

High-Throughput Near-Infrared Reflectance Spectroscopy for Predicting Quantitative and Qualitative Composition Phenotypes of Individual Maize Kernels

Gertraud Spielbauer,¹ Paul Armstrong,² John W. Baier,¹ William B. Allen,³ Katina Richardson,⁴ Bo Shen,³ and A. Mark Settles^{1,5}

ABSTRACT

Cereal Chem. 86(5):556–564

Near-infrared reflectance (NIR) spectroscopy can be used for fast and reliable prediction of organic compounds in complex biological samples. We used a recently developed NIR spectroscopy instrument to predict starch, protein, oil, and weight of individual maize (*Zea mays*) seeds. The starch, protein, and oil calibrations have reliability equal or better to bulk grain NIR analyzers. We also show that the instrument can differentiate quantitative and qualitative seed composition mutants from normal siblings without a specific calibration for the constituent affected. The analyzer does not require a specific kernel orientation to predict composition

or to differentiate mutants. The instrument collects a seed weight and a spectrum in 4–6 sec and can collect NIR data alone at a 20-fold faster rate. The spectra are acquired while the kernel falls through a glass tube illuminated with broad spectrum light. These results show significant improvements over prior single-kernel NIR systems, making this instrument a practical tool to collect quantitative seed phenotypes at high throughput. This technology has multiple applications for studying the genetic and physiological influences on seed traits.

Scoring plant phenotypes is a technological bottleneck for functional genomics and systems biology (reviewed in Long et al 2008). Phenotypes can be scored at qualitative or quantitative levels with the quantitative phenotypes being more useful for genome-wide or systems analyses. Consequently, a large number of “omics” technologies have been developed to quantify transcript, protein, and metabolite levels. These technologies generally require chemical extraction of a tissue sample to quantify the compounds of interest. Seed phenotypes pose a particular challenge for “omics” analysis. A primary phenotype of seeds is the accumulation of storage molecules, including starch, storage proteins, and oils. Each of these classes of organic molecules has differing chemical properties and requires a different strategy to analyze. For most species, the chemical analysis of any one class of storage molecule requires one or more seeds, effectively destroying the individual plant under analysis.

Near-infrared reflectance (NIR) spectroscopy is a nondestructive technology that can report the major seed storage molecules simultaneously (reviewed in McClure 2003; Osborne 2006). NIR is also a low-cost and high-speed analytical method with broad applications in plant biology (Montes et al 2007). NIR absorption bands are due to overtone and combination vibrations of C-H, N-H, O-H, and S-H functional groups, which enable the prediction of diverse organic compounds. NIR spectra from biological materials have multiple overlapping absorbance patterns due to the complex mixture of organic compounds in these samples. Multivariate statistical approaches are required to interpret NIR spectra from biological samples. Cereal grain spectra typically are measured from fine ground powders or as bulk whole grains (Orman and Schuman 1991). NIR data are collected from these types of samples by placing them in cups or cuvettes with a defined surface area and path length. Both ground and whole grain methods

utilize multiple seeds and can only give average composition estimates for the seed sample. Single-seed NIR is necessary to measure phenotypes in segregating populations.

NIR transmittance and reflectance spectroscopy on intact seeds has been widely used for classifying seeds for particular attributes. Maize kernels were classified according to characteristics such as starch composition (Campbell et al 2000), hardness (Robutti 1995), avidin (Kramer et al 2002), or mycotoxin levels (Pearson et al 2001; Dowell et al 2002). Applications for measuring wheat attributes include wheat classes (Delwiche and Massie 1996), color (Wang et al 1999), insect infestation (Rigday and Chambers 1996), hardness (Maghirang and Dowell 2003), starch composition (Delwiche et al 2006, 2009), or vitreousness (Dowell 2000; Wang et al 2002).

The prediction of constituent concentrations using NIR spectroscopy on intact single seeds has been most successful for plants with small seeds and relatively uniform distribution of seed constituents, such as rapeseed (Velasco et al 1999a; Velasco and Möllers 2002; Hom et al 2007), wheat (Delwiche 1998; Delwiche and Hruschka 2000), and sunflower achenes (Velasco et al 1999b; Velasco et al 2004). Cereal grains have a starchy endosperm and an oil-rich embryo. The maize embryo is larger than embryos of other cereal crops, resulting in an asymmetric distribution of protein, starch, and oil within the kernel. Compositional asymmetry results in distinct NIR spectra when the germinal or abgerminal side of the kernel is presented to the spectrometer (Orman and Schumann 1992; Weinstock et al 2006; Janni et al 2008). Maize kernels are also variable in size and shape and present differing surface areas and path lengths for spectral collection. Near-infrared transmittance (NIT) spectroscopy can account for the asymmetrical distribution of seed constituents because the spectra are collected from light passing through whole kernels. However, attempts to predict moisture and oil using single-kernel NIT have had mixed success (Finney and Norris 1978; Orman and Schumann 1992; Cogdill et al 2004).

An alternative approach is to obtain consistent reflectance spectra. Baye et al (2006) developed calibrations utilizing spectra collected from the abgerminal side of the kernel, while Weinstock et al (2006) used hyperspectral imaging to select NIR data from relevant sections of the kernel. Although these approaches yield acceptable composition predictions, they are low-throughput. The kernels need to be hand-placed in a specific orientation on the spectrometer or hand-placed in custom spacing grids to ensure

*The e-Xtra logo stands for “electronic extra” and indicates that Supplemental Figure 1 appears online.

¹ Horticultural Sciences Dept. and Plant Molecular & Cellular Biology Program, University of Florida, Gainesville, FL 32611.

² USDA-ARS, Grain Marketing Production Research Center, Manhattan, KS 66502.

³ Pioneer Hi-Bred International Inc., A DuPont Company, Johnston, IA 50131.

⁴ Florida Agricultural & Mechanical University, Tallahassee, FL 32307.

⁵ Corresponding author. Phone: 352-392-7571. E-mail: settles@ufl.edu

even spacing for hyperspectral images. Janni et al (2008) developed a system that uses an airstream to tumble individual kernels during spectral acquisition and obtain average reflectance over the whole kernel surface. This system gave accurate percent oil predictions in single seeds. However, the tumbling kernel analyzer requires 12 sec to collect a single spectrum and is not engineered for continuous data collection. Armstrong (2006) developed a NIR instrument that collects a spectrum from a seed as it falls through a glass tube. The glass tube NIR instrument accurately predicts soybean seed protein levels and can collect spectra from 10 seeds/sec.

