

(1) UN PROCEDE TOTALEMENT AUTOMATIQUE D'IDENTIFICATION DES VARIETES DE BLE PAR TRAITEMENT INFORMATIQUE DES ELECTROPHOREGRAMMES PROTEIQUES (\*)

Jean-Claude AUTRAN et Philippe ABBAL

Laboratoire de Technologie des Céréales I.N.R.A.  
9 Place Viala. 34060 Montpellier Cedex, France

Après électrophorèse, les protéines du blé et particulièrement les fractions protéiques de réserve présentent un remarquable polymorphisme intervariétal.

(2) PROTEINES DE RESERVE DU BLE: REMARQUABLE POLYMORPHISME INTERVARIETAL

---> CARACTERISATION DES GENOTYPES EN GENETIQUE ET EN SELECTION

---> IDENTIFICATION DES VARIETES AU NIVEAU DU COMMERCE DU GRAIN

Ce polymorphisme (indépendant des facteurs agronomiques, climatiques ou culturaux) a été largement exploité pour caractériser et discriminer les génotypes de blé dans les études de génétique ou de sélection, ainsi que, depuis une douzaine d'années, pour l'identification des variétés de blé (et d'autres céréales), dans les lots commerciaux, au niveau des laboratoires de l'industrie (meunerie, semoulerie), à l'exportation ou la répression des fraudes. C'est dans l'optique de l'identification des variétés que le présent travail a été conduit.

Historiquement, ont été utilisées: la SGE des gliadines (3) (illustration des du polymorphisme et des différences qualitatives et quantitatives), la PAGE des gliadines (4), la SDS-PAGE des protéines totales (HMW, LMW, gliadines) (5), ces deux dernières étant officiellement utilisées en France. D'autres techniques ont également un potentiel intéressant comme l'IEF (6) ou la 2-DIM (7) mais ne sont pas utilisées en routine dans les laboratoires de la profession céréalière.

En pratique, ces procédés reposent donc sur l'existence de différences qualitatives et quantitatives dans les bandes électrophorétiques. On a ainsi constitué des catalogues de diagrammes schématisés (BUSHUK Manuel: comparaison croix/mobilités Marquis) (8) (9). On procède alors au repérage de constituants discriminants et, grâce à l'établissement de sortes de clés chimiotaxonomiques (10), on retrouve, en suivant une série de dichotomies, le nom de la variété inconnue.

Toutefois, en dépit des innovations considérables qui se sont produites en matière d'informatisation de la saisie et du traitement des données électrophorétiques, innovations qui ont déjà beaucoup profité au secteur clinique et médical, l'identification des variétés de céréales reste encore aujourd'hui, en France et dans le monde, essentiellement artisanale: manuelle et visuelle. Revenir à D3 (PAGE).

Dans les très rares cas où un ordinateur a été introduit dans le système, (3-4 labos de recherche au monde, aucun labo de contrôle) c'est

(\*) Exposé présenté au Colloque 1987 de la Société Française d'Electrophorèse, Lyon, 12-14 Novembre 1987

uniquement dans la phase finale, pour comparer des données qu'un technicien a préalablement décodé (calcul des mobilités relatives et estimation des intensités) visuellement puis entré au clavier. Ceci requiert du personnel hyperqualifié, qui possède en mémoire les principales caractéristiques du diagramme (11), capable de retenir chaque année les diagrammes des nouvelles variétés, capable d'identifier les bandes devant être utilisées comme témoins de mobilité relative et de décider par exemple si un composant présent à l'état de traces doit être considéré comme une bande ou appartenant au bruit de fond.

Nous avons donc oeuvré dans le sens d'une informatisation du système de lecture, conduisant à un diagnostic automatique de la variété, de manière à le rendre utilisable même par des non spécialistes des électrophorogrammes de blé.

Il est certain que l'établissement d'un système totalement automatique et informatisé de lecture des diagrammes et de diagnostic des variétés présente de multiples difficultés. On sait, naturellement, que des systèmes de saisie très sophistiqués sont aujourd'hui au point, des algorithmes de détection et de quantification des pics ont été décrits et sont disponibles dans le commerce. Mais nous savons bien que la variabilité expérimentale des nombreux paramètres de l'électrophorèse ne permet pas toujours une excellente répétabilité du pouvoir de résolution du gel. Il est très difficile de parvenir à une évaluation automatique qui soit parfaitement répétable dès lors qu'on a affaire à des pics multiples ou non gaussiens ou à des épaulements. Même lorsque la résolution est excellente, il existe toujours un niveau ultime dans la finesse des séparations au delà duquel une évaluation automatique n'est plus répétable. D'ailleurs tous les systèmes commerciaux permettent des interventions manuelles: addition de pics non séparés, élimination de pics artefacts,...

