (1990) n'a pas observé de corrélation entre distances biochimiques (obtenues par E2D de protéines foliaires) et distances agromorphologiques. Par contre chez le maïs, Damerval *et al* (1987) ont observé une corrélation significative entre une distance agromorphologique et une distance biochimique quantitative (intensité des protéines) obtenues par E2D de protéines foliaires. Cependant, cette corrélation disparaissait lorsqu'ils considéraient la même distance agromorphologique et une distance biochimique qualitative (présence-absence des protéines). Ces résultats ont été confirmés par Léonardi (1989).

La relation qui peut exister entre caractères biochimiques et phénotypiques, n'est pas, dans la majorité des cas, une relation directe, mais une relation relevant plutôt de phénomènes de liaison génétique. Il s'ensuit une instabilité de ce type de relation, car la recombinaison peut provoquer un changement des liaisons mises en évidence à un moment donné entre les marqueurs protéiques et les caractères morphologiques.

Cependant, la principale explication de l'absence de corrélation entre ces 2 types de distances tient au fait que les marqueurs protéiques ne concernent que 9 locus répartis sur 6 paires chromosomiques, tandis que les caractères agromorphologiques sont déterminés par des gènes répartis sur l'ensemble du génome (à ce niveau, on peut remarquer que le calcul d'une

distance agromorphologique pose toujours le problème du choix du nombre et de la nature des caractères à prendre en compte (Lefort-Buson *et al*, 1988). De plus, les caractères agromorphologiques sont peu héréditaires car leur déterminisme est souvent polygénique et leur expression fortement influencée par le milieu, ce qui rend leur utilisation pour un calcul de distance plus délicate que celle des marqueurs protéiques.

Distances biochimiques, agromorphologiques et coefficient de parenté

Comme le montrent les résultats des tableaux II et III, les liaisons entre distances biochimiques et coefficient de parenté sont très faibles (les coefficients de corrélation de rangs de Spearman sont compris entre $-0,015$ et $-0,111$ pour exp 1, et $0,133$ et $-0,006$ pour exp 2).

Il en est de même pour les liaisons entre distances agromorphologiques et coefficient de parenté (pour exp 1, $\rho = -0,071$ pour la distance absolue et $\rho = -0,056$ pour la distance du χ^2).

On peut noter cependant que, pour expl, tous les signes des coefficients sont négatifs : globalement, pour un échantillon de taille importante, on a donc bien une augmentation des distances biochimiques et agromorphologiques lorsque l'apparement diminue.

Les résultats bibliographiques concernant ces relations sont également variables selon les espèces considérées (Branlard et Chevalet, 1984; Cox *et al*, 1985a,b). Les marqueurs utilisés ici ont tous été choisis en fonction de critères particuliers : les protéines de réserve ne correspondent qu'à un nombre très restreint de locus, mais dont la complexité est à l'origine d'un polymorphisme important; les caractères agronomiques et morphologiques sont sous la dépendance de locus plus répartis sur les chromosomes, mais qui ne représentent cependant qu'une partie du génome et sont soumis à une forte influence du milieu. Quant aux généalogies, elles ont été utilisées pour calculer des coefficients de parenté qui mesurent la probabilité d'identité allélique à un locus donné. Cependant, le calcul de l'apparement entre individus à l'aide du coefficient de parenté de Malécot nécessite des hypothèses de fécondation aléatoire et d'absence de sélection qui sont respectées dans le domaine de la génétique animale, mais non dans un schéma d'amélioration des plantes.

Par ailleurs, les généalogies utilisées ne sont pas toujours connues de façon certaine. En effet, un contrôle des généalogies effectué sur 29 des 306 variétés de expl, a montré qu'au moins 7 d'entre elles étaient fausses. Ce contrôle a consisté en une comparaison des diagrammes gluténines HPM des individus avec ceux de leurs parents ou grand-parents. La même étude faite sur l'ensemble des blés français inscrits au catalogue officiel des variétés et analysés au laboratoire, a montré que 32% des diagrammes observés n'étaient pas compatibles avec les généalogies déclarées (Picard, 1990).

