Journal of Cereal Science **35** (2002) 29–38
doi:10.1006/jcrs.2001.0400, available online at http://www.idealibrary.com on IDEAL®

Identification of Waxy Wheat by Near-infrared Reflectance Spectroscopy

S. R. Delwiche* and R. A. Graybosch†

* *USDA-ARS, Beltsville Agricultural Research Center, Beltsville, MD, U.S.A.; and* † University of Nebraska, *USDA-ARS, Department of Agronomy, Lincoln, NE 68583, U.S.A.*

Received 23 January 2001

Abstract

Newly developed low amylose wheats (*Triticum aestivum* L.) have unique processing characteristics, and thus allow millers to blend defined levels of amylose in mixes requiring low-amylose flour. The amount of amylose synthesised during grain fill is dependent on the expression of three structural genes that encode isoforms of granule-bound starch synthase (GBSS). Lines possessing null alleles at the three waxy (*wx*) loci produce starch that lacks amylose. While such wheats are readily identified by iodine staining, their identification in wheat marketing and production systems would be facilitated by the use of rapid, spectral methods. The present study was undertaken to determine the feasibility of using near-infrared (NIR) spectroscopy to identify waxy wheats, and differentiate them from partial waxy and wild-type phenotypes. Nearly 200 lines from each of two harvest years, with a range of zero (waxy) to three (wild-type) active genes, were ground and scanned (1100–2500 nm) in NIR reflectance. Linear or quadratic discriminant functions of the scores from principal component decomposition cross validation demonstrated that within a crop year, near-perfect separation of fully waxy (27 of 27 samples from Year 1, and 23 of 24 samples from Year 2 correctly identified) from non-waxy lines (165 of 165 samples from either Year 1 or Year 2 correctly identified) was achievable. Further classification among the three non-waxy classes was more difficult, with an average overall accuracy of 60%. Misclassifications were most often assignments into neighbouring gene classes (e.g. 1-gene line assigned to the 2-gene class). The method should prove useful in the identification of waxy wheats, or of blends of waxy and non-waxy cultivars.
© 2002 Elsevier Science Ltd

Keywords: wheat, starch, amylose, waxy, classification, near-infrared.

INTRODUCTION

Starch in plant endosperm is synthesised within specialised organelles called amyloplasts. The two macromolecules that comprise starch, namely amylose and amylopectin, are chemically similar, but differ in the degree of branching of the D-glucosyl units, which are the core to these molecules. Amylose, primarily composed of straight chains of α (1→4) linked D-glucosyl units, is synthesised by an enzymatic process that differs from the mechanism responsible for amylopectin formation. The granule-bound starch synthase (GBSS), also known as the waxy protein[1], is thought to be the primary enzyme responsible for amylose synthesis. Because of the allohexaploid nature of common wheat (*Triticum aestivum* L.), there exist three unique loci (*wx-A1*, *wx-B1*, and *wx-D1*) that contain the genes that encode GBSS. Each locus encodes a separate isoform of GBSS. In its native (wild-type) state, common wheat possesses all three isoforms. By natural mutation or conventional breeding practices, some lines may possess a null allele at one locus, while others may carry two or, in the extreme case (very rare), three null alleles. Generally, the greater the number of active GBSS isoforms, the greater will be the amylose content[2–5]. Wheats that carry three null alleles are termed 'waxy', a term borrowed from

* *Corresponding author. E-mail: delwiche@ba.ars.usda.gov.*

 © 2002 Elsevier Science Ltd

the literature on maize. Lines with one or two null alleles are often referred to as 'partial waxy' wheats[2,6].

Waxy wheat breeding programs are currently underway in Australia, Canada, Japan and the United States. Possible uses of waxy wheats include (1) a stock material for blending by wheat millers for the purpose of achieving prescribed processing characteristics, (2) flour for Asian noodlemaking, and (3) a substitute for waxy maize starch for industries such as modified food starches, papermaking and adhesives[7]. Conventional methods, such as SDS-PAGE for separation of GBSS isoforms as a means to identify waxy and partial waxy lines[6], are expensive, difficult and time consuming, and therefore not readily amenable to either wheat breeding programs or to various stages of the wheat marketing and production system. Enzyme-linked immunosorbent assay (ELISA) tests for GBSS are possible[5,8]; however, the samples still require a preliminary starch extraction procedure.

Still another alternative to the assessment of waxyness is the measurement of the actual concentration of amylose. Measurement of amylose content in wheat starch is conventionally performed by reliance on the reaction between amylose and iodine to form a blue complex. By means of standard curves from substances of known amylose content (e.g. potato starch amylose), amylose content is then determined either by colorimetry or by potentiometric titration. The greatest disadvantage to this procedure is the length of time needed for the formation of the amylose-iodine complex. Even with the introduction of $CaCl_2$ in a starch-iodine-dimethyl sulfoxide solution to permit low-temperature (65–70 °C) gelatinisation, and sonication to enhance solubilisation of the gelatinised starch, analysis time is 30–60 min per sample[9]. Furthermore, the precision of this procedure, as measured by the standard deviation of repeated assays, is typically 2–4% for maize samples of normal (<25% amylose) or high (up to 70%) amylose content[9]. Assuming a similar level of precision for wheat, such an error would make it difficult to successfully develop a secondary procedure, such as near-infrared (NIR) reflectance spectroscopy, for amylose content determination. Because commercial wheat samples typically have a narrower range of amylose content, such as 20–34%[5], quantifying the proportion of amylose in starch by current chemical methods might be less possible than categorising samples into levels corresponding to the number of active GBSS genes. Therefore, the objectives of this study were three-fold: (1) to examine the feasibility of using NIR reflectance spectroscopy to identify waxy wheat samples, (2) to determine whether NIR could differentiate wheat samples based upon the number of active GBSS genes, and (3) to understand the chemical basis for the NIR classification models. Of particular interest is the relationship between amylose content and gene class, and whether this relationship is the sole basis of the NIR classification model. A successful NIR-based classification procedure has immediate application in both plant breeding and wheat marketing and production.

