

Predicting the Number of Dominant R Alleles in Single Wheat Kernels Using Visible and Near-Infrared Reflectance Spectra

D. Wang,¹ F. E. Dowell,^{1,2} and R. E. Lacey³

ABSTRACT

Cereal Chem. 76(1):6-8

An optical radiation measurement system was used to measure reflectance spectra of single wheat kernels from 400 to 2,000 nm. A total of 18 wheat samples with 0, 1, 2, or 3 R alleles for red grain color was used for this study. The results indicated a linear relationship between the degree of the red pigmentation and the number of R alleles. The highest coefficient of determination (R^2) was 0.78 in the wavelength region of 500–1,700 nm when a four-class partial least squares model was used. The high-

est classification of red genes was 78.4%. For two-class models, differentiating samples with 0 R alleles and 1 R allele from samples with 3 R alleles had the highest success rate of 100 and 98.8%, respectively. The number and combination of R alleles had a significant effect on wheat kernel color. These relationships may be useful to wheat breeders in estimating the number and location of R alleles.

Wheat is classed as either red or white, depending on the color of the seed coat. These two basic colors, as well as certain variations within each color, are commonly considered in the classification of wheat for grading purposes (Evers and Bechtel 1988). Some reported advantages of white wheat over red wheat are higher flour extraction rate, greater aesthetic appeal of whole wheat products, more valuable bran, and better scoring on the basis of flour standards (Paulsen and Heyne 1981). One disadvantage of white wheat is its susceptibility to preharvest sprouting, whereas most red wheats are resistant (Moss et al 1972). Some research indicated a positive relationship between red kernel coat color and resistance to sprouting (Gfeller and Svejda 1960, Khan and Strand 1977, DePauw and McCaig 1983).

Mature wheat kernels vary from light buff or yellow to red-brown according to the absence or presence of red pigmentation in the seed coat and environmental conditions. The red seed coat in hexaploid wheat is controlled by separated genetic loci (R-A1, R-B1, and R-D1), and thus, color can vary among red cultivars (Miyamoto and Everson 1958, Metzger and Silbaugh 1970, Freed et al 1976). Recessive alleles at three nonlinked loci are needed to produce a colorless seed coat (white seeds) and red color is dominant to white. White-grained *Triticum aestivum* and amber-grained *T. durum* cultivars carry recessive white alleles at each locus (McIntosh and Cusick 1987). The amber color of some *T. durum* wheats is due to the pigments in the endosperm showing through clear seed coat and fruit coats and not to seed coat pigmentation.

Baker (1981) studied wheat seed coat color by visual methods. The differences in kernel color were enhanced by soaking in 5% sodium hydroxide (NaOH). Baker reported that three dominant R alleles controlling seed coat color act in an additive way, such that additional R alleles result in intensification of the red color. DePauw and McCaig (1983) reported that the intensity of the color of kernels soaked in NaOH appeared to be determined by the number of R alleles rather than by environmental factors. Flintham (1993) reported that the de-

gree of red pigmentation increased with the number of R alleles (1 to 3). This gives the possibility of transferring 1 R allele to white wheat to enhance sprout resistance while having a minimal effect on flour color. A nondestructive and objective method of determining the number of R alleles in wheat would benefit wheat breeders and grain inspectors. Near-infrared (NIR) reflectance spectroscopy is a method that meets this requirement. The main advantages of NIR reflectance spectroscopy technology are that sample preparation is simple, analysis is fast and precise, and many constituents can be analyzed simultaneously. The objectives of this study were to: 1) investigate the feasibility of the classification of the number of R alleles in wheat based on visible and NIR reflectance spectra of single wheat kernels; and 2) investigate the effect of the number of R alleles and the combination of R alleles on wheat kernel color.

MATERIALS AND METHODS

A total of 18 wheat samples with known 0, 1, 2, and 3 R alleles provided by the USDA-ARS National Small Grain Collection Laboratory and the USDA-ARS Wheat Quality Laboratory were used in this study (Table I). Because the number of kernels for each sample was limited, each wheat sample was represented by 10 kernels. All samples were stored at ambient temperature in air-tight containers.

