Computer-Assisted Method for Identifying Wheat Cultivars from Their Gliadin Electrophoregrams[1]

G. L. LOOKHART, B. L. JONES, D. E. WALKER,[2] S. B. HALL, and D. B. COOPER[3]

Cereal Chem. 60(2):111-115

ABSTRACT

A computer program was developed to aid in identifying wheat cultivars by comparing their gliadin electrophoregrams with those of standard varieties stored in the computer memory. The main output of the program is a relative percent similarity plot, which shows the similarity of the electrophoregram of the cultivar in question to that of each of the standard cultivars. In addition, the program lists the names of those cultivars whose electrophoretic patterns are most similar to that of the "unknown," in order of decreasing similarity. The weighting factors used for calculating similarities are listed and explained. Examples of the use of the program for identifying "unknown" cultivars are presented.

Many authors (Wrigley et al 1982), have reported the identification of wheat cultivars by their gliadin electrophoregrams (electrophoretic band patterns), obtained from either starch or polyacrylamide gel electrophoresis (PAGE). Catalogs have been published that list the gliadin PAGE patterns of the major wheat cultivars grown in several countries: Australia (Wrigley and Shephard 1974); Canada (Zillman and Bushuk 1979); France (Autran and Bourdet 1975); Italy (Dal Belin Peruffo et al 1981); and the United States (Jones et al 1982). Cultivar identification necessitates comparing the electrophoregram of an unidentified cultivar with each of the 90 or so standard electrophoretic patterns listed in a cultivar catalog. Making such comparisons manually is very time consuming. Fortunately, computers can quickly and reliably compare the patterns of unidentified cultivars with the stored standard electrophoregrams. Two separate approaches to identify wheat cultivars via computer analysis of gliadin electrophoregrams have been published (Bushuk et al 1978, Wrigley 1980, Wrigley et al 1981). The first method (Bushuk et al 1978) is very complex and requires major computer investment. The second method (Wrigley 1980, Wrigley et al 1981) is not flexible enough for analyzing the large genetic variations in wheat cultivars in the world today. This article presents a very flexible but relatively inexpensive and simple method of using a computer to

[1]Cooperative investigations, Agricultural Research Service, U.S. Department of Agriculture and the Department of Grain Science and Industry, Kansas State University. Contribution 82-360-J, Department of Grain Science and Industry, Kansas Agricultural Experiment Station, Manhattan 66506.

Mention of firm names or trade products does not constitute endorsement by the USDA over others not mentioned.

[2]Research chemists and electronic engineer, respectively, U.S. Grain Marketing Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Manhattan, KS 66502.

[3]Research assistants, Department of Grain Science and Industry and Department of Biochemistry, respectively, Kansas State University, Manhattan 66506.

aid in identifying "unknown" wheat cultivars from their gliadin electrophoregrams.

MATERIALS AND METHODS

Electrophoregram Preparation

Wheat samples (3 g) were ground, extracted, and subjected to PAGE, and the resulting gels were stained to produce visible electrophoregrams (Lookhart et al 1982).

Determination of the Relative Mobilities of Gliadin Bands

The destained PAGE gels were photographed, and reversal negatives of the gels were produced as reported previously (Lookhart et al 1982). The final reversals (consisting of black bands on a transparent background) were scanned with 500-nm light by using a Kratos SD 3000 spectrodensitometer operated in the transmittance mode. The output from the densitometer was plotted as density versus time by a Hewlett-Packard 3385 printer-plotter, which automatically printed the time each band was encountered. For best results, the reversals were scanned at 25 mm/min, and the printer-plotter was run at the same speed. That gave a density-versus-distance plot that was the same size as the reversal, making the bands on the reversals and the peaks on the printer-plotter scan easy to match up. The density plot gave semiquantitative information about the density of each band and a precise measure of how far down the gel each protein band had run. Because some bands that showed up on the densitometer traces were not integrated separately (shoulders on large peaks, etc.), the distances those bands had moved had to be measured manually. In each electrophoregram, some protein material did not enter the gel but remained at the origin of the electrophoregram and served as an indicator of the point of sample application.