EXPERIMENTAL

Wheat

Year 1

Wild-type and partial waxy samples were drawn from breeders' advanced lines harvested in 1998 in field plots at Lincoln and Sidney, Nebraska, U.S.A., as a component of a USDA-ARS wheat breeding program. Samples included partial waxy and wild-type wheats derived from the following crosses: MT8713/NE87612; NE90476/Ike; NE90616/Ike; and SD88137/Ike. In addition, samples of the cultivars 'Redland' (wild-type), 'TAM202' (*wx-B1* null) and 'Ike' (*wx-A1* null + *wx-B1* null) also were included. Partial waxy and wild-type lines were identified in the F_3 generation by snapping heads, determining GBSS status by SDS-PAGE on a portion of the grains from each head[5], and advancing the remaining portion to F_4. Waxy wheat lines were derived from the cross Kanto107/BaiHuo. The waxy lines were derived from single grains of F_2 seed (produced by greenhouse-grown plants) that were visually identified for the waxy trait and checked later with iodine solution. Because of poor adaptation to the Nebraska climate, waxy samples were grown during 1998 in southern California, outside Brawley, CA, U.S.A. Altogether, the samples ($n = 192$, ~ 10 g each) possessed between 0 (waxy) and 3 (wild-type) active GBSS genes. The number of samples assigned to each gene class was as follows: 27 gene class 0 (triple null), 53 gene class 1 (*wx-A1* null + *wx-B1* null), 73 gene class 2 (*wx-A1* null or *wx-B1* null) and 39 gene class 3 (wild-type).

Year 2

The samples for the second year were the progeny of the first. Three fewer waxy samples were available for NIR analysis, making a total of 189 samples for Year 2. All samples were grown in field plots in southern California.

Methodology

Wheat samples were held at room ambient condition for several days before grinding. Each sample was separately ground on a laboratory scale cyclone grinder (Udy Corp., Ft. Collins, CO, U.S.A.). Duplicate successive spectra (1100–2498 nm, 2-nm wavelength resolution, 32 scans/spectrum) were made on separate fillings of a standard ring cell loaded into a reflectance spectrophotometer (Model 6500, Foss-NIR Systems, Inc., Silver Spring, MD, U.S.A.) equipped with a rotating sample attachment. Reflected energy was referenced to corresponding readings (collected before each sample) from a ceramic disk, and stored as $\log_{10}(1/R)$. Prior to classification modelling, the duplicate spectra of each sample were averaged.

Apparent amylose content of Year 1 samples was measured by iodine-binding blue complex colorimetry, using slight modifications to the method of Knutson and Grove[9]. To enhance precision of diluting the sample for complex formation, the DMSO-iodine reagent volume was adjusted by using a positive displacement pipettor. A standard calibration curve was developed using serial dilutions of a crude wheat amylose extract obtained by the procedure of Klucinec and Thompson[10]. All samples were measured in duplicate, with the second set of assays completed approximately 1 month after the first set. The repeatability of the chemical procedure was gauged by analysis of a control sample inserted approximately every eight sample assays.

Classification modelling

SAS (SAS Institute Inc., Cary, NC, U.S.A.) procedures PRINCOMP, STEPDISC, and DISCRIM[11] were sequentially used for (1) reducing the spectral data to 15 principal components (PCs), (2) identifying by stepwise search the components that were most important in separating the gene classes, and (3) determining by cross-validation the optimal number of components to use in a discriminant function for gene classification. All samples from a given year were used in one set for these three procedures. With the assumption that PCs were multivariate normally distributed in each gene class, and with the pooling of the covariance matrix across the gene classes, a linear discriminant function was used during the third procedure. Alternatively, a quadratic discriminant function also was used, based on no pooling of the covariance matrices.

The novelty of this procedure compared to previous research using principal component analysis (PCA) is that the components used in a classification model were not restricted to be in sequential order (i.e. PC 1, PC 2, . . ., PC 15). Rather, the optimum discriminant function was determined by a one-sample-out cross validation procedure, whereupon the number and order of the components were those which resulted in the highest rate of correctly classified samples, averaged over the four gene classes.

RESULTS AND DISCUSSION

With the exception of baseline variation and a multiplicative effect caused by differences in particle size and packing density, the spectra of all samples were remarkably similar, as shown for Year 1 in Figure 1. To accentuate any chemically based spectral differences caused by gene class, second derivatives (calculated as the second central difference with a gap size of 10 nm) were determined for all waxy and wild-type samples (gene classes = 0 and 3, respectively). In the region of greatest non-water spectral differences (1950–2350 nm), the mean second derivative spectrum for each of these classes, along with a one-sided standard deviation envelope (minus side for the wild-type mean, plus side for the waxy mean), indicated that spectral variation within a class was often as large as that between classes (Fig. 2).