Here we describe a modification to the glass tube NIR instrument that integrates seed weight measurements with spectral acquisition. In contrast to our previous work (Baye et al 2006), hand-orientation of kernels is not necessary. Instead of acquiring NIR spectra from just one side of the kernel, this kernel analyzer averages the reflectance from the top and bottom of the glass tube to reduce variability due to the orientation of the kernel as the spectrum is collected. The objectives of this research were to determine whether the glass tube NIR instrument can be used to develop reliable predictions for percent starch, protein, and oil levels in single maize kernels and to evaluate its feasibility for separating kernels from segregating ears into mutant and normal kernel classes. Earlier work suggested that calibrations for relative composition of single seeds are technically challenging (Baye et al 2006). Our results show that this spectral acquisition design overcomes many of the technical challenges for simultaneously measuring protein, starch, and oil and is an effective seed phenomics technology.

MATERIALS AND METHODS

Maize Samples for Calibrations

Kernels were sampled from 120 different maize ears from 84 maize genotypes/accessions. The majority of seeds were sampled from the germplasm accessions used to develop the maize nested association mapping panel (Yu et al 2008), including environmental replicates that showed differences in kernel composition based on bulk NIR analysis (E. Buckler, *personal communication*). Additional kernel composition variation was included by sampling the Illinois long-term selection strains with altered protein and oil levels (Moose et al 2004), kernel mutants with known effects on starch and protein (e.g., *sh2*, *o2*, *bt1*), and seed mutant lines from the UniformMu transposon-tagging population (McCarty et al 2005). The specific UniformMu seed mutants were selected based on longitudinal hand-sections that showed visual differences in the relative sizes of starchy endosperm, vitreous endosperm, and embryo. Finally, several inbred lines commonly used for genetic studies were included. Three kernels per ear were randomly selected, resulting in a set of 360 kernels. Due to the amount of meal required for starch, protein, and oil analysis for each constituent, a separate set of kernels was used to collect NIR and chemical composition analytical data.

Seed Weight and Near-Infrared Data Collection

Individual kernels were manually dropped onto a microbalance consisting of a MK4 microbalance head and a Stabal control unit (CI Electronics, Salisbury, UK). The kernel weight was automatically recorded using custom software written in Microsoft Visual Basic 6.0. After a stable seed weight was recorded, the software triggered a solenoid to open an airstream and blow the kernel into the glass tube NIR instrument. The NIR instrument collected the spectra essentially as described in Armstrong (2006). Briefly, the kernel fell through a glass tube (12 mm × 60 mm), which was illuminated by multiple halogen lamps. Reflected light was collected through two 400-micron fiber-optic cables positioned at the top and bottom ends of the glass tube. The fiber-optic cables were attached to an InGaAs array based spectrometer (NIR-256-1.7T1.,

Control Development, South Bend, IN). A single NIR spectrum was recorded with a 40-msec integration time. Reflectance values were recorded at 1-nm intervals between 907 and 1689 nm and absorbance values were calculated as $\log(1/R)$. The custom Microsoft Visual Basic 6.0 program then centered each spectrum to an arbitrary mean of 1. A dark background and a reference spectrum (Spectralon, Labsphere, North Sutton, NH) were measured before recording kernel spectra. Each kernel was measured three times to allow an average of the three NIR spectra to be used for model development.

Constituent Analysis

After NIR data collection, the single kernels were transferred to 2-mL microcentrifuge tubes with two 7.9-mm steel ball bearings and ground for 5–10 min in a MiniBeadbeater-96 (BioSpec Products, OK). Starch was determined by an enzymatic hydrolysis of the maize meal with thermostable α -amylase and amylogucosidase followed by a colorimetric determination of glucose with a glucose oxidase-peroxidase (GOP) system (Karkalas et al 1985). α -Amylase solution (1.5 mL, Sigma-Aldrich A3403, from *Bacillus licheniformis*, 19,896 U/mL; diluted 1:15 in 50 mM MOPS buffer, pH 7.0) and 2 mL of water were added to 50 mg of maize meal, mixed thoroughly, and the suspension was kept at 85°C for 40 min under constant stirring. After cooling to room temperature, the volume was adjusted to 100 mL with water, and 1 mL of this solution was combined with 1 mL of amylogucosidase solution (Sigma-Aldrich A1602 from *Aspergillus niger*, 1888 U/mL; diluted 1:200 in 200 mM sodium acetate buffer, pH 4.5). The mixture was incubated overnight at 55°C. Glucose was determined by mixing 0.2 mL of the sample with 1.5 mL of GOP reagent (700 U glucose peroxidase, Sigma-Aldrich P8125 from horseradish, 113 U/mg; 10,000 U of glucose oxidase, Sigma-Aldrich G6641 from *Aspergillus niger*, 21,200 U/g; 0.22 mol *p*-hydroxybenzoic acid, and 0.40 mmol *p*-aminoantipyrine were dissolved in 1 L of 1M phosphate buffer, pH 7.5) and incubating at 55°C for 30 min. Sample absorbance was measured at 510 nm against 0.2 mL of water mixed with 1.5 mL of GOP reagent using a Beckman DU-50 series spectrophotometer. A reagent blank absorbance was determined by a parallel extraction without the addition of maize meal. The reagent blank absorbance was subtracted from each sample. The glucose content in the extract was multiplied by 0.9 for the calculation of starch. Each kernel was assayed between two and four times and an average of the analytical replicates was used for the calibration.

For protein analysis, ground maize meal was dried for five days at 60°C and the moisture content was determined by weighing samples before and after drying. Five seeds were deleted from the protein set due to errors in recording weights either before or after drying, leaving 355 samples for calibration development. Total N content was measured from the dried material by combustion analysis with a CN analyzer (NCS 2500, CE instruments, Milan, Italy) and the protein content was calculated as $N \times 6.25$. The protein content on a fresh weight basis was used for calibration.

Seed oil content was measured as described previously (Zheng et al 2008). Oil content of individual air-dried seeds was determined with a PCT-20/20B NMR analyzer (Process Control Technology, Ft. Collins, CO). Each kernel was measured in three replicates and the average oil value was used for the calibration. Pure maize oil was used as a standard.

Calibration and Validation

Partial least squares (PLS) regression models for the prediction of starch, protein, and oil were developed using The Unscrambler 9.8 (Camo Software, Oslo, Norway). Before calibration, the absorbance values from 910 to 1689 nm were pretreated with several approaches. Individual spectra and the average of the three NIR spectra from individual kernels were used for model development. Both data sets gave similar calibrations and only the av-

erage data are reported below. In addition, calibrations were developed for either mean-centered or log(1/R) spectra with both data sets yielding similar calibrations. Only the calibrations for the mean-centered spectra are reported here. The average spectra and corresponding analytical data used for the calibrations are available upon request. The effects of several spectral filters were also tested. First- and second-derivative spectra were calculated using a range of gap and segment sizes (5–30 nm). The gap and segment sizes had little effect on calibration model performance. Results are only reported for first derivatives calculated as $\lambda_n = (x_{n+10} - x_{n-10})/20$ and second derivatives calculated as $\lambda_n = (x_{n+20} - 2x_n + x_{n-20})/400$. Standard normal variate (SNV) transformation and multiplicative scatter correction (MSC) were calculated in The Unscrambler program (default settings).