Single wheat kernel reflectance spectra from 400 to 2,000 nm at 2-nm intervals were collected with an optical radiation measurement system (Oriel Corporation, Stratford, CT). The optical radiation measurement system consisted of a radiation source system with a 100W quartz tungsten halogen lamp (model 6333), a control system (RS-232 Merlin, model 70100), a chopper (model 75152), a monochromator (model 77250 1/8-m), a lead sulfide detector (model 70131), and software described by Wang et al (*in press*).

A single wheat kernel was suspended horizontally from the germ end by a vacuum tube and the crease side of the kernel was viewed. The wavelengths within the 500–1,700 nm range were used because poor sensor sensitivity and low energy levels resulted in excessive noise outside of this range, and the highest classification accuracy was obtained in this region. The transformation of reflectance spectra to $L^*a^*b^*$ color space was based on the Commission International de l'Eclairage (CIE 1931) weighted-ordinate integration method and the color difference formula (Hunter and Harold 1987). The standard observer functions (x, y, z) (CIE 1931) were used to convert the spectral curves into tristimulus values (X, Y, Z) that properly identify the color of an object in terms of a mixture of primary colors that visually match. In the $L^*a^*b^*$ color space, L^* varies from 0 (black) to 100 (perfect white); a^* ranges from –100 to 100 and measures green when negative and red when positive;

¹ Grain Marketing and Production Research Center, ARS, USDA, Manhattan, KS 66502. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.

² Corresponding author. E-mail: fdowell@usgml.ksu.edu

³ Agricultural Engineering Dept., Texas A&M University, College Station, TX 77843-2117.

and b^* , which also varies from -100 to 100, is a measure of blue when negative and yellow when positive.

All data were analyzed using partial least squares (PLS) regression. PLS is a multivariate data analysis technique designed to handle intercorrelated regressors. PLS can be used to resolve complicated, multivariate system problems by a sequence of simple least-squares regressions and uses the concentration information during decompositional processing. This causes spectra containing higher constituent concentrations to be weighted more heavily than those with low concentrations and places as much concentration information as possible into the first few loading vectors. In actuality, PLS takes advantage of the correlation relationship that already exists between the spectral data and the constituent concentrations. A more complete discussion of PLS is included in Martens and Naes (1990) and Bjørsvik and Martens (1992). The software package Grams/32 (PLS-IQ, Galactic Industries Corp. 1996) was used for PLS analysis. Four-class and two-class PLS models were developed to classify the number of dominate R alleles. For four-class models, the samples with 0, 1, 2, and 3 R alleles were assigned constant values of 1.0, 2.0, 3.0, and 4.0, respectively. A kernel was considered to be correctly categorized if the predicted value lay on the same side of the midpoint as its assigned value. PLS model performance was based on the multiple coefficient of determination (R^2) and standard error of cross validation (SECV) for each model. During performing the diagnostics, only one sample was removed from the training set in each pass. The number of factors used was the minimum required to give the maximum multiple coefficient of determination. Analyses were conducted on the absorbance spectra ($\log 1/R$) and on the first and second derivatives of the absorbance spectra. The calculation of the first derivative and second derivative were based on the Savitisky-Golay method (Savitisky and Golay 1964). A fifth-order polynomial with 25 points was used to calculate the first and second derivatives.

TABLE I
Wheat Samples with Known R Alleles^a

Location of R alleles	Cultivar	n	Accession	Country
1 R allele				
R-A1b R-B1a R-D1a	Red Bobs	10	Citr 6255	Canada
R-A1b R-B1a R-D1a	Diamant II	10	PI 190489	Sweden
R-A1a R-B1b R-D1a	Grana	10	PI 383348	Poland
R-A1a R-B1b R-D1a	Supreme	10	Citr 8026	Canada
R-A1a R-B1a R-D1b	Mardler	10	PI 447427	UK
R-A1a R-B1a R-D1b	Apollo	10	Citr 10075	Canada
2 R alleles				
R-A1b R-B1b R-D1a	Avalon	10	PI 446910	UK
R-A1b R-B1b R-D1a	Bersee	10	PI 168661	France
R-A1a R-B1b R-D1b	Khurvof	10		USA
R-A1a R-B1b R-D1b	Sperber	10	PI 476813	Germany
R-A1b R-B1a R-D1b	Brigand	10	PI 447424	UK
R-A1b R-B1a R-D1b	Bezostaya 1	10	Citr 15158	Russian
3 R alleles				
R-A1b R-B1b R-D1b	Arin	10	Citr 15207	Germany
R-A1b R-B1b R-D1b	Banco	10	PI 260896	Sweden
White wheat				
R-A1a R-B1a R-D1a	Hiller	10		USA
R-A1a R-B1a R-D1a	Rely	10		USA
R-A1a R-B1a R-D1a	SRS#2	10		USA
R-A1a R-B1a R-D1a	Klasic	10		USA