Quelques explications peuvent être avancées pour expliquer ces discordances. D'une part, il peut y avoir eu des problèmes d'alofécondation ou des erreurs de manipulation (interversion ou mélange d'échantillons) lors des différentes phases de multiplication, ce qui a pu conduire à une modification de la variété au cours du temps. D'autre part, il peut y avoir eu une erreur au niveau du diagramme gluténine attribué aux lignées ancêtres, qui pouvait ne pas être complètement fixé au moment de leur déclaration (les multiplications ultérieures auraient pu alors conduire à l'élimination du diagramme déclaré).

CONCLUSION

La diversité génétique mesurée au sein d'échantillons de lignées résulte de l'origine génétique et géographique du matériel étudié. La quantification de cette diversité peut être nécessaire, notamment pour la prédiction de la valeur hybride (sur cet aspect, voir la suite de cette étude : Picard *et al*, 1992). Cette quantification peut être faite en utilisant des marqueurs du génome qui, pour être intéressants, doivent correspondre à un nombre de locus important et présenter un large polymorphisme. Les 2 échantillons que nous avons étudiés présentent une bonne diversité génétique. Cependant, les différentes distances calculées ne sont pas corrélées. Ces distances mettent en jeu un nombre de locus important, qui sont indépendants. La diversité existant à un locus n'est pas corrélée à celle existant à un autre des locus analysés ici.

Si ces distances donnent des mesures différentes de la diversité génétique, elle peuvent toutefois s'avérer complémentaires. Il pourrait alors être intéressant de créer un indice de distance les regroupant, ce qui permettrait un marquage d'une plus large portion du génome. Dans ce cas

se poserait le problème du poids à donner à chacune de ces distances.

Ce travail nous a également permis de souligner le large polymorphisme des protéines de réserve (gluténines HPM et gliadines). L'extension à une autre famille de protéines, les gluténines de faible poids moléculaire, et l'utilisation d'une séparation par électrophorèse bidimensionnelle (E2D) (Picard, 1990) permettraient une étude encore plus fine de ce polymorphisme. L'étude des gels obtenus en E2D par un analyseur d'images, offrirait la possibilité de calculer une distance biochimique quantitative, par une comparaison automatisée de l'intensité des protéines.

Enfin, nous avons vu une autre utilisation possible des marqueurs biochimiques : ils permettent de juger de la variabilité de telle ou telle généalogie.