The spectra and analytical data were sorted according to respective constituents and every third sample was removed to generate an external validation set. Chemometric models were first developed on the calibration set of 240 spectra and analytical values. The models were evaluated through cross-validation using eight randomly selected segments. One-eighth of the calibration set was removed and predicted using a model based on the remaining samples. The process was iterated until all samples were removed and predicted. An optimal model for each pretreatment was selected using default software settings to minimize the number of PLS factors and the prediction error sum of squares. The cross-validation predictions were evaluated using the coefficient of multiple determination (R^2) and the standard error of cross-validation (SECV). The optimal model was used to predict the

TABLE I
Analytical Reference Values of Kernels Used to Calibrate the NIR Spectrometer

Constituent	<i>n</i>	Mean	SD	SDR ^a	Range	<i>r</i> ² with Seed Wt
Relative (% fresh weight)						
Starch	360	55.3	7.2	2.20	17.3–70.0	0.20
Protein	355	13.8	2.7	0.36	4.1–29.8	0.030
Oil	360	4.2	2.2	0.08	0.7–19.1	4 × 10 ⁻⁴
Absolute (mg/kernel)						
Starch	360	127.8	47.4	5.20	15.0–255.5	0.94
Protein	355	31.7	10.9	0.77	8.7–63.9	0.76
Oil	360	9.5	5.7	0.16	1.2–50.0	0.26
Seed weight	360	227.6	72.6	0.70	69.2–423.6	na

^a Standard deviation of repeatability.

TABLE II
Partial Least Squares (PLS) Regression Statistics Using Optimal Spectra Data Pretreatments

Constituent Unit	Constituent	Spectra Data Pretreatment ^c	PLS Factors	Cross-Validation ^a		External Validation ^b	
				<i>R</i> ²	SECV	SEP	SDR
Relative (% fresh weight)							
	Starch	1Der+MSC	7	0.66	4.24	3.72	3.19
	Protein	2Der+MSC	6	0.88	0.93	0.81	0.72
	Oil	1Der+MSC	7	0.86	0.79	0.79	0.66
Absolute (mg/kernel)							
	Starch	1Der	7	0.85	18.34	18.20	16.27
	Protein	1Der	8	0.89	3.65	3.82	3.51
	Oil	1Der	10	0.85	2.19	2.74	2.36
	Seed weight	1Der	7	0.86	27.16	27.60	25.63

^a *R*², coefficient of multiple determination; SECV, standard error of cross-validation.

^b SEP, standard error of prediction; SDR, standard deviation of repeatability.

^c 1Der, first derivative; 2Der, second derivative; MSC, multiplicative scatter correction.

TABLE III
Partial Least Squares (PLS) Model Statistics of Protein Values Using Multiple Spectral Data Pretreatments

Constituent Unit	Pretreatment ^c	PLS Factors	Cross-Validation ^a		External Validation ^b	
			<i>R</i> ²	SECV	<i>R</i> ²	SEP
Relative (% fresh weight)						
	None	11	0.82	1.14	0.75	1.34
	MC	12	0.80	1.21	0.86	1.03
	MC, 1Der	8	0.81	1.17	0.87	1.00
	MC, 2Der	8	0.81	1.18	0.87	0.99
	MC, SNV	11	0.88	0.92	0.87	0.97
	MC, MSC	10	0.86	0.99	0.88	0.95
	MC, 1Der, MSC	8	0.89	0.90	0.89	0.90
	MC, 2Der, MSC	6	0.88	0.93	0.91	0.81
Absolute (mg/kernel)						
	None	11	0.87	3.93	0.80	4.79
	MC	4	0.88	3.86	0.87	3.94
	MC, 1Der	8	0.89	3.65	0.88	3.82
	MC, 2Der	5	0.88	3.86	0.83	4.63
	MC, SNV	6	0.77	5.24	0.73	5.72
	MC, MSC	9	0.81	4.81	0.83	4.51
	MC, 1Der, MSC	7	0.84	4.40	0.85	4.28
	MC, 2Der, MSC	6	0.82	4.58	0.82	4.66

^a *R*², correlation of multiple determination; SECV, standard error of cross-validation.

^b SEP, standard error of prediction.

^c MC, mean centering; 1Der, first derivative; 2Der, second derivative; SNV, standard normal variate transformation; MSC, multiplicative scatter correction.

samples in the external prediction set, and the model performance was further evaluated based on the r^2 of the observed and predicted analytical values as well as the standard error of prediction (SEP). The standard deviation of repeatability (SDR) of the NIR measurement was estimated by applying the PLS models on the three single NIR scans of the 120 samples of the prediction sets. The repeatability error was calculated as the square root of the average variances of the predicted constituent.

Detection of Seed Composition Mutants

Homozygous seed composition mutants in defined inbred backgrounds (W22, W23, or W64A) were ordered from the Maize Genetics Cooperative Stock Center. These mutants were crossed to their respective inbred parents. The resulting heterozygous F_1 plants were self-pollinated to generate F_2 populations segregating for each mutant. Seed weight and NIR data were collected from 96 F_2 kernels for each mutant. When possible, the mutant and normal sibling seeds were segregated by visual selection and NIR data were collected from 72 normal and 24 mutant kernels that were indexed in 48-well microtiter plates. For each F_2 , the NIR absorbance values from 910 nm to 1690 nm were analyzed by principal components analysis (PCA) (SAS Institute, Cary, NC). Individual seed phenotypes were then confirmed either with a second visual score or a destructive analysis. The *bt1*, *bt2*, *sh1*, *sh2*, *su1*, and *ae* mutants were re-scored by visual inspection. The extent of vitreous endosperm development was assayed by longitudinal sections of the families segregating for the *h1*, *fl1*, *fl2*, and *o2* mutants. The *wx1* phenotype was scored by cutting the crown of the kernel and staining the endosperm with IKI solution (1% iodine, 2% potassium iodide) for 30 sec.