Upon reduction of the spectral data by PCA, a stepwise search of the most important factors, in which up to 10 PCs were permitted, resulted in the selection of PC 1 as most useful in Year 1, as well as in Year 2 (Tables I and II, respectively). With the first PC alone, the overall classification rate was nearly 50%, regardless of year or discriminant function (i.e. linear, quadratic). Interestingly, the second most important factor (PC 6 for Year 1, PC 8 for Year 2) did not coincide with that associated with the next greatest in spectral variation. Based on band assignments in

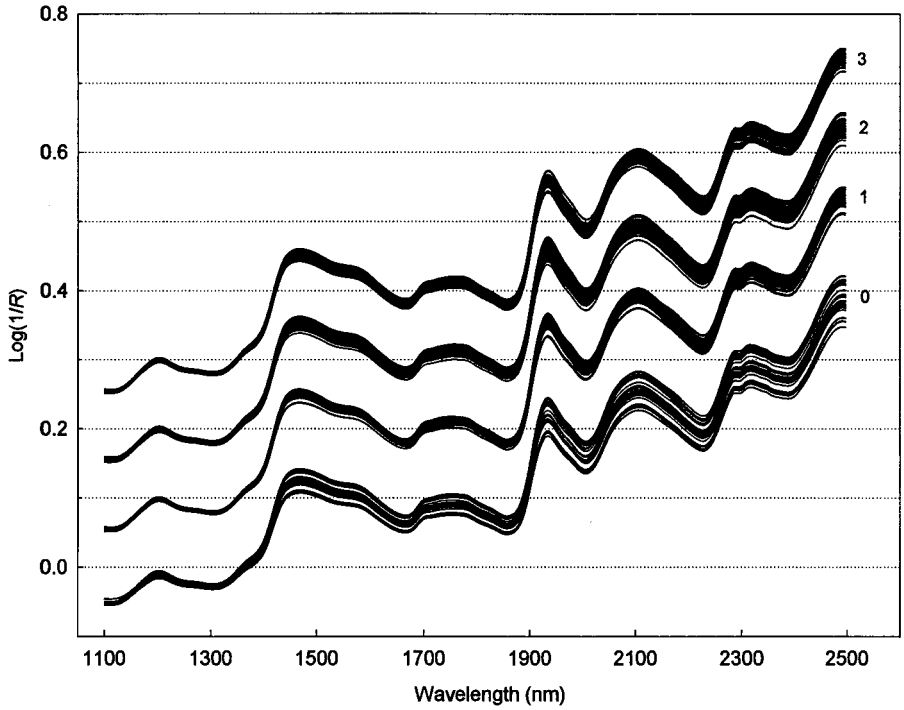

Figure 1 Raw log $(1/R)$ spectra of all Year 1 samples. Samples have been grouped according to the number of active granule-bound starch synthase (GBSS) genes (0–3) by displacing the spectra 0·1 units upward per number of active genes. For active gene number 0, 1, 2, and 3, $n=27$, 53, 73, and 39 samples, respectively.

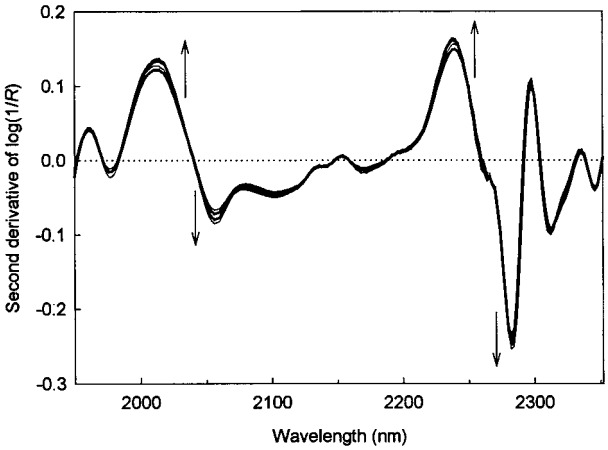

Figure 2 Mean second derivative spectra of Year 1 GBSS gene classes 0 and 3, limited to the spectral region (1950–2350 nm) that showed the greatest variation with gene class. Arrows point in the direction of increasing number of genes. Also included are mean $+1\sigma$ and mean -1σ traces for groups 0 and 3, respectively.

the literature[12], starch, protein, or their interaction were prominent in PC 6 for Year 1 and PC 8 for Year 2, especially near 1460 nm (N-H stretch first overtone), 2050 nm (N-H stretch + amide II), 2130 nm (N-H stretch + C = O stretch), 2260 nm (O-H stretch + O-H deformation), and 2382 nm (O-H deformation second overtone). With two principal components and a linear discriminant function, only one of the 27 waxy samples from Year 1 could not be correctly classified. Similarly, two waxy samples from Year 2 were incorrectly classified. As more principal components were added, the performance of the classification models generally improved through the addition of the seventh PC. At this number of PCs or greater, the average percentage of correctly classified samples ranged between 64 and 72, with little difference between the choice of discriminant function or year. By inspection, models using five to seven PCs were judged to offer the best compromise between accuracy and complexity of single-year models. With seven PCs, the actual numbers of correctly and incorrectly assigned samples by gene class are shown in Tables III and IV for Years 1 and 2, respectively. For waxy wheat, no more than two of the 27 samples from Year 1, and two of the 24 samples from Year 2, were misclassified. In each case, waxy misclassification occurred as an assignment into the neighbouring gene class.