Eight-slot gels were run with a standard, Marquis, which was normally applied in slots 1, 4, and 8. Five "unknown" samples were placed in the remaining sample wells. This ensured that no unknown samples were run in the slots adjacent to the edge of the gel, where "edge effects" were sometimes noticed. By comparing the Marquis patterns across the gel, an indication of the uniformity of conditions across the gel was obtained.

The distance from the origin to the center of the Marquis reference band (the heavy single band that ran slightly farther down the gel than the heavy doublet) was assigned a relative distance value of 50 units, following the practice of Bushuk and Zillman (1978). For each band of an electrophoregram, the distance (in minutes) from the origin to the center of that band was determined by the densitometer tracing and, after subtracting the distance from scan start to the point where the sample had been applied, was taken as the band migration distance. This migration distance value was divided by the migration distance of the Marquis standard band, and the result was multiplied by 50 to give the relative migration distance of the band. The precision of the relative mobilities was generally ±0.5 units. The mobilities of each band, calculated from the duplicate electrophoretic runs, were averaged and rounded off to the nearest whole unit.

Determination of Band Densities

The reversal negatives (positives) of the gels were placed on a light box (Lookhart et al 1982), and each band was subjectively assigned a numerical value from 1 (very light) to 5 (very dark), according to the method of Zillman and Bushuk (1979). The precision of the band density assignment was ±1 unit. Three people assigned density values to all bands, and the numbers assigned by the various investigators were identical in more than 95% of the cases. Originally, an attempt was made to assign band densities on the basis of the areas under the density peaks as determined by the scanning densitometer traces. That method was not reliable, however, because of problems with bands that showed up as shoulders on large peaks and because the background varied at different positions along the gel. Also, the reversals being scanned varied somewhat because of differences in exposure times and other factors. The exposure and developing times varied among the gels because the gels were sometimes stained and/or destained for different times. Some of the problems in assigning band densities from integrator data could probably be resolved by complicated computer massaging of the data, but assigning the densities as reported above was much simpler and more reliable.

Array Preparation

The electrophoregram of each cultivar contained 18 to 28 bands and was manually converted into 18 to 28 pairs of numbers for computer manipulation. Each pair of numbers describes one band. The first number of the pair described the band's mobility (the relative distance the protein band moved into the gel, or mobility value). The mobility was determined from the densitometric trace of the reversal negative. The second number of the pair corresponded to the band's density (the relative intensity of the stained protein band, or density value). The density was subjectively determined directly from the reversal negative. The complete set of mobility density pairs for every band of a given cultivar makes up the array for that cultivar. The mobility and density values were calculated and assigned as described by Jones et al 1982. The arrays were entered into the computer memory and stored on a hard disk.

Computer

A Data General S/130 Eclipse with a 10*M* byte hard-disk storage unit and a Tektronix 4014-1 CRT terminal were used to store and manipulate the data.

RESULTS AND DISCUSSION

Programming Considerations

A computer program for cultivar identification was developed that compared the protein band array of one cultivar with the arrays of all cultivars stored in the computer memory. It also listed the degree of similarity of the unknown to each known cultivar. Two major considerations were used to determine the overall similarity of the arrays of two cultivars: how well the band mobilities matched and how well the band-staining densities of bands having the same mobilities compared. Low-mobility bands were compared first, followed by the bands with progressively higher mobilities.

Correlating Protein Band Mobilities

The computer program first compared the protein band mobilities of the unknown cultivar to those of each standard cultivar stored in the computer memory. If the mobility of a protein band of the unknown was the same as or within ±1 distance unit (1/50th of the distance from the origin to the first dark band after the Marquis doublet) of a band of the standard cultivar, the bands were considered as matching, and their staining densities were compared. The ±1 unit flexibility was allowed because band mobilities were calculated to 0.1 unit and rounded off to the nearest whole number. Mobility values of a given protein band sometimes varied by ±1 distance unit when calculated from measurements of different gels, especially for the bands that moved farthest into the gel.