Single-Kernel NIR Predicts Seed Composition

We modified the single-seed NIR instrument developed by Armstrong (2006) by integrating a microbalance to record a seed weight along with a NIR spectrum from individual seeds. The instrument was then calibrated to determine its utility for predicting maize kernel composition. NIR calibrations are most robust when developed with samples that have a full range of possible compositions (reviewed in Williams and Norris 1987). We chose maize samples to cover a broad range of kernel composition and genetic diversity. The starch, protein, and oil levels from this germplasm collection had large ranges both in terms of relative (%) and absolute (mg/kernel) values (Table I). All of the kernels were stored in controlled temperature and humidity environments, and moisture content was much less variable (7.0 and 10.9%) than the other measured constituents. The data for each constituent were separated into a calibration subset and an external validation subset for partial least squares (PLS) regression. As expected, the Illinois long-term selection lines gave extreme values for percent protein and oil. To ensure that both the calibration and validation sets have a similar range and variance, the data sets were sorted according to their constituents, and every third sample was removed to generate the validation sets. In addition, the 120 maize ears used were randomly separated into calibration and validation samples to ensure that kernels in the two sets were derived from different ears. Models developed for a random partitioning of ears to the validation set resulted in similar calibration statistics (data not shown), indicating that the sorting procedure did not lead to an overfitting of the PLS models. PLS prediction models were calcu-

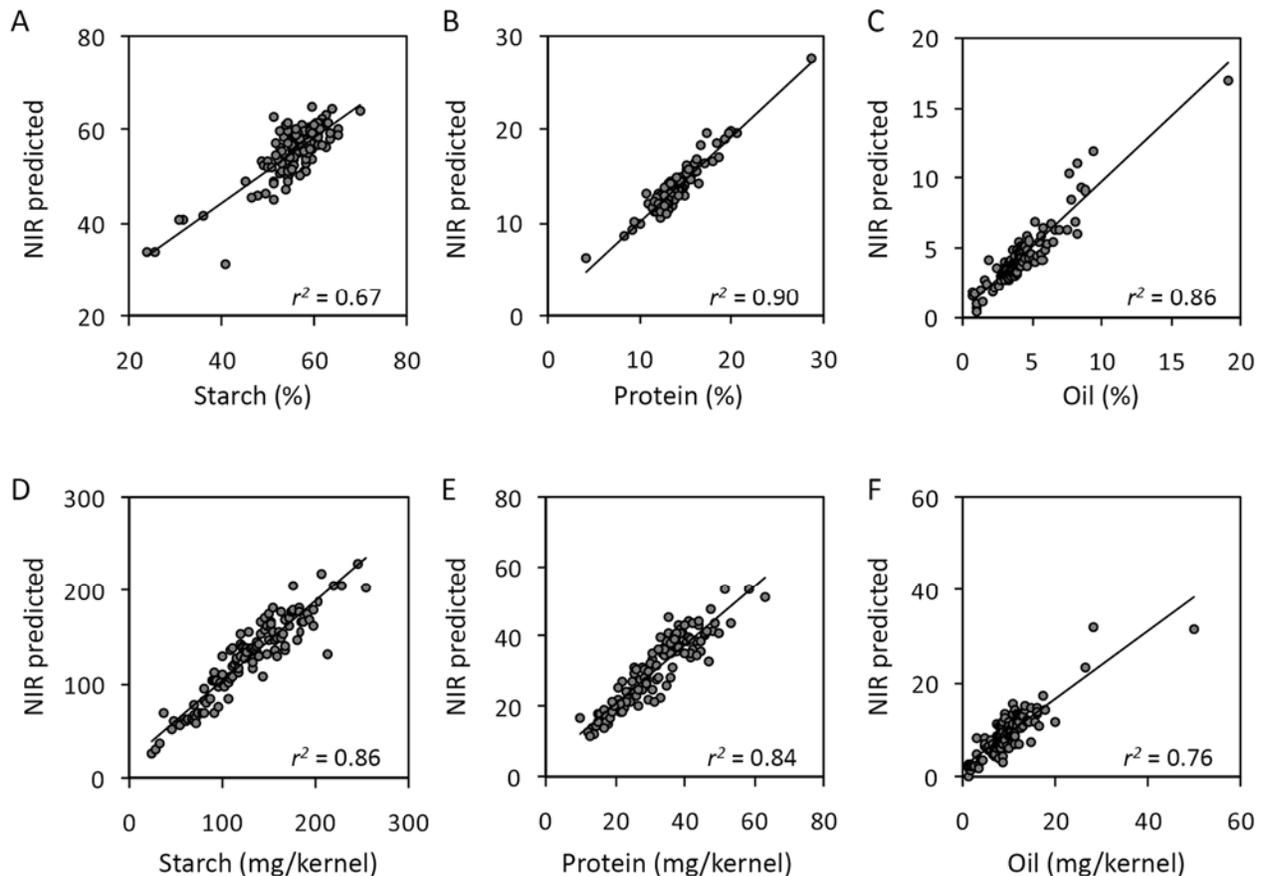


Fig. 1. Scatter plots of NIR-predicted and analytical reference values for starch, protein, and oil. Each plot shows values for the external validation set and a linear regression trend line. All NIR-predicted values used optimal PLS regressions in Table II. **A–C,** Predictions for relative constituents. **D–F,** Predictions for absolute constituents.

lated for both relative and absolute constituent data. Spectral filters can be used to remove systematic variation in the NIR spectra that is not related to analytical data (Williams and Norris 1987), and several spectral pretreatments were investigated. We tested the effects of first and second derivatives, standard normal variate (SNV) transformation, multiplicative scatter correction (MSC), as well as combinations of derivatives and MSC on PLS model performance. For all constituents, the use of spectral filters had only minor effects on PLS model performance (Table III) for the prediction of percentage and protein (mg). For the relative analytical data, the most predictive models were generated using pretreatments of first or second derivative and MSC (Table II). For absolute analytical data, first derivatives resulted in the best calibrations. Except for percent starch, the optimal PLS calibrations all have good correlations between the predicted and analytical values with R^2 values of 0.79–0.91. Similar errors in the calibration and external validation data sets indicate that the models are not over-fit to the calibration data (Table II). Scatter plots of the observed and predicted values for the external validation data sets indicate that the models do not have strong bias. Overall, the PLS results suggest that the single-seed NIR instrument can predict percent protein and oil with good accuracy and is useful for classifying kernels into groups of low, medium, and high % starch.