Il est vrai aussi que l'étude des électrophorogrammes de protéines du blé présente de multiples difficultés: les différentes variétés ne renferment absolument pas les mêmes composants et qui n'ont pas la même allure générale (comme pour les protéines du sérum): chaque variété renferme 20 à 30 bandes, prises dans un ensemble de plus de 60 bandes possibles, aucune n'étant commune à toutes les variétés cultivées, ne rendant guère possible l'utilisation d'étalons internes ou externes pour corriger les différences expérimentales de migration et obtenir des diagrammes normalisés pouvant être comparés.

Dans le contexte que nous venons de résumer, le système que nous présentons (constitué d'un densitomètre et d'un ordinateur IBM-PC) est une première tentative d'automatisation totale de la lecture des électrophorogrammes dans le but d'identifier les variétés de blé. L'objectif visé est de supprimer toute intervention humaine entre le signal start de la lecture densitométrique et l'apparition à l'écran du nom de la ou des variétés identifiées.

#### (12) PREMIERE TENTATIVE D'AUTOMATISATION TOTALE DE LA LECTURE DES ELECTROPHOREGRAMMES DANS LE BUT D'IDENTIFIER LES VARIETES DE BLE

OBJECTIF: SUPPRIMER TOUTE INTERVENTION HUMAINE ENTRE LE SIGNAL START DU DENSITOMETRE ET L'APPARITION A L'ECRAN DU NOM DE LA (OU DES) VARIETE(S) IDENTIFIEE(S)

Les solutions retenues sont les suivantes:

1) Saisie des données au moyen d'un densitomètre à laser, mais en enregistrant alternativement des diagrammes de protéines du blé (en SDS-PAGE) et des

diagrammes de marqueurs de poids moléculaire répartis entre 205000 et 29000 et encadrant ainsi l'ensemble des protéines de réserve du blé (13).

2) Digitalisation des courbes brutes (640 points, cad couples distance/voltage)

3) Développement d'un logiciel (14) qui permet (outre les opérations classiques de filtrage, d'identification des maxima des pics de la courbe, d'identification des minima par lesquels on réalise une approximation de la ligne de base laquelle est ensuite soustraite de la courbe brute, etc.)

1) de RECALIBRER LA COURBE INCONNUE EN MOBILITES RELATIVES à partir de celles des TEMOINS

2) d'identifier les MOBILITES RELATIVES et de calculer les HAUTEURS des pics de la COURBE INCONNUE

3) D'établir ainsi automatiquement le SCHEMA DE LA VARIETE INCONNUE avec une même structure de fichier que ceux antérieurement obtenus par la méthode visuelle: mobilités relatives + intensités de 1 à 5

4) De déterminer par un calcul d'INDICES DE SIMILARITE la variété qui, dans la banque, présente le diagramme le plus proche de celui de l'inconnue.

La figure (15) donne un exemple de courbe densitométrique corrigée et la figure (16), le schéma de diagramme calculé par l'ordinateur.

Ce diagramme étant comparé à tout ceux de la banque, la réponse apparaissant à l'écran est le nom de la variété détectée, ainsi que la distribution des indices de similarité avec le reste de la banque (17).

Plusieurs cas de figure peuvent intervenir: celui d'une identification sans ambiguïté, celui d'une identification de plusieurs variétés voisines (18), celui de l'absence de diagramme correspondant à celui de l'inconnu (19).

Egalement, le schéma des diagrammes des variétés les plus proches est reconstitué à l'écran à des fins de vérification (20).

#### Discussion:

La limite majeure vient naturellement encore de la répétabilité du niveau de résolution des diagrammes. Contrairement à un diagramme issu de plusieurs lectures à partir desquelles on a calculé une moyenne et sur lesquelles un spécialiste a affiné la séparation manuellement, un diagramme obtenu de façon entièrement automatique risque a priori d'être davantage sensible aux fluctuations expérimentales. (L'identification peut ainsi s'avérer impossible ou erronée à partir de diagrammes présentant une résolution insuffisante et surtout variable).

