Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy

Linda Dykes  
*Texas A & M University - College Station*, ldykes@ag.tamu.edu

Leo Hoffmann Jr.  
*Texas A & M University - College Station*

Ostilio Portillo-Ridriguez  
*Texas A & M University - College Station*

William L. Rooney  
*Texas A & M University - College Station*

Lloyd W. Rooney  
*Texas A & M University - College Station*

Follow this and additional works at: [http://digitalcommons.unl.edu/intsormilpubs](http://digitalcommons.unl.edu/intsormilpubs)
1. Introduction

Sorghum (*Sorghum bicolour* (L.) Moench) is the fifth largest cereal crop in the world after wheat, maize, rice, and barley (FAO, 2007). It is grown primarily in hot and dry regions and is used as a food in Africa and Asia and as a feed and industrial grain in the Americas and Australia (Dykes et al., 2005, Waniska and Rooney, 2000). There has been a recent increase in the use of sorghum as a food in the United States because of its gluten-free characteristic as well as other potential health benefits which include slow digestibility, cholesterol-lowering, anticarcinogenic, and anti-inflamatory properties (Bralley et al., 2008, Burdette et al., 2010, Dykes and Rooney, 2006, Moraes et al., 2012, Turner et al., 2006, Yang et al., 2009). These sorghums can be utilized in a wide array of food products such as breads, cakes, cookies, extrudates, tortillas, and tortilla chips (Dykes and Rooney, 2006, Taylor et al., 2006).

Grain sorghum genotypes vary extensively based on pericarp color, kernel size and shape, pericarp thickness, presence/absence of a pigmented testa for condensed tannins, presence of the spreader gene, endosperm color, and secondary plant color (Rooney, 2000, Waniska and Rooney, 2000). Based on pericarp color, sorghum genotypes can be grouped into four categories: white, lemon-yellow, red, and black. Black sorghums are genetically red but turn black in the presence of sunlight during maturation (Dykes et al., 2009). Sorghum has a wide variety of sorghum phenolics and their composition is affected by the genotype (Dykes et al., 2005, 2009, 2011). For example, sorghums with a pigmented testa contain condensed tannins while black sorghums have the highest levels of 3-deoxyanthocyanidins among all sorghums (Beta et al., 1999, Dlamini et al., 2007, Dykes and Rooney, 2006, Dykes et al., 2005, 2009). Both condensed tannins and 3-deoxyanthocyanidins have potential health benefits such as antioxidant and anticarcinogenic properties (Devi et al., 2011, 2012).
Hagerman et al., 1998, Shih et al., 2007, Turner et al., 2006, Woo et al., 2012, Yang et al., 2009).

One of the goals of the Texas A&M AgriLife Research Sorghum Program is to develop sorghums with high levels of phenolics such as condensed tannins and 3-deoxyanthocyanidins. Several methods have been used to determine sorghum phenolic levels. These include colorimetric methods [i.e. Folin–Ciocalteu, Prussian blue, ferric ammonium citrate (International Organization for Standardization), vanillin/HCl, and butanol–HCl methods], which determine relative sorghum phenolic levels among genotypes (Beta et al., 1999, Chiremba et al., 2012, Dlamini et al., 2007, Dykes et al., 2005, Waniska and Rooney, 2000). High performance liquid chromatography (HPLC) coupled with photodiode array (PDA), fluorescent, and/or mass spectroscopy (MS) detectors identify and quantify specific phenolic compounds (Chiremba et al., 2012, Dykes et al., 2009, 2011, Gu et al., 2003, Svensson et al., 2010). All aforementioned methods require sample preparation (i.e. grinding), the use of chemical solvents, technical skills, and are time-consuming. Consequently, they are not practical for screening hundreds, if not thousands, of samples in the short time required for selection in a breeding program.

To mitigate this issue, near-infrared (NIR) spectroscopy has been used to measure composition of phenolic levels in foods/beverages such as rice, cocoa, wine, grapes, apples, and tea (Chen et al., 2008, Cozzolino et al., 2004, 2008, Janik et al., 2007, Pissard et al., 2013, Whitacre et al., 2003, Zhang et al., 2008). NIR prediction is a quick, inexpensive, and non-destructive method to screen samples for desired components assuming that calibration curves can be developed that accurately relate data collected from lab-based assays to the reflectance data obtained from an NIR scan.