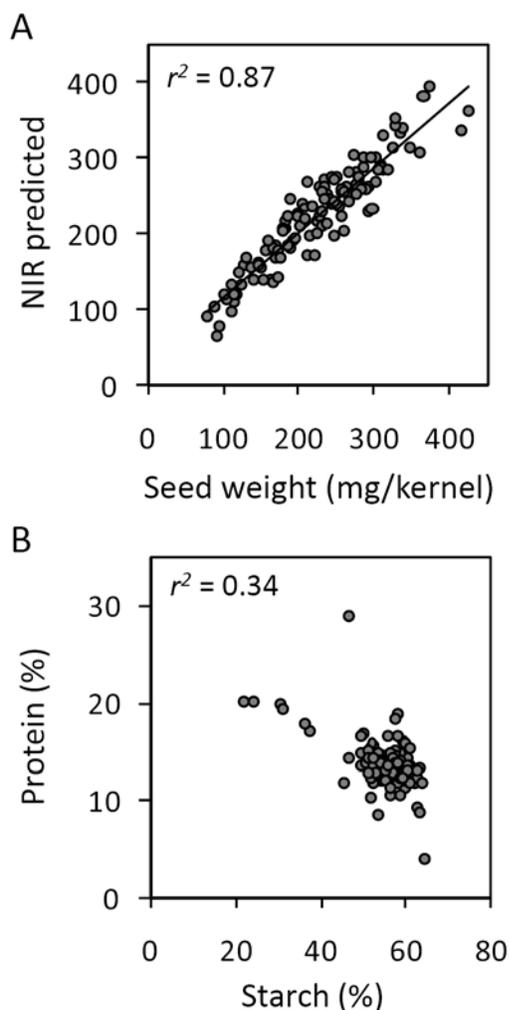


Fig. 2. **A**, Scatter plot and linear regression trend line of the NIR-predicted and analytical reference values for seed weight. NIR-predicted values are from the external validation set using the optimal PLS regression described in Table II. **B**, Scatter plot showing the relationship between average starch and protein reference values of the 120 sampled ears.

Importantly, the percent protein and percent starch calibrations are similar to previously reported NIR calibrations for bulk whole maize grain (Orman and Schumann 1991). The oil calibration reported here is significantly improved over NIR bulk grain predictions (Orman and Schumann 1991) and near-infrared transmission spectroscopy of single grain (Orman and Schuman 1992) but gives lower accuracy when compared with the lower throughput tumbling-kernel NIR system (Janni et al 2008).

Single-Kernel NIR Reports Both Kernel Size and Percent Constituents

PLS calibrations for absolute (mg/kernel) protein and oil content performed similar to percent calibrations, but absolute starch could be predicted with higher accuracy than percent starch. The spectra also predict seed weight with similar levels of accuracy to other absolute constituents (Table II, Fig. 2A). These results raise the issue of whether the PLS calibrations primarily use seed weight to predict absolute seed compositions. Absolute levels of starch have a high correlation with seed weight (Table I), and a linear regression with seed weight can predict the absolute levels of starch better than the PLS models. However, absolute protein and oil are not as well correlated with seed weight, and the PLS models predict absolute protein and oil with far greater accuracy than is possible with a linear regression on seed weights. In addition, seed weight has low correlations with the relative levels of protein, oil, and starch (Table I), and the relative levels of the constituents are not correlated to each other (e.g., Fig. 2B). Yet, the NIR spectra can predict relative levels of all of the major storage molecules (Table II). These observations indicate that single-kernel NIR reports both seed weight and relative levels of the seed storage molecules independently. It was not obvious how the NIR spectra could report seed weight, absolute constituents, and relative constituents simultaneously. We investigated the basis of this phenomenon in more detail.

The relative absorbance and transmittance of near-infrared radiation is based on the Beer-Lambert law (Siesler et al 2002). The Beer-Lambert law states that the radiation absorbed by a substance is related to the product of the molar absorptivity, the concentration, and the sample path length. Absorbance values are additive, and a spectrum of a complex sample mixture represents a weighted sum of the spectra of the pure constituents. The molar absorptivity for each kernel constituent is constant, while the constituent concentration is variable from sample to sample. In bulk or ground grain NIR analysis, the light scattering of diffuse reflectance does not allow the Beer-Lambert Law to be applied directly to the spectra. Nevertheless, a fixed volume of grain or meal is placed in the spectrometer to maintain a constant path length (e.g., Orman and Schumann 1991). In single-kernel NIR, a spectrum is recorded from an intact kernel. Maize kernels show wide ranges in size, which affects path length. Larger, heavier kernels will have longer path lengths than smaller, lighter kernels.

To separate the effects of changes in constituent concentration from changes in sample path length, we compared spectra from kernels with similar weight but different relative chemical composition (Fig. 3A). We also compared spectra from kernels with different weights but similar percent composition (Fig. 3B). In both comparisons, the spectra show differences in absorbance pattern, illustrating that both constituent concentrations and sample path lengths affect the spectra.

MSC and SNV are two approaches to remove effects due to light scattering and differences in spectroscopic path lengths (Geladi et al 1985; Barnes et al 1989). MSC and SNV spectra pretreatments gave the best PLS calibration statistics for relative constituents, but had negative effects on calibrating absolute constituents and seed weight (Table III and data not shown). These results suggest that SNV and MSC remove relevant information about absolute composition. Based on these observations, we suggest that path length effects are the primary source for seed

weight information in the NIR spectra. Constituent concentration absorbance effects are likely to be the source for predicting the relative levels, and a combination of these two spectroscopic effects is needed to predict absolute levels of seed constituents.

Single-Kernel NIR Differentiates Seed Composition Mutants

Our PLS calibrations indicate that the glass tube NIR instrument is sensitive to a large range of possible changes in seed composition. We next investigated whether it was possible to identify seed phenotypic variation with the glass tube instrument when no calibration or analytical reference data are available. Several groups have reported that qualitative changes in grain composition are possible to identify with NIR data using principal component analysis (PCA) (Campbell et al 2000; Jacobsen et al 2005; Delwiche et al 2006; Munck 2007). Most of these studies compared seeds from different plants. Spectra from independent ears are influenced by environment and background effects, both of which could alter chemical components and hence NIR spectral data. It is difficult to determine whether the underlying cause of any NIR spectral differences observed in plant-to-plant comparisons are due to multiple environmental and genetic background effects or due to the qualitative phenotype of interest. As a hypothetical example, an apparent difference between lines that

have high versus low amylose content may be driven by genetic background differences in protein content.

Single-seed NIR provides the opportunity to control for genetic background and environmental effects by comparing seeds from a single ear of maize. We tested a series of maize seed composition mutants known to reduce the amount of starch (*bt1*, *bt2*, *sh1*), to alter starch composition (*wx*, *ae*, *h1*, *su1*), or to change protein composition (*fl1*, *fl2*, *o2*) (reviewed in Gibbon and Larkins 2005; Balconi et al 2007). Each mutant was in a defined inbred background, and we developed segregating F₂ families by crossing to the corresponding inbred. We investigated a variety of spectral pretreatments before PCA. For all samples, we found that PCA of mean-centered spectra yields two principal components (PC1 and PC2) that explain at least 90% of the variance within the spectra from an individual F₂ family.