(21) LIMITE ACTUELLE: REPETABILITE DU NIVEAU DE RESOLUTION DES DIAGRAMMES

APPLICATION POSSIBLE A DES LABORATOIRES DE CONTROLE CEREALIER TRAVAILLANT SUR UN NOMBRE LIMITE DE VARIETES

PERSPECTIVES: SAISIE DES DONNES PAR CAMERA CCD OU SCANNER

Le système proposé résulte donc d'un compromis entre un maximum de finesse et de précision dans la détermination du schéma de diagramme (pas

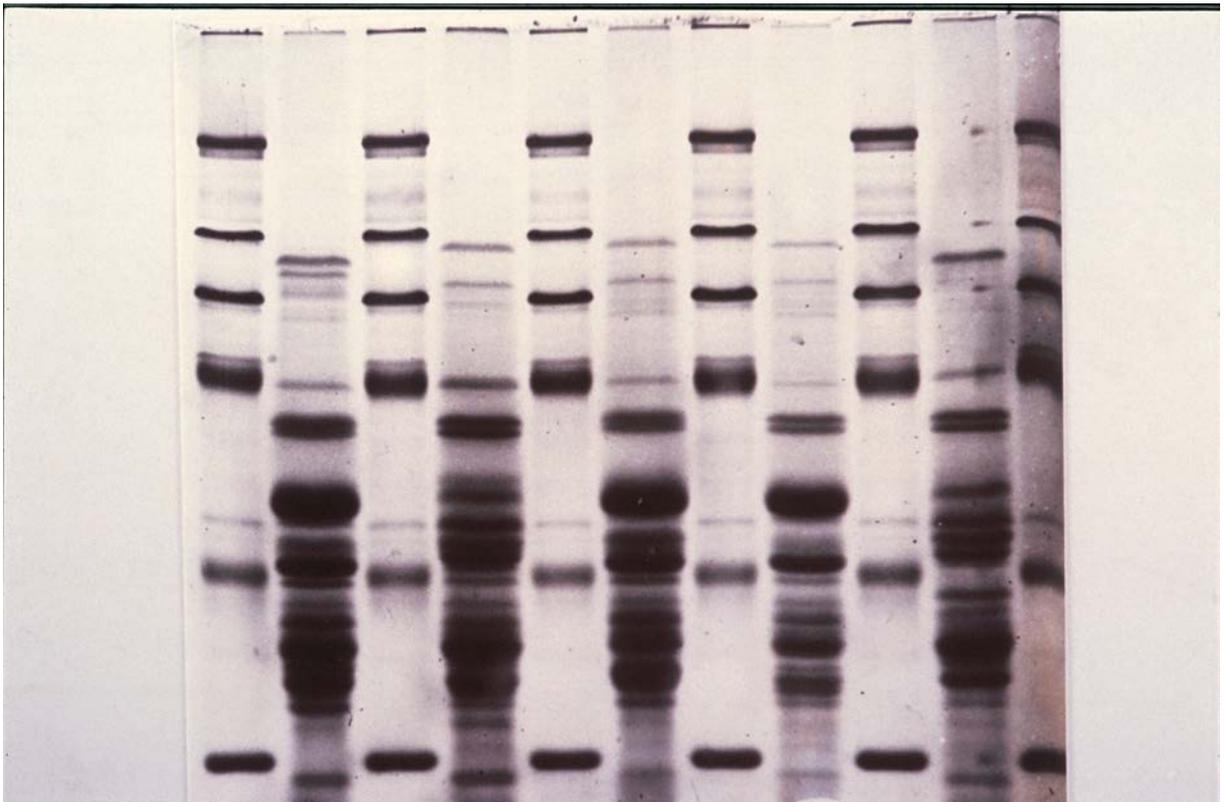
assez de précision ne permet pas de discriminer suffisamment de variétés) et un maximum de répétabilité (trop de précision est en effet illusoire compte tenu des fluctuations expérimentales et des caractéristiques propres aux diagrammes du blé).