Correlating Protein Band Densities

After matching bands were detected, the densities of the matched bands were compared and a "match value" (MV) was assigned, as defined in Table I. It reflected the probability that the bands contained identical proteins. The larger the MV, the more likely the bands are to contain the same protein species, whereas those with small MVs are more likely to contain proteins that have different primary structures but that happen to migrate to the same position on the gel. The more similar the matching bands are in density, the higher the MV, and the more weight is given that match in determining overall (total array) similarity. MVs of individual band pairs listed in Table I were calculated from the formula: $MV = \Sigma Di/[Db(Db - Da)]$, where the $(Db - Da)$ term is replaced by 1 when $Db = Da$. In that equation, Di is the density of each

individual band being considered, Db is the density of the darker-staining band of the two being compared, and Da is the density of the lighter-staining band. For example, if bands of densities 2 and 3 are being compared, then $MV = (3+2)/[3\times(3-2)] = 5/3 = 1.67$, as is listed in Table I. The formula we devised for calculating MVs placed more emphasis on partial matches between denser bands than it did on partial matches of less dense bands, since the darker bands (3, 4, or 5) were more accurately determinable than the lighter bands (1 and 2). This was discussed in our earlier paper (Jones et al 1982). Many different formulas were investigated, and the reported formula worked better than any other.

Calculating Overall Similarities

To compare an unknown cultivar against a catalog of standard cultivars, a single value needed to be calculated that reflected the relative overall similarity between the band pattern (array) of the unknown and that of each of the standard cultivars. That was done by first summing the MVs for all matching protein bands of the unknown and standard cultivars. That gives a value called the overall match value (OMV). Thus, OMV = ΣMV for each set of cultivars being examined.

An adjustment was made to the OMV to give more weight to those bands whose mobility values were identical, and less weight to those bands whose mobility values differed by ±1 unit. That adjustment was made for each unknown-standard comparison by dividing the number of bands that had identical mobility values by the sum of the number of bands that have identical mobilities plus those that were within ±1 mobility unit (partial match) from each other. In addition, if a band of the unknown exactly coincided with one band of a standard cultivar, and was within ±1 unit of a second band, the ±1 unit partial match was ignored in making the calculations. The mobility adjustment factor (MAF) was: MAF = Σ bands with identical mobilities/Σ bands with identical or partial match mobilities. The MAF was then multiplied by the OMV, and the resulting product was raised to the 0.1 power to give the band correlation factor (BCF), thus, $BCF = [(MAF) \times (OMV)]^{0.1}$. The exponent value of 0.1 was arrived at by trial and error. That value was used to compress the scale so that BCF values are reduced to manageable proportions. The BCF was then linearly normalized to give relative percent similarities (RS) that ranged between 100% (identical band patterns) and 0% (complete dissimilarity). The RS values are not, however, strict percentages, since they give more weight to bands with identical mobilities than to those with partial match mobilities.

Computer Output: Cultivar Similarity

The arrays of a typical durum, Rugby, and bread wheat, Tam W-101, were individually compared, using the computer program, against the arrays of all of the cultivars stored in the computer catalog. Graphs of the relative percent similarities of Rugby and Tam W-101 versus the 88 standard cultivars are shown in Figs. 1 and 3, respectively. The program automatically lists (Figs. 1 and 3) the names of those cultivars that are most similar to the "unknown" data that were entered. In these cases, the data from standards were entered, so the similarity was 100% for Rugby and Tam W-101, respectively. The listing is truncated after the first 12 cultivars because they are the most important for identification. The program is written so that truncation can be easily varied from 0 to 88.

In Fig. 1, Rugby is 100% similar to Botno because their electrophoregrams are identical. Cultivars that are closely related have very similar electrophoregrams (Jones et al 1982). For example, the high degree of similarity (76% or greater) of the first seven wheat cultivars in Fig. 1 is indicative of the genetic similarity of the durums, all of which lack the D genome. That similarity is also apparent in the mobility-density graph shown in Fig. 2, which is a computer simulation of the band patterns of the gels of the 12

TABLE I
Density Match Value

Peak Density[a]		Match Value[b]	Peak Density		Match Value
Da	Db		Da	Db	
1	1	2.00			
1	2	1.50	3	3	2.00
1	3	0.65	3	4	1.75
1	4	0.41	3	5	0.80
1	5	0.30			
2	2	2.00	4	4	2.00
2	3	1.67	4	5	1.80
2	4	0.75			
2	5	0.47	5	5	2.00

[a] Db = more densely stained band; Da = less densely stained band. When Db = Da the (Db−Da) is set equal to 1.