The scatter plots of PC1 and PC2 showed clear differentiation of mutant and normal kernels for all of the composition mutants tested (Fig. 4 and Supplemental Figure 1). However, the extent of overlap between mutant and normal spectra varied by mutant genotype. Most starch biosynthetic mutants including *bt1*, *bt2*, and *sh2*, as well as the *su1* starch quality and *o2* protein quality mutant showed separation between mutant and normal kernels with no overlap (Fig. 4A and Supplemental Figure 1A-D). The *fl1*

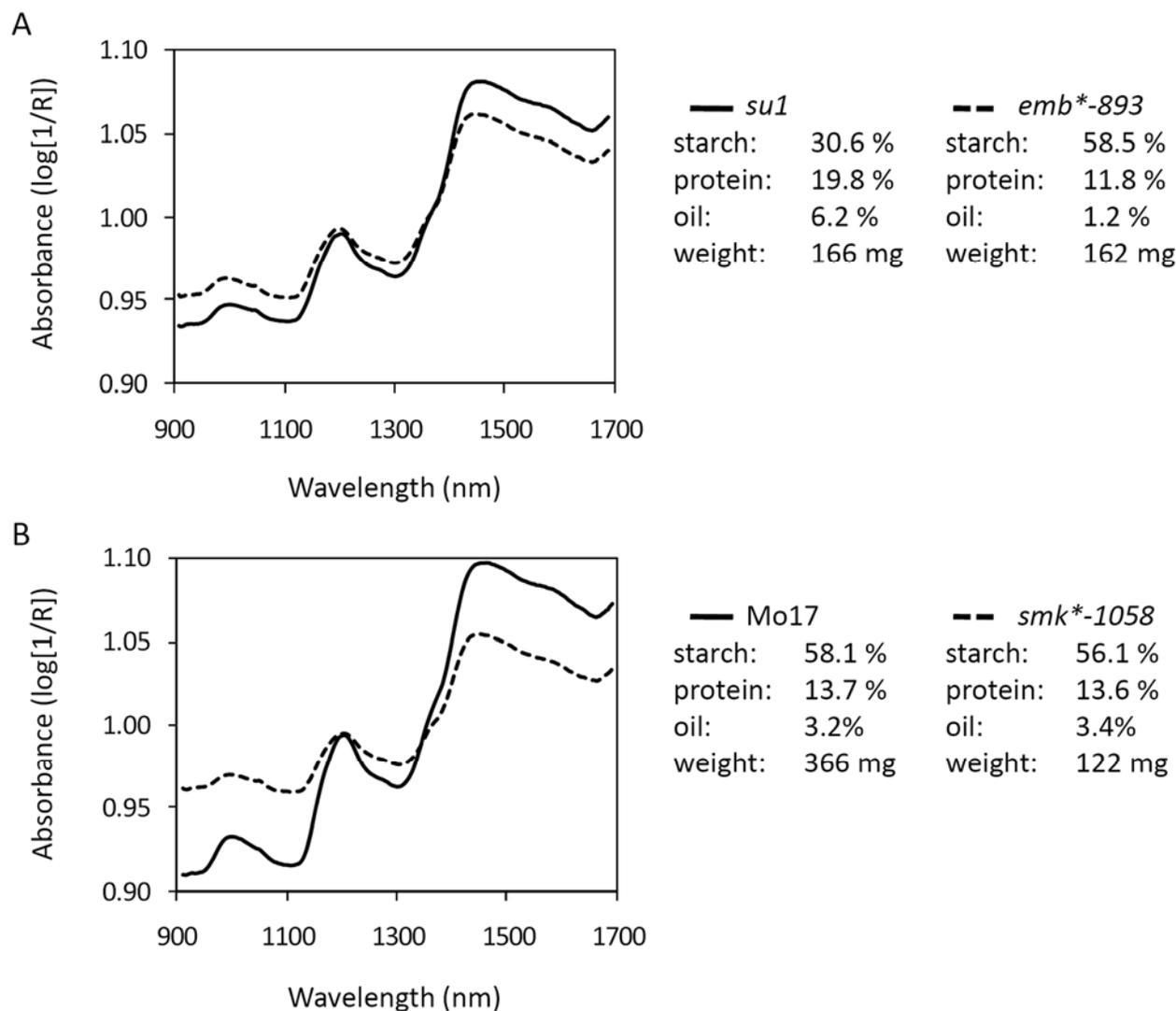


Fig. 3. Examples of single-kernel NIR spectra. Each line represents average NIR spectra of nine kernels from an individual maize ear. Three kernels each were derived from the starch, protein, and oil calibration sets. Constituent percentages were averaged from three kernels, and seed weights are the average of all nine kernels. **A**, Comparison of two genotypes with similar seed weights but different relative amounts of starch, protein, and oil. **B**, Comparison of two genotypes with different seed weights but similar relative amounts of starch, protein, and oil.

and *fl2* mutants as well as the *sh1* and *h1* mutants showed distinct mutant and normal kernel classes, but the borders between these classes had overlapping kernels (Fig. 4B and Supplemental Figure 1E).

Finally, *wx* and *ae* did not have apparent mutant and normal classes (Fig. 4C and Supplemental Figure 1). However, the mutant kernels for these loci clustered and the NIR spectra could be

used to greatly enrich for mutant or normal kernels. Although the starch biosynthetic mutants show a mutant class with reduced seed weight (Fig. 4E), the protein and starch quality mutants do not affect seed weight (Fig. 4D, 4F). Combined, these data indicate that the glass tube NIR instrument can be used to identify or sort kernels based on quantitative changes in starch content or qualitative phenotypes for which there are no calibrations.