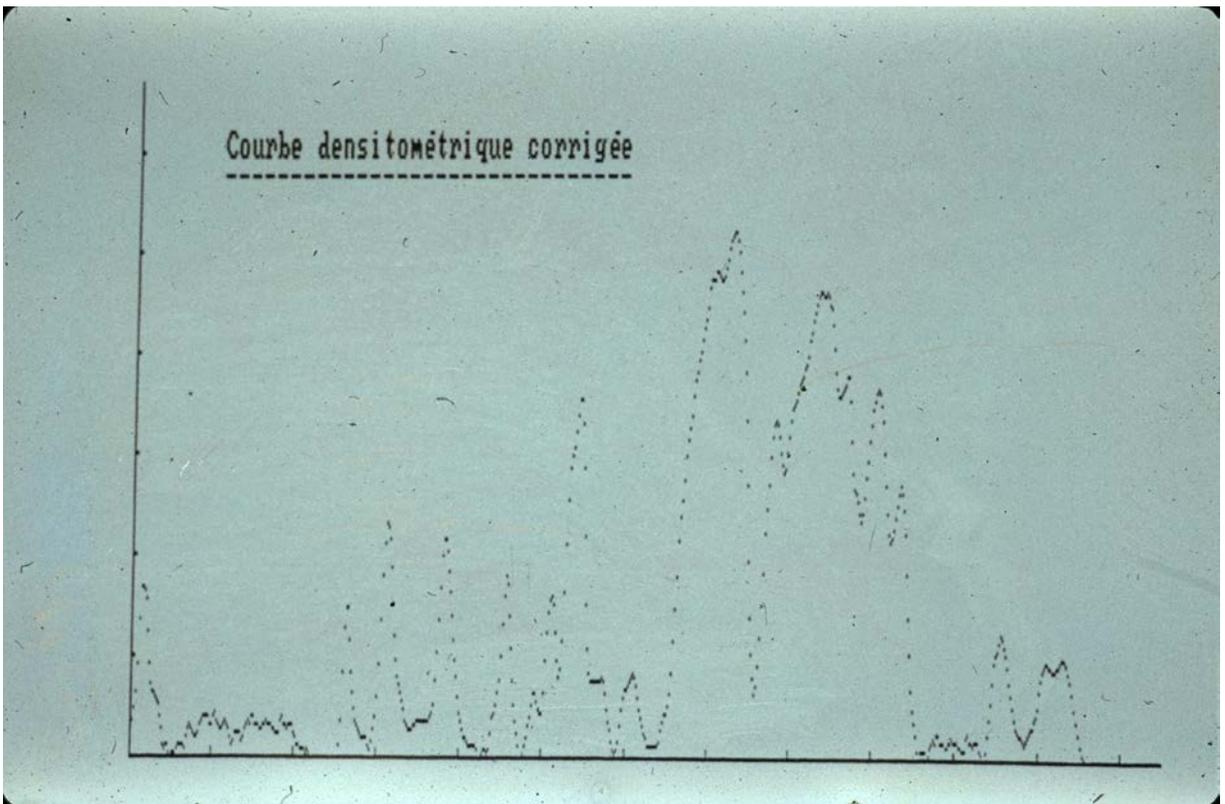
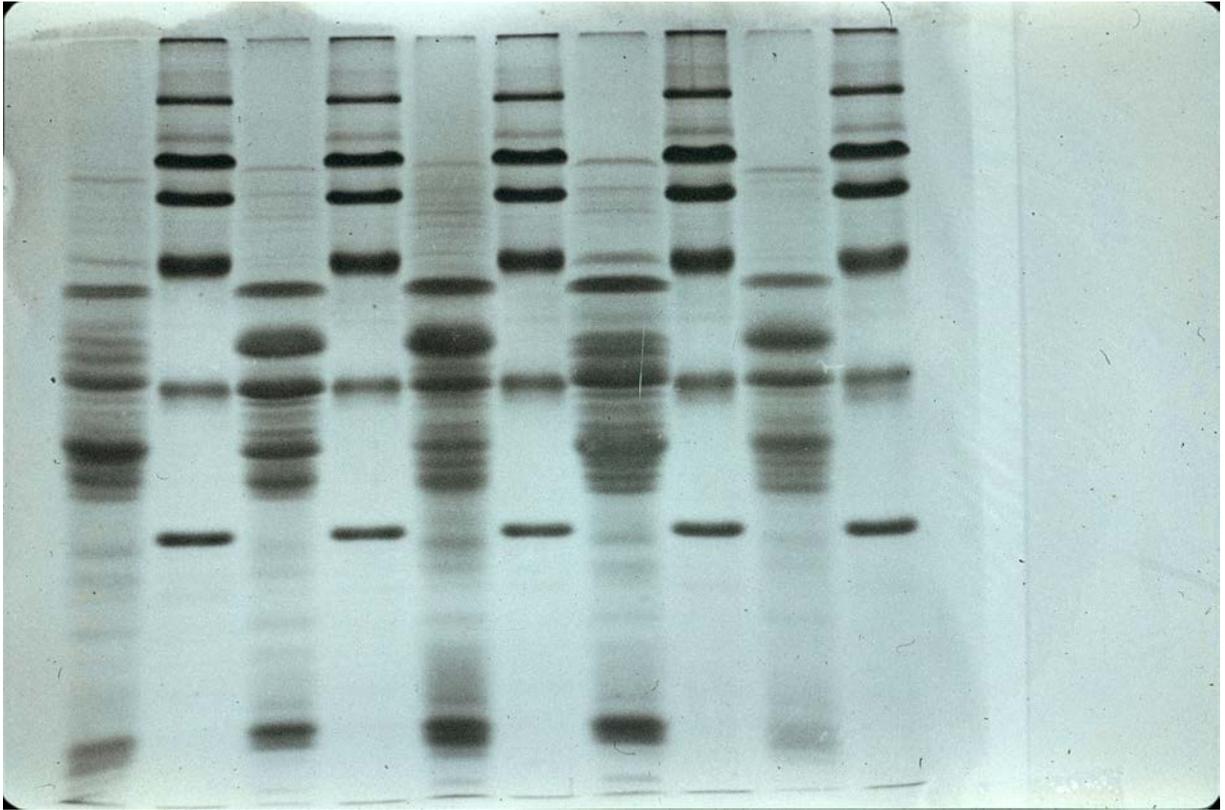
Ceci a orienté les choix réalisés pour l'établissement de l'algorithme de calcul qui fait le plus souvent appel à des solutions simples (approximation aux valeurs entières les plus proches, correction de ligne de base très simplifiée, 5 niveaux seulement d'intensité - de hauteur - des pics, souplesse dans le calcul de l'indice de similarité permettant à des bandes non identiques mais très voisines d'être prises en compte).