The Texas A&M AgriLife Research Sorghum Program has been developing such calibration curves to predict total phenols, condensed tannins, and 3-deoxyanthocyanidin levels in grain sorghum using NIR spectroscopy. The objectives of this study were to determine the range of variation in total phenols, condensed tannins, and 3-deoxyanthocyanidin levels in a diverse set of grain sorghum genotypes and to assess the predictive value of an NIR curve to estimate these compounds in sorghum grain.

2. Materials and methods

2.1. Materials

A total of 287 grain sorghum samples were evaluated for total phenols, condensed tannins, and 3-deoxyanthocyanidins. For purposes of NIR calibration curve development, these were divided into calibration (224) and validation (63) sets (Table 1). These samples included grains from numerous lines and hybrids and they were collected from research trials conducted in Texas (Weslaco, Corpus Christi, College Station, and Halfway), USA. These samples represented the phenotypic diversity in sorghum grain appearance, which include pericarp color, the presence/absence of the pigmented testa layer and spreader gene, kernel size and shape, pericarp thickness, endosperm color, and secondary plant color.

Table 1. Basic color and tannin classifications of sorghum samples used for the calibration and validation sets.

<table>
<thead>
<tr>
<th>Calibrationset</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-tannin</td>
<td>Tannin</td>
</tr>
<tr>
<td>Pericarp sorghum</td>
<td>sorghum</td>
</tr>
<tr>
<td>White</td>
<td>29</td>
</tr>
<tr>
<td>Lemon-yellow</td>
<td>19</td>
</tr>
<tr>
<td>Red</td>
<td>48</td>
</tr>
<tr>
<td>Black</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

2.2. Chemicals and reagents

Gallic acid, catechin hydrate, vanillin, and Folin–Ciocalteu reagent were obtained from Sigma (St. Louis, MO, USA). Methanol and HCl were purchased from VWR (West Chester, PA, USA). Water was purified using the Simplicity® UV water purification system (Millipore, Billerica, MA, USA). All solvents were reagent grade.

2.3. Reference analysis

Whole grains were ground for 2 min using a Braun KSM2 coffee grinder (Gillette Co., MA, USA) prior to analysis. All samples were ground to achieve a particle size that passed through a 500-μm sieve. Samples (0.10–0.30 g) were extracted in 25 mL 1% HCl/methanol (v/v) for 2 h while shaking at low speed using an Eberbach shaker (Eberbach Corp., MI, USA). The extracts were then centrifuged at 2790 g for 15 min in a Sorvall SS-34 centrifuge (DuPont Instruments, Wilmington, DE, USA) and were decanted. To avoid oxidation, extracts were stored in the dark at −20 °C and analyses were performed within 24 h.

Total phenols were measured using the Folin–Ciocalteu method of Kaluza et al. (1980). Total phenol results were expressed as mg gallic acid equivalent (GAE)/g. The 3-deoxyanthocyanidins were measured using the colorimetric method of Fuleki and Francis (1968). The absorbance of each acidified methanol extract was read at 485 nm and 3-deoxyanthocyanidin concentrations were expressed as abs/mL/g. Condensed tannins were measured using the modified vanill/HCl assay as described by Price et al. (1978). Results were expressed as mg catechin equivalents (CE)/g. Details on the aforementioned procedures are found in Dykes et al. (2005). All analyses were done in triplicate.

2.4. NIR scanning

NIR spectra were obtained using a FOSS XDS NIR spectrometer (FOSS North America, Eden Prairie, MN, USA). Each whole grain sample was packed (2.54 cm deep) into a rectangular-shaped sample cup (15.24 cm × 3.81 cm × 5.08 cm, length × width, depth) and was scanned twice using the wavelength range of 400–2500 nm. Collection of the spectra data was done using the ISIscan software (Version 3.10.05935) set at 32 readings per scan and stored at 2 nm intervals. The average of two scans was used for the equation development.

The calibration model was developed using the WinISI software (version 4.0.03770) that supports the modified partial least-squares (mPLS) method which creates a calibration equation for each of the variables separately. In the development process, several equations were tested which varied in derivative math treatment. After evaluating the quality parameters of each equation [coefficient of determination (r²), standard error in cross-validation (SECV), and coefficient of determination for cross-validation (1-VR values)], the equation with 2,6,4,1 math treatment provided the best results with the highest r² and 1-VR and lowest SECV values.

Accuracy, precision, and robustness of the method were determined using 63 samples in a validation set (Table 1). Like the samples used for the calibration set, samples varied based on pericarp color, presence of a pigmented testa and spreader gene, pericarp thickness, endosperm color, and secondary plant color. Scanning and analysis were as previously described.