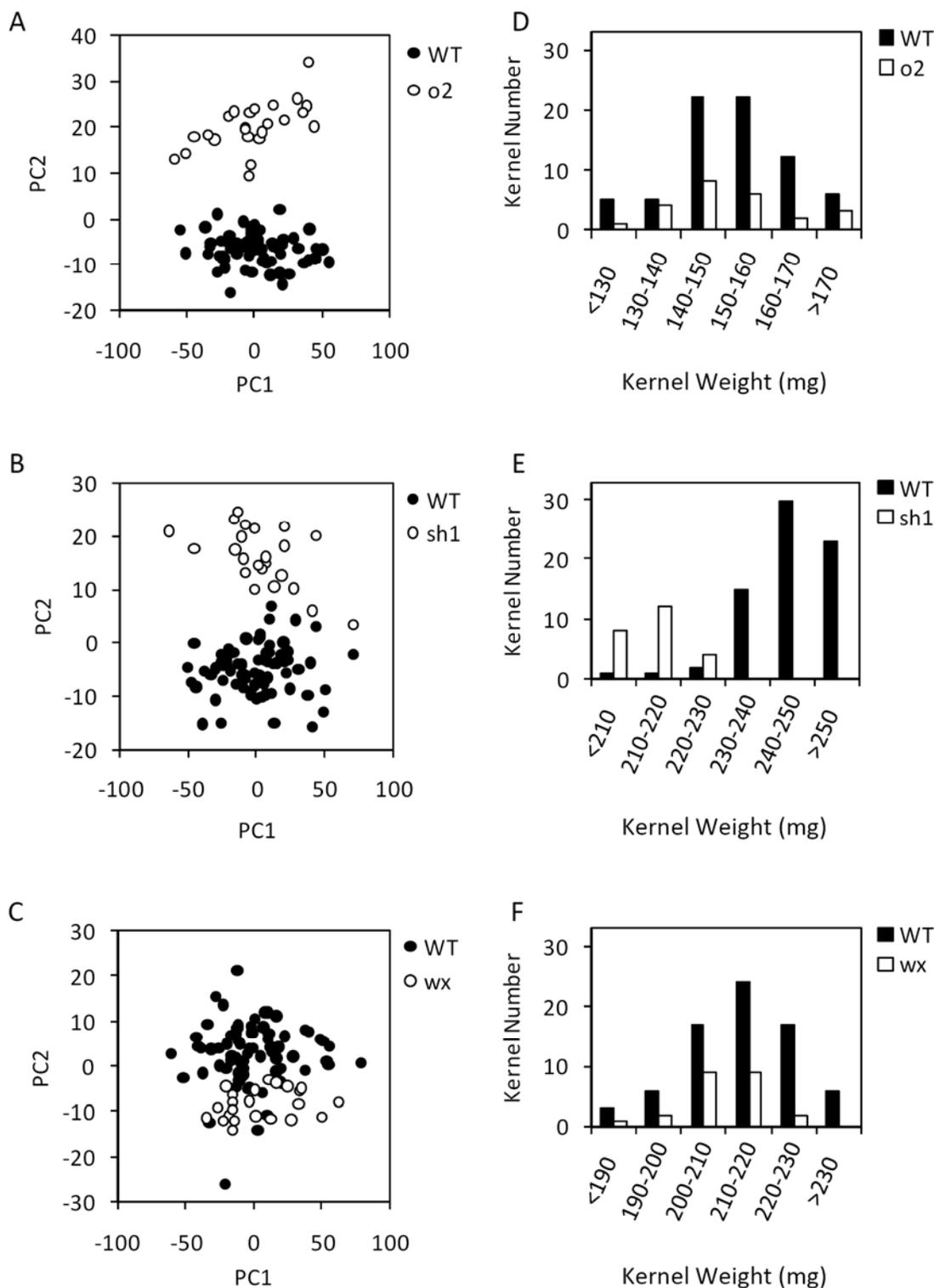


Fig. 4. Differentiation of seed composition mutants. Segregating families were generated for 11 known seed composition mutants. NIR and seed weight data were collected from 96 kernels for each mutant. Kernels were indexed in 48-well microtiter plates. Mutant and normal phenotypes were confirmed with destructive analysis when needed. **A–C,** Scatter plots of the first two components (PC1 and PC2) from principal component analyses for individual segregating families. **D–F,** Histograms of seed weight for each family in A–C.

CONCLUSIONS

We have shown that acquiring NIR spectra from single maize kernels as they fall through an illuminated glass tube provides a quantitative and comprehensive prediction of seed composition and weight. The PLS calibrations presented here indicate that the glass tube NIR instrument provides equivalent or improved predictions to bulk, whole grain NIR analyzers, but at single-kernel resolution. Similar to earlier studies (Ormann and Schumann 1991; Baye et al 2006), we found it difficult to obtain accurate percent starch predictions. We have also shown that single-seed spectra can be used to identify kernels with qualitative changes in seed composition with mutations that alter endosperm structure (*o2*, *fl1*, *fl2*, *h1*) being the simplest to differentiate.

These results indicate that our glass tube NIR spectrometer provides a platform to study the maize seed phenome. Importantly, this design combines good accuracy with high throughput data collection. The instrument has the potential to collect NIR spectra at a rate of 10 kernels/sec. This technology enables genetic screens, QTL or association analysis, and breeding for composition phenotypes at the single-kernel level. Based on earlier work (Armstrong 2006) on soybean, the glass tube instrument can provide seed phenotype information for multiple plant species. Appropriate scaling of the glass tube should allow this technology to be extended to plant species with seeds of different sizes.

ACKNOWLEDGMENTS

We thank the USDA-ARS North Central Regional Plant Introduction Station, Ed Buckler, and Steve Moose for providing seed stocks of diverse maize germplasm as well as Danielle Watts and Matthew Cohen for providing access to the C/N analyzer. We also thank the anonymous reviewers for their helpful comments. This work was supported by the National Science Foundation Plant Genome Research Project award DBI-0606607.