Table I Average percentage of correctly classified samples by cross validation, based on linear or quadratic discriminant function applied to principal component analysis scores

Year 1

Number of principal components	Principal components selected[b]	Average percentage of correctly classified samples[a]	
		Linear discriminant	Quadratic discriminant
1	1	42·3	51·6
2	1, 6	59·2	57·1
3	1, 6, 8	58·8	54·1
4	1, 6, 8, 4	59·7	56·4
5	1, 6, 8, 4, 2	56·8	59·4
6	1, 6, 8, 4, 2, 10	65·1	64·0
7	1, 6, 8, 4, 2, 10, 7	68·0	66·0
8	1, 6, 8, 4, 2, 10, 7, 12	67·9	65·8
9	1, 6, 8, 4, 2, 10, 7, 12, 9	71·5	65·6
10	1, 6, 8, 4, 2, 10, 7, 12, 9, 11	71·2	63·3

[a] Value determined as the average of the percentage of correctly classified samples within each of the four active starch synthase categories by leave-one-out cross validation.
[b] Stepwise selection used to determine best combination of principal components.

Table II Average percentage of correctly classified samples by cross validation, based on linear or quadratic discriminant function applied to principal component analysis scores

Year 2

Number of principal components	Principal components selected[b]	Average percentage of correctly classified samples[a]	
		Linear discriminant	Quadratic discriminant
1	1	47·2	46·5
2	1, 8	63·7	59·3
3	1, 8, 5	67·6	66·0
4	1, 8, 5, 12	67·4	65·6
5	1, 8, 5, 12, 6	70·6	68·1
6	1, 8, 5, 12, 6, 9	69·6	71·5
7	1, 8, 5, 12, 6, 9, 13	70·4	72·2
8	1, 8, 5, 12, 6, 9, 13, 11	69·0	71·6
9	1, 8, 5, 12, 6, 9, 13, 11, 14	68·5	70·0
10	1, 8, 5, 12, 6, 9, 13, 11, 14, 10	68·2	71·7

[a] Value determined as the average of the percentage of correctly classified samples within each of the four active starch synthase categories by leave-one-out cross validation.
[b] Stepwise selection used to determine best combination of principal components.

For the partial waxy (gene classes 1 and 2) and wild-type wheats (gene class 3), misclassification occurred with greater frequency than misclassification of waxy samples. Misclassification usually represented assignment into the neighbouring partial waxy or wild-type wheats, with favour given to the wild-type wheats (class 3) in the case of the gene class 2 samples. Samples possessing two active genes were most difficult to classify.

A scores plot for Year 1 PCs 1 and 6 [Fig. 3(a)] and Year 2 PCs 1 and 8 [Fig. 3(b)] demonstrates that spectral variation within the waxy samples was at least as large as that variation among the three other gene classes combined. The patterns of the two scores plots are remarkably similar. The samples tended to plot in the same location from the first year to the second, suggesting that the models were responding to an effect caused by genotype more so than that caused by environment, especially considering the difference in geographical origin (Nebraska vs California) of the

Table III Cross validation of samples, using seven-factor principal component models with linear or quadratic discriminant function

Year 1 ($n = 192$ total)

Actual gene class	Discriminant function[b]	Assigned gene class[a] 0	1	2	3
0	L	**27**	0	0	0
	Q	**25**	2	0	0
1	L	0	**40**	11	2
	Q	0	**39**	10	4
2	L	0	15	**31**	27
	Q	0	15	**34**	24
3	L	0	2	16	**21**
	Q	0	3	16	**20**

[a] Gene class refers to the number of active genes (0 = triple null allele, . . ., 3 = wild-type) that encode granule-bound starch synthase. Diagonal values (in bold) represent correct assignments.
[b] L = linear, Q = quadratic.

Table IV Cross validation of samples, using seven-factor principal component models with linear or quadratic discriminant function

Year 2 ($n = 189$ total)

Actual gene class	Discriminant function[b]	Assigned gene class[a] 0	1	2	3
0	L	**22**	2	0	0
	Q	**23**	1	0	0
1	L	0	**44**	8	1
	Q	0	**42**	9	2
2	L	0	13	**35**	25
	Q	0	12	**38**	23
3	L	0	3	13	**23**
	Q	0	1	14	**24**

[a] See footnote to Table III.
[b] L = linear, Q = quadratic.