3. Results and discussion

3.1. Range of variation in the sorghum genotypes

Significant variation for all three traits existed among the samples, and both the validation and calibration sets were comparable.
in both mean and range (Table 2). For the calibration set, reference values for total phenols ranged from 1.35 to 37.73 mg GAE/g with a mean of 8.86 mg GAE/g. For the validation set, reference values for total phenols ranged from 1.31 to 22.71 mg GAE/g with a mean of 6.29 mg GAE/g. The higher total phenol levels were from genotypes with a pigmented testa, which contained condensed tannins.

For condensed tannins, high variability was also observed in both calibration and validation sets. Condensed tannin levels in the calibration set ranged from ND to 101.03 mg CE/g with a mean of 14.83 mg CE/g. For the validation set, condensed tannin levels ranged from ND to 51.10 mg CE/g with a mean of 7.10 mg CE/g. The non-detectable (ND) levels of condensed tannins were due to the absorbance readings being less than 0.001. The highest condensed tannin levels were in genotypes with a pigmented testa and spreader gene.

The sorghum 3-deoxyanthocyanidin concentrations also varied in both sample sets. For the calibration set, 3-deoxyanthocyanidin concentrations ranged from 0.36 to 229.64 abs/mL/g with a mean of 43.11 abs/mL/g. For the validation set, concentrations ranged from 0.58 to 152.69 abs/mL/g with a mean of 20.49 abs/mL/g. As was previously reported (Dykes et al., 2005, 2009), the highest 3-deoxyanthocyanidin levels were from black sorghum genotypes, while the lowest levels came from sorghums with tan secondary plant color. All these data suggested high variability among genotypes which were acceptable to perform calibration curve generation and validation.

The variation present within this set of sorghum germplasms demonstrated that this set of germplasm represented the variation that had been reported previously. Based on previous reports (Dlamini et al., 2007, Dykes and Rooney, 2006, Dykes et al., 2005, 2013), total phenol concentrations greater than 15 mg GAE/g, condensed tannin concentrations above 40 mg CE/g and 3-deoxyanthocyanidin concentrations higher than 100 abs/mL/g were considered high. In the current set of samples, there were numerous entries at both the non-detectable to low levels and at the extreme high range of reported values for these traits. Not only did this extend the reported ranges for these compounds, but they also served as a diverse and large sample set for the development of an NIR curve to estimate these concentrations even more efficiently.

### 3.2. Visible and NIR (Vis/NIR) spectra

Typical Vis/NIR spectra of black, red, lemon-yellow, and white sorghums are illustrated in Figure 1. In the visible region of the spectrum (400–700 nm), absorption was affected by the pericarp color of whole sorghum grains where absorptions increased as pericarp color darkened. In the near-infrared region of the spectra, absorption bands at around 1450 nm (OH stretch first overtone) and 1930 nm (OH stretch and deformation combination) were due to water (Cozzolino et al., 2008, Zhang et al., 2008). Phenolic compounds (i.e. flavonoids, condensed tannins) absorb at bands 1415–1512 nm, 1650–1750 nm, and 1955–2035 nm (Cozzolino et al., 2004, Zhang et al., 2008). NIR spectra trends for non-tannin (i.e. Tx2953) and tannin sorghums (i.e. Black PI Tall) did not differ.

### 3.3. Calibration and validation of the NIR curves

Calibration curves for each trait were developed from 224 varieties of varying genotypes. Statistics for the calibration curves are listed in Table 3. Calibration curve correlations (NIR prediction vs. reference value) for total phenols, condensed tannins, and 3-deoxyanthocyanidins were $r = 0.98, 0.97,$ and 0.99, respectively.

The lower correlation for condensed tannins was due to the NIR curve occasionally predicting a low level of condensed tannins.

### Table 2. Means and ranges of concentrations of total phenols, condensed tannins and 3-deoxyanthocyanidins from both calibration and validation sets of sorghum samples.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Calibration set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>Range</td>
</tr>
<tr>
<td>Total phenols (mg GAE/g)</td>
<td>224</td>
<td>1.35–37.73</td>
</tr>
<tr>
<td>Condensed tannins (mg CE/g)</td>
<td>224</td>
<td>ND–101.03</td>
</tr>
<tr>
<td>3-Deoxyanthocyanidins (abs/mL/g)</td>
<td>224</td>
<td>0.36–229.64</td>
</tr>
</tbody>
</table>

GAE, gallic acid equivalents; CE, catechin equivalents; $n$, number of varieties; ND, not detected (absorbance was less than 0.001); SD, standard deviation.