[b] Match value = Σ Di/[Db(Db − Da)], where Di = individual peak density.

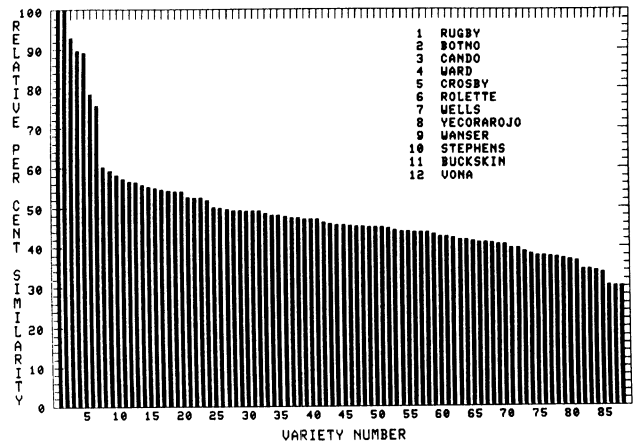

Fig. 1. Relative percent similarity plot for Rugby.

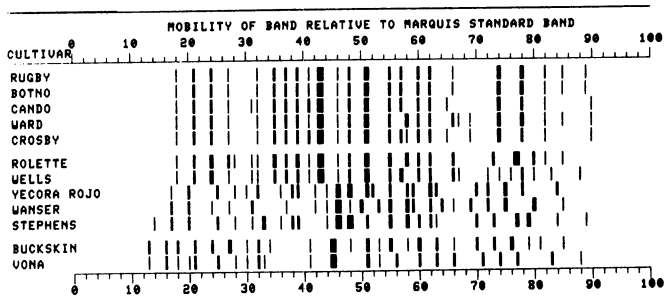

Fig. 2. Mobility-density plot (band patterns) for the top 12 cultivars listed in Fig. 1. Each pattern approximates the electrophoregram of a given cultivar. The mobility units are indicated on the horizontal axis. The relative band intensities (densities) are defined by the widths of the lines. Single width (narrow) line = 1s, double width = 2s, triple width = 3s, quadruple width = 4s, and quintuple width = 5s.

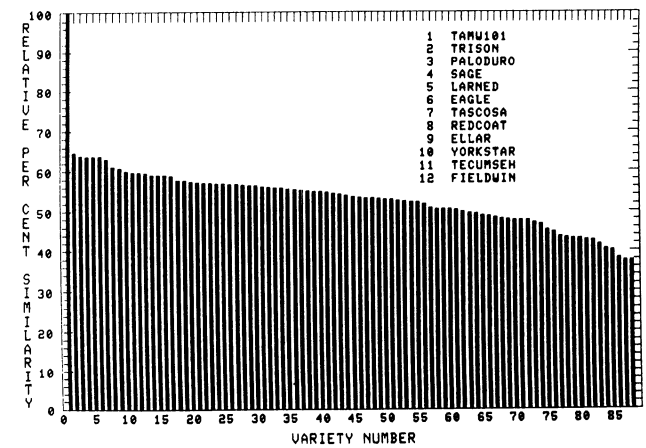

Fig. 3. Relative percent similarity plot for Tam W-101.

cultivars showing the greatest similarities to the "unknown." In all of the durum wheat cultivars, Rugby through Wells, over 90% of the bands in the first 60 mobility units match exactly.

The percent similarity graph for Tam W-101 (Fig. 3), shows that the electrophoregram of this cultivar is quite different from the electrophoregram of any other cultivar in the computer memory, the nearest cultivar showing only 65% similarity. However, the six cultivars with electrophoregrams most similar to that of Tam W-101 were all hard red winter wheats, like Tam W-101. In addition, only 2% relative similarity separated the first six cultivars. In general, cultivars of a given wheat class were more similar to other members of that class than to cultivars of differing classes. As shown in Fig. 3, however, where soft wheat (Redcoat), hard red spring wheat (Ellar), and three white wheats (Yorkstar, Tecumseh, and Fieldwin) were included in the 12 cultivars most similar to the hard red winter Tam W-101, the similarities sometimes overlap the various classes. Fortunately, the very high degree of similarity of band arrays found in the durums examined was not present in the other wheat classes. It is this diversity of band patterns that has made possible the identification of wheat cultivars from their mobility-density arrays.