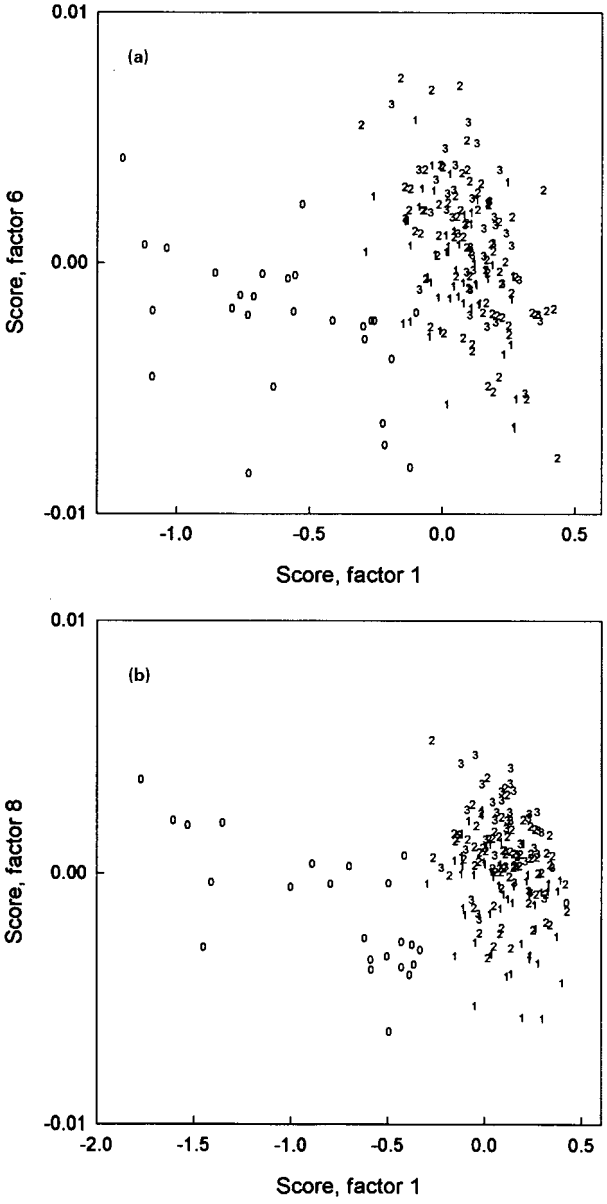

Figure 3 Sample scores of the two most significant principal components for classification of the number of active GBSS genes. Plot symbols (0, 1, 2, 3) represent GBSS gene classes. (a) = Year 1 (factors 1 and 6), (b) = Year 2 (factors 1 and 8).

non-waxy samples for Years 1 and 2. However, this is not to say that Year 2 spectra exactly resembled Year 1 spectra. A plot of the average spectrum for each year reveals that the second year's spectra were additively and multiplicatively offset from the first year's spectra (Fig. 4). The reason for this offset was most likely a difference in moisture contents between the two years' samples, which tends to cause a non-linear scatter effect[13].

To better estimate the performance of a classification model, the Year 2 samples were used as a test set to which a model developed with Year 1 spectra was applied. Misclassifications were least when the five-factor model with linear discriminant function was used. Table V displays the test results of applying a five-factor model, developed from Year 1 spectra, to the Year 2 test set. As a spectral pre-treatment to minimise overall yearly differ-

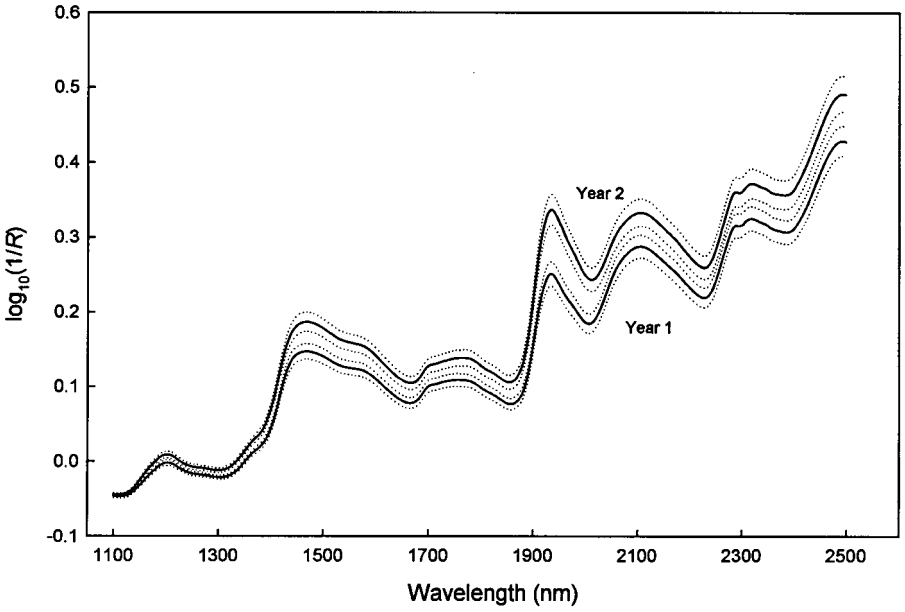

Figure 4 Mean spectrum (solid line) for all samples in each year, with ±one s.D. envelope (dotted line).

ences in the spectra, all spectra were multiplicatively scatter corrected[14] to the mean spectrum of the Year 1 samples. The effect of the scatter correction is displayed in Figure 5, whereby most of the non-chemically based spectral differences between the two years is removed, as seen by the overall similarity between the Year 1 mean spectrum and the mean of the multiplicatively scatter corrected Year 2 spectra. The major spectral difference that remains lies in the 1900–2000 nm water band region. Although the overall accuracy (62·2% correct on average) is slightly less than those of the cross validations within each year, the pattern for misclassification was similar, this being that misclassified samples were most often assigned to the neighbouring gene class. Only one out of 165 non-waxy samples was erroneously classified as waxy; however, six of the 24 waxy samples were misclassified.

In general, the possible reasons for the ability of NIR to classify wheat by the number of active GBSS genes include spectral sensitivities to particle size variation, amylose vs amylopectin concentration, and GBSS concentration. We suspect that the last possibility, GBSS concentration, is unlikely because of its low level of abundance compared to the endosperm storage proteins present within the ground sample and to the amylose and amylopectin concentrations. In an effort to measure protein content of starch granules from different GBSS genotypes, no statistical differences attributable to the GBSS group effect were noticed, with each being approximately 0·2% protein (Graybosch, unpublished). Therefore, focus was given to the two other possibilities.