RÉFÉRENCES

- Bahrman N, Thiellement H (1987) Parental genome expression in synthetic wheats (*Triticum turgidum* sp, *T tauschii* sp) revealed by two-dimensional electrophoresis of seedling proteins. *Theor Appl Genet* 74, 218-223
- Behl RK (1981) Studies on genetic divergence heterosis and gene action in hexaploid triticale. *Thesis abstracts* 7, 2, 150-151
- Behl RK, Singh VP, Paroda RS (1985) Genetic divergence in relation to heterosis and specific combining ability in triticale. *Indian J Genet Plant Breed* 45, 2, 368-375
- Benzecri JP (1973) *L'analyse des données. 1. La taxonomie*. Dunod, Paris
- Branlard G, Chevalet C (1984) Sur la diversité des blés tendres cultivés en France. *Agronomie* 4, 933-938
- Branlard G, Picard B, Courvoisier C (1990a) Electrophoresis of gliadins on long acrylamide gels : method and nomenclature. *Electrophoresis* 11, 4, 310-314
- Branlard G, Autran JC, Rousset M, Dardevet M, Koenig J (1990b) *Catalogue des sous-unités de hauts poids moléculaires des gluténines des blés*. INRA, 56 p
- Bushuk W, Zillman RR (1978) Wheat cultivar identification by gliadins electrophoregrams. 1 - Apparatus, method and nomenclature. *Can J Plant Sci* 58, 505-515
- Camussi A, Spagnoletti Zeuli PL, Melchiorre P (1983) Numerical taxonomy of Italian maize populations: genetic distances on the basis of heterotic effects. *Maydica* 38, 411-424
- Colas des Francs C, Thiellement H (1985) Chromosomal localization of structural genes and regulators in wheat by 2-D electrophoresis of ditelosomic lines. *Theor Appl Genet* 71, 31-38
- Courvoisier C (1984) Polymorphisme des protéines de réserve du blé tendre. Thèse doctorat d'état univ Paris-Sud Orsay, 135 p
- Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985a) Relationship between coefficient of parentage and genetic similarity indices in the soybean. *Crop Sci* 25, 529-532
- Cox TS, Lookhart GL, Walker DE, Harrel LG, Albers LD, Rodgers DM (1985b) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electrophoresis patterns. *Crop Sci* 25, 1058-1063
- Damerval C, Hébert Y, De Vienne D (1987) Is the polymorphism of protein amounts related to phenotypic variability? A comparison of two-dimensional electrophoresis data with morphological traits in maize. *Theor Appl Genet* 74, 194-202
- Gallili G, Feldman M (1984) Mapping of glutenin and gliadin genes located on chromosome 1B of common wheat. *Mol Gen Genet* 193, 293-295
- Klose J (1982) Genetic variability of soluble proteins studied by two-dimensional electrophoresis on different inbred mouse strains and on different mouse organs. *J Mol Evol* 18, 315-328
- Leammli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685
- Lefort-Buson M (1986) Hétérosis chez le colza oléagineux (*Brassica Napus* L) : analyse génétique et phénotypique. Thèse univ Paris-Sud Orsay, 228 p
- Lefort-Buson M, Hébert Y, Damerval C (1988) Les outils d'évaluation de la diversité génétique et phénotypique. *Agronomie* 8, 3, 173-178
- Léonardi A (1989) Variabilité et hérédité des quantités de protéines chez le maïs. Analyse par électrophorèse bidimensionnelle et relation avec la variabilité agromorphologique. Thèse univ Paris VII, 170 p
- Lookhart GL, Jones BL, Walker DE, Hall SB, Cooper DB (1983) Computer-assisted method for identifying wheat cultivars from their gliadin electrophoregrams. *Cereal Chem* 60, 111-115
- Malécot G (1948) *Les mathématiques de l'hérédité*. Masson, Paris, 63 p
- Martinez WOJ, Goodman MM, Timothy DH (1983) Measuring racial differentiation in maize using multivariate distance measures standardized by F₂ populations. *Crop Sci* 23, 775-781
- Payne PI (1983) Breeding for protein quantity and protein quality in seed crops. In: *Seed proteins* (Daussant J, Mosse J, Vaughan J, eds) Academic Press, London, 223-253
- Payne PI, Corfield KG, Blackman JA (1979) Identification of a high molecular weight subunit of glutenin whose presence correlates with bread-making quality in wheats of related pedigree. *Theor Appl Genet* 55, 153-159

- Picard B (1990) Étude de la diversité biochimique et génétique du blé tendre : application à la connaissance de l'hétérosis. Thèse univ Clermont II, 121 p + annexes
- Picard B, Branlard G, Oury FX, Berard P, Rousset M (1992) Étude de la diversité biochimique et génétique du blé tendre. II. application à la prédiction de l'hétérosis. *Agronomie*, à paraître
- Prasad SK, Singh TP (1986) Heterosis in relationship to genetic divergence in maize (*Zea mays* L). *Euphytica* 35, 3, 919-924
- Price S, Shumaker KM, Kahler AL, Allard RW, Hill JE (1984) Estimates of population differentiation obtained from enzyme polymorphisms and quantitative characters. *J Hered* 75, 141-142
- Rodgers DM, Murphy JP, Frey KJ (1983) Impact of plant breeding on the grain yield and genetic diversity of spring oats. *Crop Sci* 23, 737-740
- Seguin M (1990) Étude de l'hétérozygotie chez le blé tendre par l'électrophorèse bidimensionnelle des protéines totales et relation avec des caractères agronomiques. Thèse univ Paris-Sud Orsay, 211 p
- Shamsuddin AKM (1985) Genetic diversity in relation to heterosis and combining ability in spring wheat. *Theor Appl Genet* 70, 3, 306-308
- Singh VP, Behl RK, Chowdhury RK (1985) Genetic divergence vis-à-vis heterosis in hexaploid triticale. *Crop Improv* 12, 1, 46-48
- Vienne (de) D, Damerval C (1985) Mesures de la divergence génétique. 3 - Distances calculées à partir de marqueurs moléculaires. In: *Les distances génétiques*. INRA, 39-57
- Zivy M, Thiellement H, Vienne De D, Hofmann JP (1984) Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional electrophoresis. 2 - Genetic differences between two lines and two groups of cytoplasm at five developmental stages or organs. *Theor Appl Genet* 68, 335-345