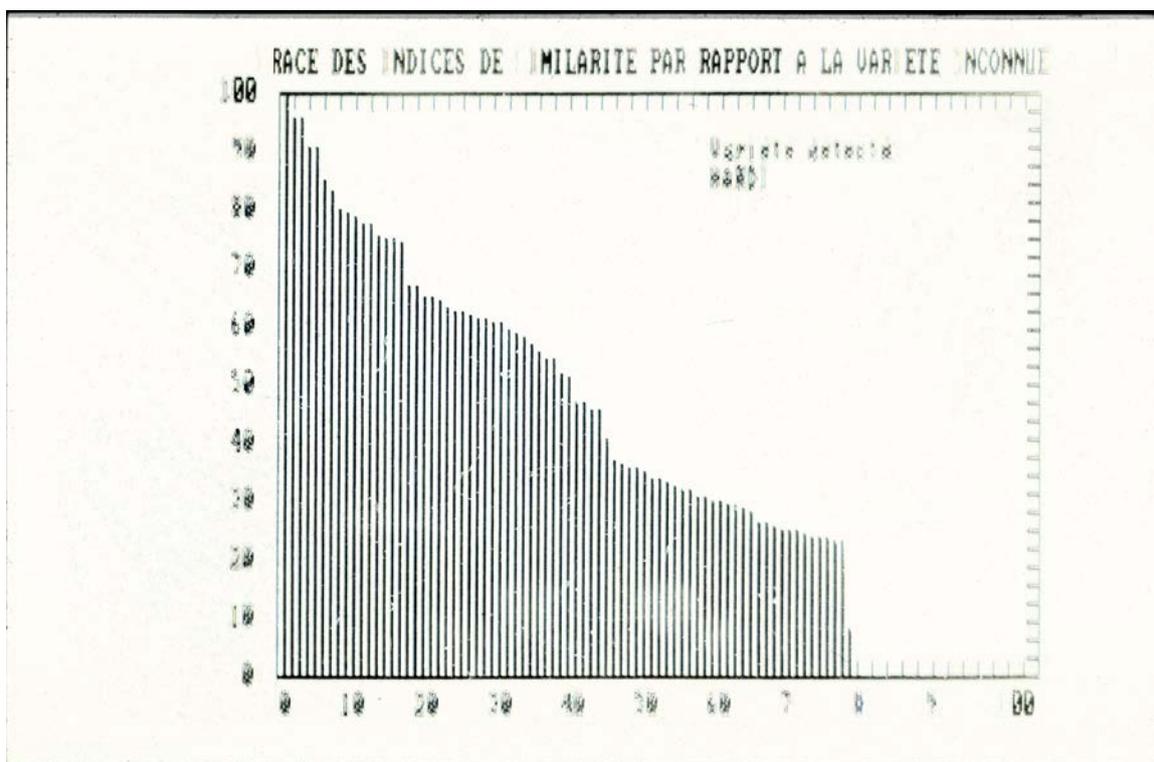
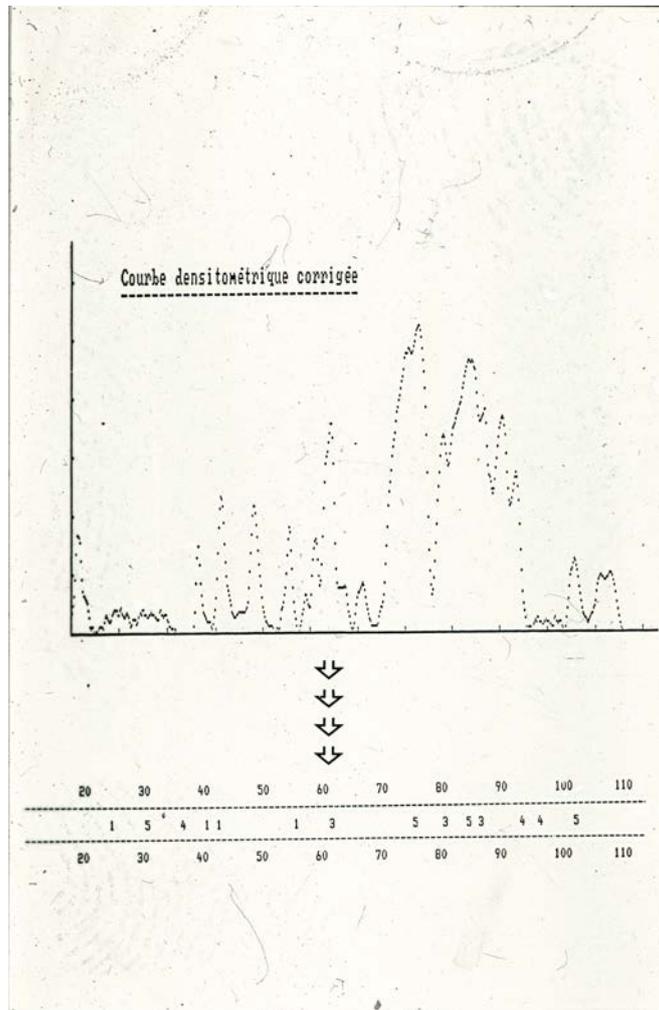
Dans sa version actuelle, le système ne semble donc pas adapté pour la caractérisation d'un grand nombre de géotypes trop étroitement apparentés tel que le Catalogue complet des 200 variétés de blé françaises.