Table V Test set results. Calibration model: linear discriminant function applied to five-factor principal components of multiplicative scatter corrected Year 1 spectra ($n = 192$)

Test set: Year 2 samples ($n = 189$)

	Assigned gene class[a]			
Actual gene class	0	1	2	3
0	**18**	5	0	1
1	1	**46**	5	1
2	0	22	**23**	28
3	0	10	9	**20**

[a] See footnote to Table III.

Short of actual physical measurement of the ground wheat particles, the particle size effect can be investigated by developing an NIR model that has incorporated a spectral pre-treatment to reduce the effect of particle size variation (such as multiplicative scatter correction), then comparing the classification results of this new model to those of the non-scatter-corrected model. One could reason that the particle size distribution has a large

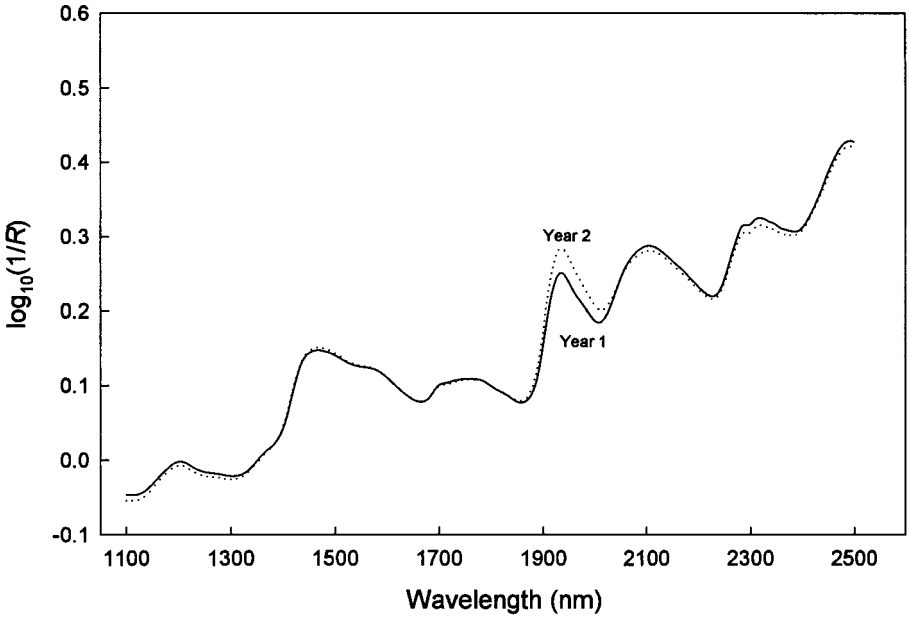

Figure 5 Mean spectrum for all samples in Year 1 and mean spectrum for all samples in Year 2 spectra that were multiplicatively scatter-corrected to the Year 1 mean.

effect on GBSS classification provided that the scatter-corrected NIR model results in a poorer grouping of the samples. Conversely, if little difference in classification ability occurs between the two models, the particle size effect would likely not be the primary basis for GBSS classification by NIR. When this experiment was performed on Year 1 samples, the results for the scatter-corrected model (average overall cross validation accuracy ranging from 47·2 to 72·6%, depending on the number of PCs, data not shown) were equivalent to the non-corrected model (42·3 to 71·2% range, Table I). Hence, it seems likely that GBSS classification by NIR is most likely based on a spectral sensitivity to amylose, amylopectin, or their interaction with other chemical constituents. Starch lipid content also might have contributed to NIR classification. Yasui *et al.*[15] found starch lipid content of waxy wheats derived from Kanto107 and BaiHuo to be significantly less than their non-waxy parents. Recently, the same researchers[16] found total grain starch to be higher, and total grain fat and β-glucan content to be elevated, in two waxy mutants derived from Kanto107. In the latter case, however, the waxy lines were induced by mutagenesis, so it is possible that the observed changes in grain biochemical components might have been derived from additional mutations. Although not evaluated in the current study, β-glucan content is to a certain extent measurable by NIR reflectance spectroscopy[17], particularly at the longer (>2200 nm) wavelengths[18].

If it is assumed that the spectral basis of the classification models is primarily associated with the level of amylose, the ultimate accuracy of such models is limited to the extent of the relationship between GBSS gene class and amylose content. The actual relationship, as shown for Year 1 samples in Figure 6, is such that only the waxy wheats had amylose contents distinctly different than the other gene classes. Some samples assigned to the gene class 0 (waxy) had amylose contents ranging from 3–9%. These samples could have encountered a low level of outcrossing in field plots, been the result of mechanical mixtures during harvest operations, or might represent 'leaky' backgrounds in which either some limited amylose synthesis occurs or some long-chain amylopectin molecules bind iodine. Indeed visual inspection of waxy seed stained with iodine demonstrated some purplish foci, generally in the area around the crease in the grain. The variation in amylose content within each of the other three classes, with a standard deviation ranging from 1·75% (gene class 2) to 2·19% (gene class 3) was sufficiently large to cause overlap across these classes, considering the narrow range of class means (18·7 to 22·0%). This overlap is caused by a dosage