^a At each locus, the white allele is assigned a and the red allele is assigned b .

The correlation between the degree of red pigmentation and the number of dominate R alleles (0, 1, 2, or 3) was tested using four-class PLS classification models. Table II shows the PLS calibration results in the wavelength region of 500–1,700 nm. The highest R^2 was 0.78 and the highest average classification accuracy was 78.4% when the first derivative was used. Results showed that kernels with 0, 1, or 2 R alleles had a higher percentage of correct classification than the kernels with 3 R alleles. This is probably because wheat kernels with 2 and 3 R alleles had similar L^* values, or brightness, even when a^* values were not the same (Table III). The data pretreatment had no significant effect on the classification accuracy.

The correlation between the degree of red pigmentation and the number of R alleles was also tested using two-class PLS classification models (Table IV). Results showed that differentiating wheat samples with 0 R alleles from wheat samples with R alleles was the easiest with the classification accuracies of 99% for 0 R alleles versus 1 R allele and 100% for 0 R alleles versus 2 or 3 R alleles. The R^2 increased as the number of R alleles between the two classes increased. The highest R^2 (0.91) was obtained from classification of 0 R alleles and 3 R alleles when $\log (1/R)$ and the first derivative were used. For the classification between wheat samples with R alleles, differentiating wheat samples with 1 R allele from wheat samples with 3 R alleles was the easiest with the classification accuracies of 98.8% for $\log (1/R)$ and the first derivative, and 96.7% for the second derivatives. Differentiating between 1 R allele and 2 R alleles, and between 2 R alleles and 3 R alleles, was comparatively more difficult. The highest average classification accuracy of 1 R allele compared to 2 R alleles was 87.5%, and the highest average classification accuracy of 2 R alleles compared to 3 R alleles was 89.2%. These results indicate some overlapping in the degree of red pigmentation between 1 and 2 R alleles and between 2 and 3 R alleles. This conclusion is also supported by L^* , a^* and b^* values in the $L^*a^*b^*$ color space in Table III.

The color variations of wheat kernels affected by the number of R alleles and combination of R alleles measured as L^* , a^* , and b^* values are summarized in Table III. White wheat (0 R alleles) had larger

TABLE III
Color Variation of Wheat Kernels with Number of R Alleles and Locations as Measured by $L^*a^*b^*$ Color Values

No. and Location of Red Genes	L^*	a^*	b^*
No. of R alleles ^a			
0 R alleles	52.30 ^{ab}	4.66 ^d	14.82 ^a
1 R alleles	45.35 ^b	6.15 ^c	11.47 ^c
2 R alleles	41.78 ^c	6.98 ^b	11.67 ^c
3 R alleles	41.65 ^c	7.63 ^a	12.44 ^b
Location of 1 R allele			
R-A1b R-B1a R-D1a	43.80 ^c	6.09 ^b	10.59 ^b
R-A1a R-B1b R-D1a	47.24 ^a	7.26 ^a	12.57 ^a
R-A1a R-B1a R-D1b	44.98 ^b	5.66 ^c	11.00 ^b
Location of 2 R alleles			
R-A1b R-B1b R-D1a	42.46 ^a	6.37 ^b	10.91 ^b
R-A1b R-B1a R-D1b	40.01 ^b	7.80 ^a	11.19 ^{ab}
R-A1a R-B1b R-D1b	42.15 ^a	6.93 ^b	11.60 ^a

^a At each locus, the white allele is assigned a and the red allele is assigned b .