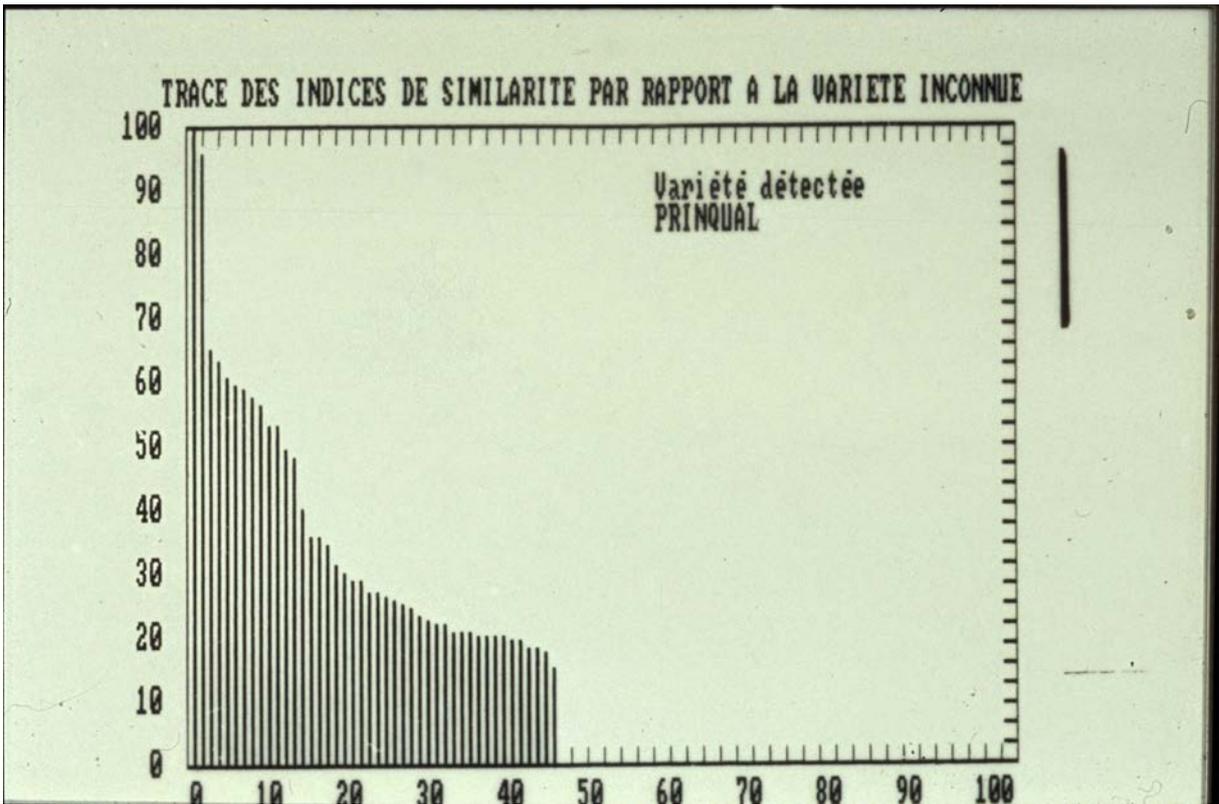
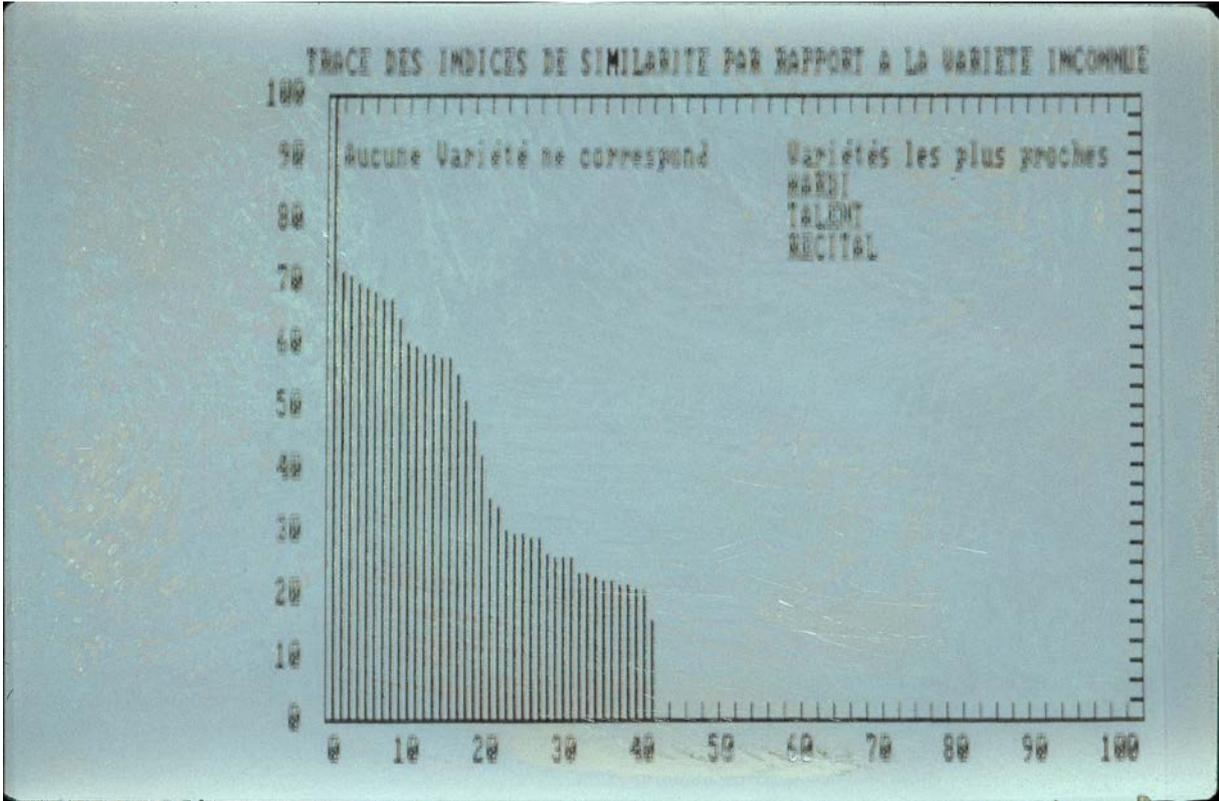
L'avantage de l'automatisme complet peut par contre convenir pleinement aux laboratoires qui ne possèdent pas de personnel d'une haute technicité, tels que les laboratoires de contrôle de la profession céréalière et qui n'ont chacun affaire qu'à des lots commerciaux d'une région bien déterminée lesquels ne renferment qu'un nombre limité (10 ou 20) de variétés.