### Table 3. NIR calibration curve statistics for the prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghums.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>$n$</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
<th>SEC</th>
<th>1-VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total phenols (mg GAE/g)</td>
<td>214</td>
<td>ND–27.98</td>
<td>8.28</td>
<td>6.57</td>
<td>1.03</td>
<td>1.41</td>
</tr>
<tr>
<td>Condensed tannins (mg CE/g)</td>
<td>212</td>
<td>ND–65.92</td>
<td>13.60</td>
<td>17.44</td>
<td>3.93</td>
<td>5.35</td>
</tr>
<tr>
<td>3-Deoxyanthocyanidins (abs/mL/g)</td>
<td>212</td>
<td>ND–192.73</td>
<td>39.62</td>
<td>51.03</td>
<td>7.27</td>
<td>9.97</td>
</tr>
</tbody>
</table>

GAE, gallic acid equivalents; CE, catechin equivalents; $n$, number of varieties; ND, not detected (absorbance was less than 0.001); SD, standard deviation; SEC, standard error of calibration; SECV, standard error in cross-validation; 1-VR, coefficient of determination for cross-validation.
present in non-tannin sorghums. Correlation plots of the NIR predictions of the validation set revealed a positive relationship between NIR estimates and reference values (Figure 2). The correlation for total phenols was $r = 0.93$, which indicated the method can effectively predict total phenols. Correlations for condensed tannins and 3-deoxyanthocyanidins were slightly lower at $r = 0.81$ and 0.82, respectively, which indicated greater variation between the reference values and NIR predictions. The reduced correlations were typically observed in samples with low or non-detectable concentrations of condensed tannins or 3-deoxyanthocyanidins, thus making them more difficult to quantify. Genetic and environmental variations as well as genotype × environment interactions will influence phenolic compound concentrations at all levels (Dykes et al., 2011, Taleon et al., 2012).

![Figure 2](image)

Figure 2. NIR prediction versus reference value for sorghum (a) total phenols (mg GAE/g), (b) condensed tannins (mg CE/g), and (c) 3-deoxyanthocyanidins (abs/mL/g) from the validation set ($n = 63$).

Accuracy at the lower concentrations was also influenced by the wide range of wavelengths included in the calibration curves. For improved accuracy in estimating these values, it may be useful to develop independent curves for low concentration germplasm (i.e., non-tannin white sorghum) as well as for high concentration germplasm (i.e., tannin or dark-colored sorghum). To develop such curves will require additional germplasm screening and testing.

For this study, the goal was to identify methodology applicable across a wide range of germplasm grown in different environments. Herein, the correlations had value as ANOVA showed $p$-values less than 0.001 and the estimates were within a range of variation. This indicated that NIR should be effective for pre-screening to establish relative (low, medium and high) concentrations of these compounds, which can then be followed by further analysis to confirm their exact values. Finally, principle component analysis (PCA) showed that both total phenols and condensed tannins had similar prediction patterns and they were highly correlated, suggesting that an attempt to breed for increment in values of total phenols will increase the value of the condensed tannins and vice versa (Figure 3).

![Figure 3](image)

Figure 3. Principal component analysis (PCA) score plot of calibration and validation sets.

4. Conclusion

The resultant calibration curves indicate that NIR spectroscopy has good potential to predict total phenol, condensed tannin, and 3-deoxyanthocyanidin concentrations in whole grain sorghum. This method has many advantages over the conventional methods (i.e. colorimetric assays, HPLC, etc.) because it is quick, inexpensive, non-destructive, and does not require the use of chemical solvents. Accuracy of the methodology was established using a validation set which indicated that estimates of total phenols were very good; estimates of condensed tannins, and 3-deoxyanthocyanidin concentrations were slightly less accurate. While not perfect, the estimates establish a potential range in the composition. In a breeding program, these estimates can be used effectively for rapid screening to eliminate those that do not possess the trait of interest. Those identified as desirable can then be further analyzed to confirm the initial estimates. Ultimately, this method will benefit plant breeders to rapidly screen sorghums for high levels of health-promoting components.
Acknowledgments — We appreciate long-term support from USDA-ARS, USAID INTSORMIL CRSP, and the United Sorghum Producers Checkoff.

References