^b Values in the same column within number and location data and not followed by the same letter are significantly different at $P < 0.05$.

TABLE II
Classification Accuracy (% correct) of Number of R Alleles and Spectral Parameters Using Four-Class Partial Least Squares (PLS) Models (500–1,700 nm)

Spectral Parameters	No. of Kernels	Correct Classification of R Alleles (%)					R^2	No. of PLS Factors
		0	1	2	3	Avg		
Log (1/R)	180	80.0	83.0	80.0	65.0	77.0	0.78	8
First derivative	180	90.0	87.0	76.6	60.0	78.4	0.77	8
Second derivative	180	80.0	80.0	76.6	75.0	77.9	0.70	11

TABLE IV
Classification Accuracy (% correct) of Number of R alleles and Spectral Parameters Using Two-Class Partial Least Squares (PLS) Models (500–1,700 nm)

Spectral Class Parameters	No. of Kernels	Correct Classification of R Alleles (%)					R ²	No. PLS Factors
		0 R	1 R	2 R	3 R	Avg		
Log (1/R)								
0 vs. 1 R	100	100	98.3			99.0	0.81	6
0 vs. 2 R	100	100		100		100	0.86	6
0 vs. 3 R	60	100			100	100	0.91	6
1 vs. 2 R	120		86.7	88.3		87.5	0.55	9
2 vs. 3 R	80			90.0	75.0	82.5	0.49	8
1 vs. 3 R	80		98.3		100	98.8	0.81	8
First derivative								
0 vs. 1 R	100	100	98.3			99.0	0.83	4
0 vs. 2 R	100	100		100		100	0.89	4
0 vs. 3 R	60	100			100	100	0.91	4
1 vs. 2 R	120		85.0	81.6		83.3	0.49	5
2 vs. 3 R	80			98.3	80.0	89.2	0.50	4
1 vs. 3 R	80		98.3		100	98.8	0.81	7
Second derivative								
0 vs. 1 R	100	100	98.3			99.0	0.84	5
0 vs. 2 R	100	100		100		100	0.81	5
0 vs. 3 R	60	100			100	100	0.86	5
1 vs. 2 R	80		75.0	70.0		72.5	0.23	3
2 vs. 3 R	80			96.7	75.0	85.9	0.51	5
1 vs 3 R			98.3		95.0	96.7	0.76	12

L^* and b^* values and a smaller a^* value than red wheat regardless of the number of R alleles. This indicated that white wheat can be separated from red wheat based on color variation. For red wheats, the a^* value increased as the number of the R alleles increased. This result suggested that the redness of these kernels increased as the number of R alleles increased. The average L^* value of the wheat samples with 1 R allele was higher than the wheat samples with 2 and 3 R alleles. In the $L^*a^*b^*$ color space, if the a^* value remains constant, redness will be reduced as the L^* value increases. This may cause some red wheats with 1 R allele to be difficult to separate visually from white wheat when environmental conditions influence color.

The location of R alleles had a significant effect on wheat color. For example, wheat samples with 1 R allele combined as R-A1a, R-B1b, and R-D1a had larger L^* , a^* and b^* values than the other two wheat samples with R alleles combined as R-A1b; R-B1a and R-D1a; and R-A1a, R-B1a, and R-D1b. Wheat samples with 2 R alleles combined as R-A1b, R-B1a, and R-D1b were redder than the wheat samples with 2 R alleles combined as R-A1b; R-B1b and R-D1a; and R-A1a, R-B1b and R-D1b.

In summary, a linear relationship between the degree of red pigmentation and the number of dominant R alleles was observed. For four-class models, the highest R^2 was 0.78 and the highest classification accuracy was 78.4%. For two-class models, differentiating between wheat samples with 0 R alleles and wheat samples with R alleles was the easiest with average classification accuracies of 99–100%. The classification accuracy between 1 R allele and 3 R alleles was 98.8%. The greatest average classification accuracy of 1 R allele compared to 2 R alleles was 87.5%, and the greatest average classification accuracy of 2 R alleles compared to 3 R alleles was 89.2%. The number of R alleles and the combinations of R alleles had significant effects on wheat kernel color. The redness of wheat kernels increased as the number of R alleles increased. White wheat had larger L^* and b^* values and smaller a^* values than red wheat in the $L^*a^*b^*$ color space. The average L^* value of wheat kernels with 1 R allele was larger than kernels with 2 or 3 R alleles. These relationships and color values may be useful to breeders in estimating the number and location of R alleles and to grain inspectors classifying red and white wheat.