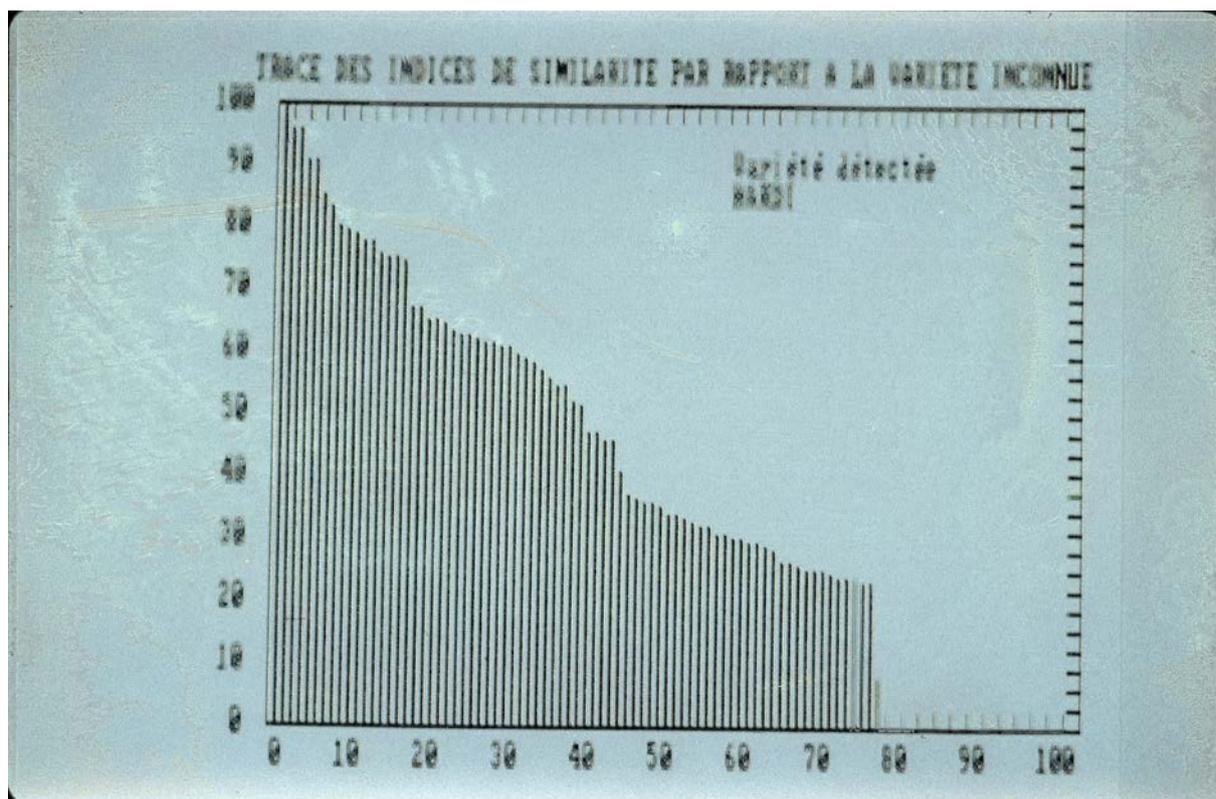
Comme un autre facteur limitant est le temps de densitométrie, nous envisageons dorénavant de saisir les données soit à partir d'un système de caméra CCD, soit à partir d'un scanner. Cela devrait d'ailleurs permettre d'améliorer la précision du fait de la possibilité de faire instantanément plusieurs saisies et d'en faire la moyenne.