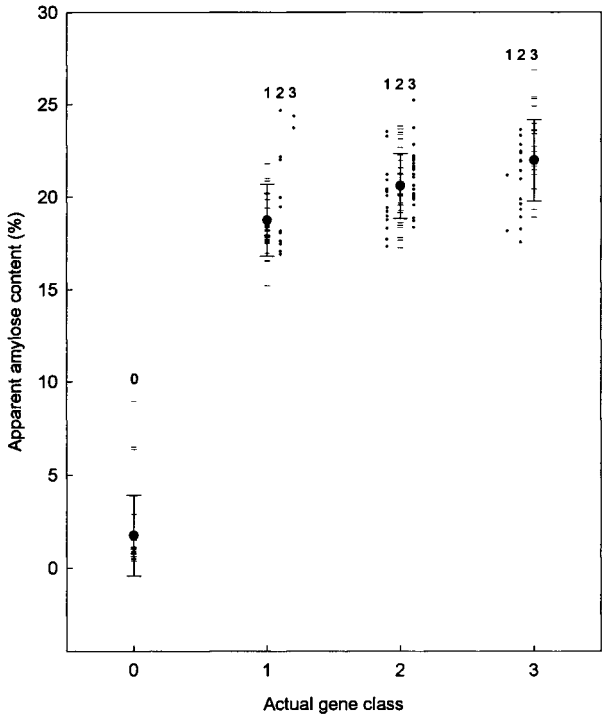

Figure 6 Relationship between amylose content and gene class for Year 1, shown as gene class means with error bars. Also included are the cross validation predictions of the seven factor with linear discriminant function model (summarised in Table III), with GBSS gene class (0, 1, 2, or 3) labelled above each cluster. Mean ± S.D. (●); misclassified individual sample (•); correctly classified individual sample (–).

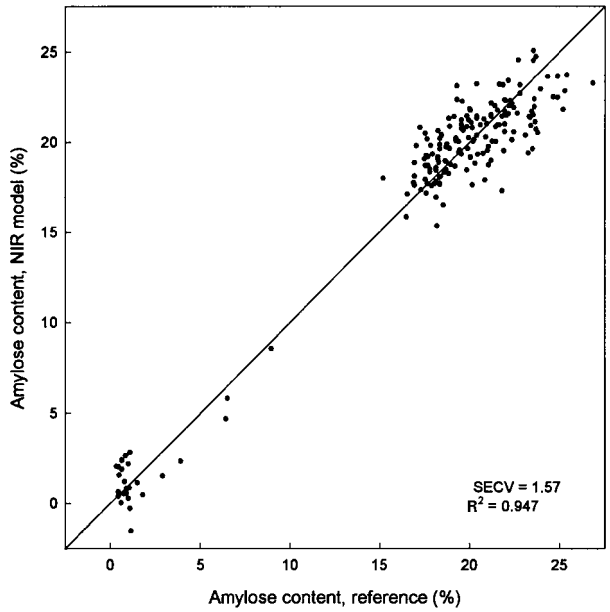

Figure 7 Cross validation predictions for a partial least squares 12-factor NIR model.

compensation response that occurs when one or more null alleles for GBSS are present, resulting in compensation by the remaining active genes to elevate their production of GBSS[2,4]. Recent research has demonstrated differential effects among the various active waxy alleles on amylose content[19,20]. Amongst the three possible double null classes, and within defined environments, some significant differences were observed. However, across environments, amylose contents of these genetic classes overlapped. In addition the differences were so slight that the changes of NIR differentiating the three possible single active GBSS genotypes seems remote.

Cross validation assignments from the seven-factor linear discriminant function model (Table III) are also displayed in Figure 6 for the purpose of identifying trends in misclassification. These trends are identified by examining an individual sample's amylose content with respect to the distribution of amylose contents within the sample's gene class. Often a misclassified sample had an amylose content that was skewed in the direction toward the gene class to which the sample was erroneously assigned. When discriminant analysis was based on the chemically determined value for amylose content, the overall average classification accuracy was 70·0%, which is marginally higher than the 68·0% level for the discriminant analysis based on seven PCs (Table III). Therefore, with the assumption that all power of an NIR discriminant model for gene class is primarily dependent on amylose content, further improvements in model accuracy may only be marginal.

Repeatability error of the reference chemical analysis procedure for amylose content, defined as the standard deviation of the amylose content readings on a control sample (measured 53 times), was 2·48%. The reason this value is greater than the standard deviations of amylose content within the gene classes is most likely because the latter values were calculated using averages from duplicate assays. Despite the large repeatability error of the reference procedure, a 12-factor partial least squares calibration for amylose content produced a standard error of cross validation of 1·57% (Fig. 7). This error is slightly higher than typical standard errors of performance of about 1·0% that have been reported for rice amylose content NIR models[21,22]. Additional research is needed to identify and quantify the degree of spectrally sensed features (from lipid, β-glucan, and protein

components) beyond that of amylose content, that might contribute to an NIR classification model for gene class.

CONCLUSIONS

NIR reflectance spectroscopy can be used to discriminate waxy wheat from partial-waxy (single or double null alleles) or wild-type wheat with near-perfect accuracy. Thus, waxy wheat samples can be easily identified both in breeding programs, and at various stages of the wheat production and marketing system, by instrumentation that already is used to measure protein content and grain hardness. Using the best NIR model, partial waxy and wild-type lines can be grouped according to the correct number of active GBSS genes at approximately 60% accuracy. The ability of NIR-based models to classify wheat lines according to the number of active GBSS genes appears to be due to spectrally sensed differences in amylose and amylopectin. However, reliance on amylose vs amylopectin concentrations prevents NIR models from achieving perfect GBSS classification accuracy of non-waxy lines because of overlap in amylose content between these gene classes.