ACKNOWLEDGMENTS

We thank H. E. Bockelman of the USDA-ARS National Small Grains Collection, Aberdeen, ID; and R. G. Sears of the Department of Agronomy, Kansas State University, Manhattan, KS, for providing wheat samples.

LITERATURE CITED

- Baker, R. J. 1981. Inheritance of seed coat color in eight spring wheat cultivars. *Can. J. Plant. Sci.* 61:719-721.
- Björsvik, H.-R., and Martens, H. 1992. Data analysis: Calibration of NIR instruments by PLS regression. Pages 159-180 in: *Handbook of Near-Infrared Analysis*. D. A. Burns and E. W. Ciurczak, eds. Marcel Dekker: New York.
- CIE. 1931. Standard sources for which the Commission Internationale de l'Eclairage in 1931 specified the spectral energy distributions. (In English) Int. Commission on Illumination: Paris.
- Evers, A. D., and Bechtel, D. B. 1988. Microscopic structure of the wheat grain. Pages 47-95 in: *Wheat Chemistry and Technology*. Y. Pomeranz, ed. Am. Assoc. Cereal Chem.: St. Paul, MN.
- DePauw, R. M., and McCaig, T. N. 1983. Recombining dormancy and white seed coat color in a spring wheat cross. *Can. J. Plant. Sci.* 63:581-589.
- Flintham, J. 1993. Grain color and spout-resistance in wheat. *Int. Symp. Pre-Harvest Sprouting Cereals*, 6th. Amer. Assoc. Cereal Chem.: St. Paul, MN.
- Freed, R. D., Everson, E. H., Ringlund, K., and Gullord, M. 1976. Seed coat color in wheat and the relationship to seed dormancy at maturity. *Cereal Res. Comm.* 4:147-149.
- Gfeller, F., and Svejda, F. 1960. Inheritance of post-harvest seed dormancy and kernel colour in spring wheat lines. *Can. J. Plant Sci.* 40:1-6.
- Hunter, R. S., and Harold, R. W. 1987. *Measurements of Appearance*. Hunter Association Laboratory: Reston, VA.
- Khan, F. N., and Strand, E. A. 1977. Investigations into the genetics of kernel color and dormancy in wheat (*Triticum aestivum* L.). *Meldinger Fra Norges Landbrukshogskole* 56:1-12.
- Martens, H., and Naes, T. 1990. Multivariate calibration by data compression. Pages 57-87 in: *Near-Infrared Technology in the Agricultural and Food Industries*. P. C. Williams and H. K. Norris, eds. Am. Assoc. Cereal Chem.: St. Paul, MN.
- McIntosh, R. A., and Cusick, J. E. 1987. Linkage map of hexaploid wheat. Pages 289-297 in: *Wheat and Wheat Improvement*. E. G. Heyne, ed. Am. Soc. Agronomy: Madison, WI.
- Metzger, R. J., and Silbaugh, B. A. 1970. Location of genes for seed coat color in hexaploid wheat, *Triticum aestivum* L. *Crop Sci.* 10:495-496.
- Miyamoto, T., and Everson, E. H. 1958. Biochemical and physiological studies of wheat seed pigmentation. *Agron. J.* 50:733-734.
- Moss, H. J., Derera, N. F., and Balam, L. N. 1972. Effect of pre-harvest rain on germination in the ear and α -amylase activity of Australian wheat. *Aust. J. Agri. Res.* 23:769-777.
- Paulsen, G. M., and Heyne, E. G. 1981. Development of hard white winter wheats for the Great Plains. *Wheat Grower* 4:26-28.
- Savitsky, A., and Golay, M. J. E. 1964. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* 36:1627-1639.
- Wang, D., Dowell, F. E., and Lacey, R. E. *In press*. Single wheat kernel color classification using neural networks. *Trans. ASAE*.

[Received April 28, 1998. Accepted October 14, 1998.]