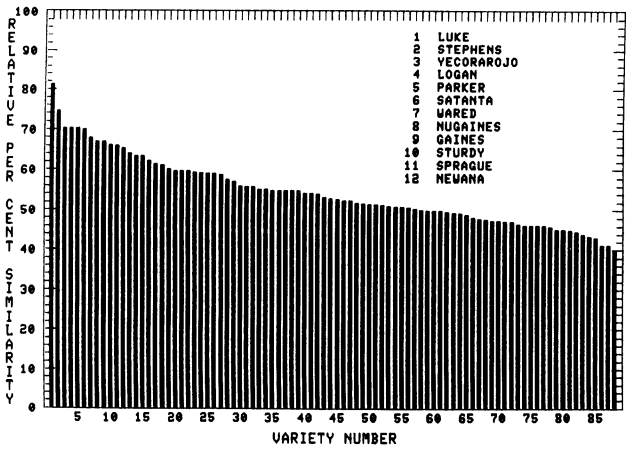

Fig. 4. Relative percent similarity plot for "unknown" 1.

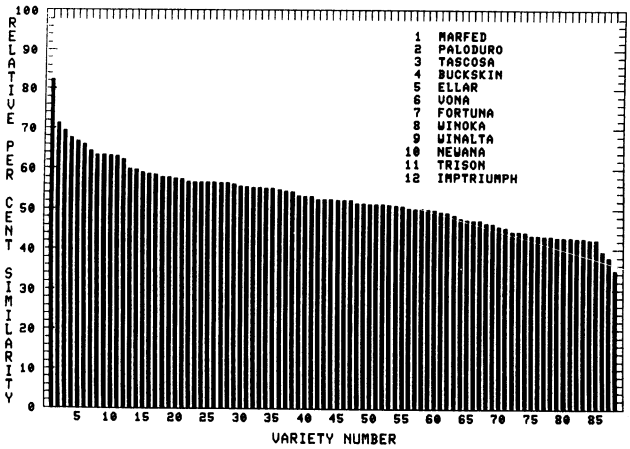

Fig. 5. Relative percent similarity plot for "unknown" 2.

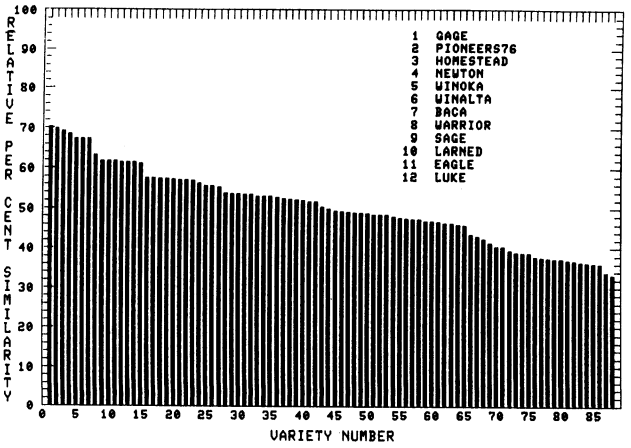

Fig. 6. Relative percent similarity plot for "unknown" 3.

Computer Output—Unknown Identification

The usefulness of the program is shown by the identification of five wheat grain samples, which were randomly selected from among the standard cultivars, the arrays of which were stored in the computer memory. The wheats were ground and extracted, and their gliadin electrophoregrams were made and processed as discussed previously. The mobility-density arrays of the unidentified test samples were individually entered into the program. The relative percent similarity graphs for unknowns 1–5 are shown in Figs. 4–8, respectively.