LITERATURE CITED

- Armstrong, P. R. 2006. Rapid single-kernel NIR measurement of grain and oil-seed attributes. *Appl. Eng. Agric.* 22:767-772.
- Balconi, C., Hartings, H., Lauria, M., Pirona, R., Rossi, V., and Motto, M. 2007. Gene discovery to improve maize grain quality traits. *Maydica* 52:357-373.
- Barnes, R. J., Dhanoa, M. S., and Lister, S. J. 1989. Standard normal variance transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43:772-777.
- Baye, T. M., Pearson, T. C., and Settles, A. M. 2006. Calibration development to predict maize seed composition using single kernel near infrared spectroscopy. *J. Cereal Sci.* 43:236-243.
- Campbell, M. R., Sykes, J., and Glover, D. V. 2000. Classification of single- and double-mutant corn endosperm genotypes by near-infrared transmittance spectroscopy. *Cereal Chem.* 77:774-778.
- Cogdill, R. P., Hurburgh, C. R., Rippke, G. R., Bajic, S. J., Jones, R. W., McClelland, J. F., Jensen, T. C., and Liu, J. 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. *Trans. ASAE* 47:311-320.
- Delwiche, S. R. 1998. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* 27:241-254.
- Delwiche, S. R., and Hruschka, W. R. 2000. Protein content of bulk wheat from near-infrared reflectance of individual kernels. *Cereal Chem.* 77:86-88.
- Delwiche, S. R., and Massie, D. R. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chem.* 73:399-405.
- Delwiche, S. R., Graybosch, R. A., Hansen, L. E., Souza, E., and Dowell, F. E. 2006. Single kernel near-infrared analysis of tetraploid (Durum) wheat for classification of the waxy condition. *Cereal Chem.* 83:287-292.
- Dowell, F. E. 2000. Differentiating vitreous and nonvitreous durum wheat kernels by using near-infrared spectroscopy. *Cereal Chem.* 77:155-158.
- Dowell, F. E., Pearson, T. C., Maghirang, E. B., Xie, F., and Wicklow, D. T. 2002. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. *Cereal Chem.* 79:222-226.
- Dowell, F. E., Maghirang, E. B., Graybosch, R. A., Berzonsky, W. A., and Delwiche, S. R. 2009. Selecting and sorting waxy wheat kernels using near-infrared spectroscopy. *Cereal Chem.* 86:251-255.
- Finney, E. E., and Norris, K. H. 1978. Determination of moisture in corn kernels by near-infrared transmittance measurements. *Trans. ASAE* 21:581-584.
- Geladi, P., MacDougall, D., and Martens, H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39:491-500.
- Gibbon, B. C., and Larkins, B. A. 2005. Molecular genetic approaches to developing quality protein maize. *Trends Genet.* 21:227-233.
- Hom, N. H., Becker, H. C., and Möllers, C. 2007. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* 153:27-34.
- Jacobsen, S., Sondergaard, I., Moller, B., Desler, T., and Munck, L. 2005. A chemometric evaluation of the underlying physical and chemical patterns that support near infrared spectroscopy of barley seeds as a tool for explorative classification of endosperm, genes and gene combinations. *J. Cereal Sci.* 42:281-299.
- Janni, J., Weinstock, B. A., Hagen, L., and Wright, S. 2008. Novel near-infrared sampling apparatus for single kernel analysis of oil content in maize. *Appl. Spectrosc.* 62:423-426.
- Karkalas, J. J. 1985. An improved enzymatic method for the determination of native and modified starch. *J. Sci. Food Agric.* 36:1019-1027.
- Kramer, K. J., Morgan, T. D., Throne, J. E., Dowell, F. E., Bailey, M., and Howard, J. A. 2000. Transgenic avidin maize is resistant to storage insect pests. *Nat. Biotechnol.* 18:670-674.
- Maghirang, E. B., Lookhart, G. L., Bean, S. R., Pierce, R. O., Xie, F., Caley, M. S., Wilson, J. D., Seabourn, B. W., Ram, M. S., Park, S. H., Chung, O. K., and Dowell, F. E. 2006. Comparison of quality characteristics and breadmaking functionality of hard red winter and hard red spring wheat. *Cereal Chem.* 83:520-528.
- McCarty, D. R., Settles, A. M., Suzuki, M., Tan, B. C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J., Messing, J., Koch, K. E., and Hannah, L. C. 2005. Steady-state transposon mutagenesis in maize. *Plant J.* 44:52-61.
- McClure, W. F. 2003. 204 years of near infrared technology: 1800-2003. *J. Near Infrared Spectrosc.* 11:487-518.
- Montes, J. M., Melchinger, A. E., and Reif, J. C. 2007. Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* 12:433-436.
- Moose, S. P., Rocheford, T. R., and Dudley, J. W. 2004. Maize selection turns 100: A 21st century genomics tool. *Trends Plant Sci.* 9:358-364.
- Munck, L. 2007. A new holistic exploratory approach to systems biology by near infrared spectroscopy evaluated by chemometrics and data inspection. *J. Chemometrics* 21:406-426.
- Ormann, B. A., and Schumann, R. A. 1991. Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. *J. Agric. Food Chem.* 39:883-886.
- Ormann, B. A., and Schumann, R. A. 1992. Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. *J. Am. Oil Chem. Soc.* 69:1036-1038.
- Osborne, B. G. 2006. Applications of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programmes. *J. Near Infrared Spectrosc.* 14:93-101.
- Pearson, T. C., Wicklow, D. T., Maghirang, F., Xie, F., and Dowell, F. E. 2001. Detecting aflatoxin in single corn kernels by transmit and reflectance spectroscopy. *Trans. ASAE* 44:1247-1254.
- Ridgway, C., and Chambers, J. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* 71:251-264.
- Robutti, J. L. 1995. Maize kernel hardness estimation in breeding by near infrared transmission analysis. *Cereal Chem.* 72:632-636.
- Siesler, H. W., Ozaki, Y., Kawata, S., and Heise, H. M. 2002. Near-infrared spectroscopy principles, instruments, applications. Wiley-VCH: Weinheim, Germany.
- Velasco, L., and Möllers, C. 2002. Nondestructive assessment of protein content in single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 123:89-93.
- Velasco, L., Möllers, C., and Becker, H. C. 1999a. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near infrared reflectance spectroscopy. *Euphytica* 106:79-85.

- Velasco, L., Pérez-Vich, B., and Fernández-Martínez, J. M. 1999b. Non-destructive screening for oleic and linoleic acid in single sunflower achenes by near-infrared reflectance spectroscopy. *Crop Sci.* 39:219-222.
- Velasco, L., Pérez-Vich, B., and Fernández-Martínez, J. M. 2004. Use of near-infrared reflectance spectroscopy for selecting for high stearic acid concentration in single sunflower seeds. *Crop Sci.* 44:93-97.
- Wang, D., Dowell, F. E., and Lacey, R. E. 1999. Single wheat kernel color classification by using near-infrared reflectance spectra. *Cereal Chem.* 76:30-33.
- Wang, D., Dowell, F. E., and Dempster, R. 2002. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. *Cereal Chem.* 79:418-422.
- Weinstock, B. A., Janni, J., Hagen, L., and Wright, S. 2006. Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl. Spectrosc.* 60:9-16.
- Williams, P. C., and Norris, K. H. 1987. Near-infrared technology in the agricultural and food industries. AACC International: St. Paul, MN.
- Williams, P. C., and Norris, K. H. 2002. Near-Infrared Technology in the Agricultural and Food Industries. 2nd Ed. AACC International: St. Paul, MN.
- Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 2178:539-551.
- Zheng, P., Allen, W. B., Roesler, K., Williams, M. E., Zhang, S., Li, J., Glassman, K., Ranch, J., Nubel, D., Solawetz, W., Bhatramakki, D., Laca, V., Deschamps, S., Zhong, G. Y., Tarczynski, M. C., and Shen, B. 2008. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* 40:269-270.

[Received March 4, 2009. Accepted June 10, 2009.]