**Table 1.** SDS-PAGE protein arrays of the 12 most extensively grown french wheat cultivars. The 3 upper patterns are from durum wheat type, the other 9 patterns are from soft wheat type. Relative band intensity is expressed in the 1 to 5 scale (1 is lightest, 5 is strongest).

	20	30	40	50	60	70	80	90	100	110							
MONDUR	1	5	4	1	1	3	5	3	5	3	4	4	5				
CANDU		5	4		2	3	4	2	3	3	4	1	1	2	4	4	
CAPDUR			5	3	2	1	4	5	3	5	3	2	3	5	4		
HARDI		5	5		3	2	2	4	2	2	3	5	1	5	4		
CAPITOLE		5	5	1	2	2	2	3	1	2	2	4	3	4	4	4	
PRINQUAL		5	5		3	1	2	2	2	1	2	5	3	5	5		
TALENT		5	5	1	3	2	2	2	2	2	2	5		5	5		
BEAUCHAMP		5	5	2	3	1	2	1	2	3	1	2	5		5	5	
ARMINDA		4	5		2	2	1	2	4	1	1	5	5	3	5	4	
CAMP RENY		4	5	1	2	1	1	2	1	1	2	4	5		4	4	4
FIDEL		3	5	2	2	1	2	3	1	1	1	5	5	1	3	5	4
FESTIVAL		4	5		2	3	1	1	2	2	2	5	2		4	4	4

DIAGRAMMES DES VARIETES LES PLUS PROCHES