The computer program identified unknown 1 as Luke (Fig. 4), since the mobility-density array for Luke was 7% more similar to that of the unknown than to the array of any other cultivar stored in the memory. Unknown 1 was later identified as Luke by the person who selected the unknowns. In every case, where one cultivar was favored by at least 5 percentage points of similarity over any other cultivar, identification was found to be correct. If the similarity advantage of the top cultivar was less than 5%, then identification could be achieved in either of two ways. The easier way, which worked most of the time, was to make a visual comparison of the densitometer scans of the unknown with those of the cultivars listed as being most similar. In cases in which more than one cultivar was very similar to the unknown, the most similar cultivars and the unknown were electrophoresed on the same gel, and a visual

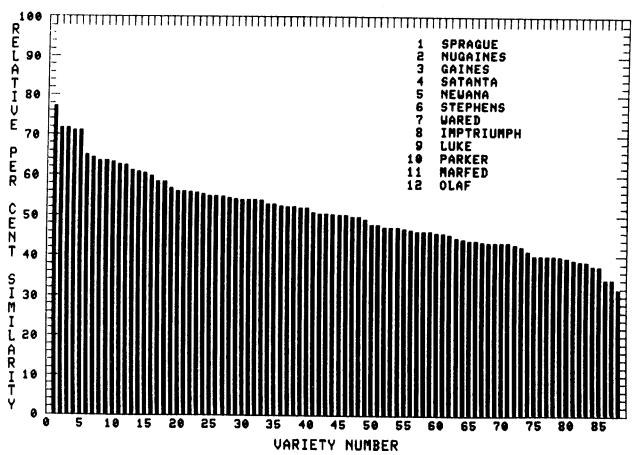

Fig. 7. Relative percent similarity plot for "unknown" 4.

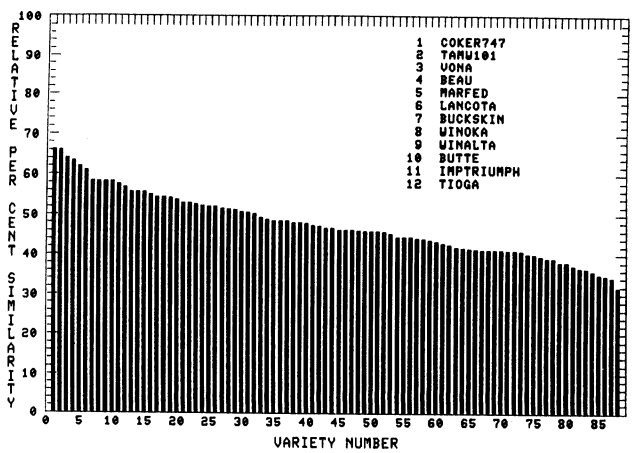

Fig. 8. Relative percent similarity plot for "unknown" 5.

discrimination was made. The fact that the relative percent similarity of Luke to unknown 1 was not 100% was a reflection of the fact that slight differences in band mobilities and band density assignments were found when a given cultivar was electrophoresed on different gels (Jones et al 1982). The cultivars most probably identical to the unknown, those within 5 percentage points of the most similar cultivar, were predominately members of the same wheat class (eg, Stephens and Luke were white wheats).

Unknown 2 was easily and accurately identified as Marfed (Fig. 5), since the array of unknown 2 was 11% more similar to the array of Marfed than to that of any other cultivar in the memory. Grouping into wheat classes was not as distinct as in Fig. 4, as Marfed was a white wheat and the three next most similar cultivars were all hard red winter wheats.

For identifying unknown 3 (Fig. 6), the program listed eight cultivars within a 5% relative similarity span. Two of the eight, Winoka and Winalta, have identical electrophoregrams, so that only seven different densitometer scans had to be compared with that of the unknown. Such comparison positively identified unknown 3 as Pioneer S-76, the second cultivar in the listing, but tied for first in relative percent similarity. The top seven cultivars listed were within 3% relative similarity. The first eight most similar cultivars were grouped into seven hard red winters and one soft red winter, Pioneer S-76.

Unknown 4 (Fig. 7) was identified by the fact that it was most similar to Sprague (77% similarity). The two next most similar cultivars showed 5% less similarity. The cultivars in the top 5% relative similarity were all white wheats. Thus, some class grouping was again found.

Unknown 5 was identified by comparing the densitometric scans of the unknown with those of the six cultivars that had relative percent similarities between 66 and 61% (Fig. 8). The comparison unambiguously identified the unknown as Tam W-101. The wheat classes found in the top six cultivars listed in Fig. 8 were three hard red winters, two soft red winters, and one white wheat.