Acknowledgements

J. D. Shaffer, USDA-ARS, Beltsville collected the spectral data. V. E. Hansen, USDA-ARS, Lincoln, Nebraska, U.S.A. performed all assays of amylose content. Research was funded in part by a grant from the Nebraska Wheat Board to R. A. Graybosch.

REFERENCES

1. Sivak, M.N. and Preiss, J. Starch synthesis in foods. In 'Seed Development and Germination', (J. Kigel and G. Galili, eds), Marcel Dekker, New York (1995) pp. 139–168.
2. Nakamura, T., Yamamori, M., Hidaka, S. and Hoshino, T. Decrease of waxy (Wx) protein in two common wheat cultivars with low amylose content. *Plant Breeding* **111** (1993) 99–105.
3. Yamamori, M., Nakamura, T., Endo, T.R. and Nagamine, T. Waxy protein deficiency and chromosomal location of coding genes in common wheat. *Theoretical and Applied Genetics* **89** (1994) 179–184.
4. Yamamori, M., Nakamura, T. and Kuroda, A. Variation in the content of starch-granule bound (Wx) protein among several Japanese cultivars of common wheat (*Triticum aestivum* L.). *Euphytica* **64** (1992) 215–219.
5. Graybosch, R.A., Peterson, C.J., Hansen, L.E., Rahman, S., Hill, A. and Skerritt, J.H. Identification and characterization of U.S. wheats carrying null alleles at the *wx* Loci. *Cereal Chemistry* **75** (1998) 162–165.
6. Nakamura, T., Yamamori, M., Hidaka, S. and Hoshino, T. Expression of HMW Wx protein in Japanese common wheat (*Triticum aestivum* L.) cultivars. *Japanese Journal of Breeding* **42** (1992) 681–685.
7. Graybosch, R.A. Waxy wheats: origin, properties, and prospects. *Trends in Food Science and Technology* **9** (1998) 135–142.
8. Rahman, S., Kosarhashemi, B., Samuel, M.S., Hill, A., Abbott, D.C., Skerritt, J.H., Preiss, J., Appels, R. and Morell, M.K. The major proteins of wheat endosperm starch granules. *Australian Journal of Plant Physiology* **22** (1995) 793–803.
9. Knutson, C.A. and Grove, M.J. Rapid method for estimation of amylose in maize starches. *Cereal Chemistry* **71** (1994) 469–471.
10. Klucinec, J.D. and Thompson, D.B. Fractionation of high-amylose maize starches by differential alcohol precipitation and chromatography of the fractions. *Cereal Chemistry* **75** (1998) 887–896.
11. SAS Institute Inc. The DISCRIM, PRINCOMP, and STEPDISC Procedures. In 'SAS/STAT User's Guide', Release 6.03 Edition, SAS Institute Inc., Cary, NC (1988) pp. 359–447, 751–771, 909–922.
12. Osborne, B.G. and Fearn, T. 'Near Infrared Spectroscopy in Food Analysis', Wiley & Sons, New York (1986) pp. 20–42.
13. Birth, G.S. and Hecht, H.G. The physics of near-infrared reflectance. In 'Near-Infrared Technology in the Agricultural and Food Indistries', (P.C. Williams and K.H. Norris, eds), American Association of Cereal Chemists, Inc., St Paul, MN (1987) pp. 1–15.
14. Martens, H. and Næs, T. 'Multivariate Calibration', Wiley, Chichester (1989) pp. 345–350.
15. Yasui, T., Matsuki, J., Sasaki, T. and Yamamori, M. Amylose and lipid contents, amylopectin structure and gelatinisation properties of waxy wheat (*Triticum aestivum*) starch. *Journal of Cereal Science* **24** (1996) 131–137.
16. Yasui, T., Sasaki, T. and Matsuki, J. Milling and flour pasting properties of waxy endosperm mutant lines of bread wheat (*Triticum aestivum* L.). *Journal of the Science of Food and Agriculture* **79** (1999) 687–692.
17. Allison, M.J., Cowe, I.A. and McHale, R. The use of infra red reflectance for the rapid estimation of the soluble β-glucan content of barley. *Journal of the Institute of Brewing* **84** (1978) 153–155.
18. Szczodrak, J., Czuchajowska, Z. and Pomeranz, Y. Characterization and estimation of barley polysaccharides by near-infrared spectroscopy. II. Estimation of total β-D-glucans. *Cereal Chemistry* **69** (1992) 419–423.
19. Miura, H., Araki, E. and Tarui, S. Amylose synthesis capacity of the three *Wx* genes of wheat cv Chinese Spring. *Euphytica* **108** (1999) 91–95.
20. Yamamori, M. and Quynh, N.T. Differential effects of Wx-A1, -B1 and -D1 protein deficiencies on apparent amylose content and starch pasting properties in common wheat. *Theoretical and Applied Genetics* **100** (2000) 32–38.
21. Villareal, C.P., de la Cruz, N.M. and Juliano, B.O. Rice amylose analysis by near-infrared transmittance spectroscopy. *Cereal Chemistry* **71** (1994) 292–296.
22. Delwiche, S.R., Bean, M.M., Miller, R.E., Webb, B.D. and Williams, P.C. Apparent amylose of milled rice by near-infrared reflectance spectrophotometry. *Cereal Chemistry* **72** (1995) 182–187.