Three of the five unknowns were unambiguously identified by the program, and the other two were tied with other cultivars for first in relative percent similarity. Both were readily identified by comparing the densitometer scans of the unknowns with the scans of the appropriate cultivars. Thus, the computer program aided in the identification of wheat varieties if they were listed in the catalog of standard cultivars. Because some of the standard cultivars have identical electrophoregrams (Jones et al 1982), they could not be differentiated. The program provided a listing of similar cultivars in decreasing order of similarity. The cultivar with the highest relative similarity was generally the unknown. Whenever there was any question about whether the cultivars showing the greatest percent similarity were the same as the unknown, it was sufficient to compare the densitometer scans of the unknown with those of the standard cultivars most similar to the unknown. Comparison of densitometric traces of gels allows one to observe subtleties that cannot be programmed into the computer, and facilitates identification. As a last resort, it may sometimes be necessary to electrophorese samples of those cultivars that the computer shows are most similar to the unknown, together with the unknown on a single gel. In that way, the slight differences found between gels made on different days can be eliminated. In general, if one cultivar is more than 5% more similar to the unknown than any other cultivar, it is the same as the unknown.

A complete listing of the gliadin electrophoregrams from the 88 most commonly grown U.S. wheat varieties (of 1979) with their computer-generated mobility-density graphs has been published (Jones et al 1982). The complete computer program (in BASIC) is available from the authors.

LITERATURE CITED

AUTRAN, J. C., and BOURDET, A. 1975. L'identification des variétés de blé: Établissment d'un tableau général de détermination fondé sun le diagramme électrophorétique des gliadins du grain. Ann. Amelior. Plant. 25:277.

BUSHUK, W., SAPIRSTEIN, H. D., and ZILLMAN, R. R. 1978. Wheat cultivar identification by computer analysis of gliadin electrophoregrams. (Abstr.) Cereal Foods World 23:496.

BUSHUK, W., and ZILLMAN, R. R. 1978. Wheat cultivar identification by gliadin electrophoregrams. I. Apparatus, method and nomenclature. Can. J. Plant Sci. 58:505.

DAL BELIN PERUFFO, A., PALLAVICINI, C., VARANINI, Z., and POGNA, N. E. 1981. Analysis of wheat varieties by gliadin electrophoregrams. I. Catalog of electrophoregram formulas of 29 common wheat cultivars grown in Italy. Genet. Agrar. 35:195.

JONES, B. L., LOOKHART, G. L., HALL, S. B., and FINNEY, K. F. 1982. Identification of wheat cultivars by gliadin electrophoresis: Electrophoregrams of the 88 wheat cultivars most commonly grown in the United States in 1979. Cereal Chem. 59:181.

LOOKHART, G. L., JONES, B. L., HALL, S. B., and FINNEY, K. F. 1982. An improved method for standardizing polyacrylamide gel electrophoresis of wheat gliadin proteins. Cereal Chem. 59:178.

WRIGLEY, C. W. 1980. The genetic and chemical significance of varietal differences in gluten composition. Ann. Technol. Agric. 29:213.

WRIGLEY, C. W., AUTRAN, J. C., and BUSHUK, W. 1982. Identification of cereal varieties by gel electrophoresis of grain proteins. Advances in Cereal Science and Technology, Vol. V. Am. Assoc. Cereal Chem., St. Paul, MN.

WRIGLEY, C. W., ROBINSON, P. J., and WILLIAMS, W. T. 1981. Association between electrophoretic patterns of gliadin proteins and quality characteristics of wheat cultivars. J. Sci. Food Agric. 32:433.

WRIGLEY, C. W., and SHEPHARD, K. W. 1974. Identification of Australian wheat cultivars by laboratory procedures: Examination of pure samples of grain. Aust. J. Exp. Agric. Anim. Husb. 14:796.

ZILLMAN, R. R., and BUSHUK, W. 1979. Wheat cultivar identification by gliadin electrophoregrams. III. Catalogue of electrophoregram formulas of Canadian wheat cultivars. Can. J. Plant Sci. 59:287.

[Received June 1, 1982. Accepted September 14, 1982]