

# Genome-Wide Association Study of Grain Polyphenol Concentrations in Global Sorghum [*Sorghum bicolor* (L.) Moench] Germplasm

Davina H. Rhodes,<sup>\*,†</sup> Leo Hoffmann, Jr.,<sup>‡</sup> William L. Rooney,<sup>‡</sup> Punna Ramu,<sup>§,||</sup> Geoffrey P. Morris,<sup>⊥</sup> and Stephen Kresovich<sup>¶</sup>

<sup>†</sup>Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, United States

<sup>‡</sup>Department of Soil & Crop Sciences, Texas A&M University, College Station, Texas 77843, United States

<sup>§</sup>International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502 324, Andhra Pradesh, India

<sup>⊥</sup>Department of Agronomy, Kansas State University, Manhattan, Kansas 66506, United States

<sup>¶</sup>Department of Genetics and Biochemistry, Clemson University, Clemson, South Carolina 29634, United States

## Supporting Information

**ABSTRACT:** Identifying natural variation of health-promoting compounds in staple crops and characterizing its genetic basis can help improve human nutrition through crop biofortification. Some varieties of sorghum, a staple cereal crop grown worldwide, have high concentrations of proanthocyanidins and 3-deoxyanthocyanidins, polyphenols with antioxidant and anti-inflammatory properties. We quantified total phenols, proanthocyanidins, and 3-deoxyanthocyanidins in a global sorghum diversity panel ( $n = 381$ ) using near-infrared spectroscopy (NIRS), and characterized the patterns of variation with respect to geographic origin and botanical race. A genome-wide association study (GWAS) with 404,628 SNP markers identified novel quantitative trait loci for sorghum polyphenols, some of which colocalized with homologues of flavonoid pathway genes from other plants, including an orthologue of maize (*Zea mays*) *Pr1* and a homologue of *Arabidopsis* (*Arabidopsis thaliana*) *TT16*. This survey of grain polyphenol variation in sorghum germplasm and catalog of flavonoid pathway loci may be useful to guide future enhancement of cereal polyphenols.

**KEYWORDS:** *Sorghum bicolor*, cereal, grain, polyphenol, flavonoid, proanthocyanidin, 3-deoxyanthocyanidin, condensed tannins, GWAS, QTL

## INTRODUCTION

Polyphenols are a large diverse group of phytochemicals that include phenolic acids, stilbenes, lignans, isoflavonoids, and flavonoids.<sup>1</sup> All flavonoids share a common C6–C3–C6 backbone structure but differ in their oxidation level, glycosylation, acylation, and hydroxyl and methyl substitutions, allowing for an enormous variety of structure and function.<sup>2</sup> In plants, flavonoid secondary metabolites are involved in growth, pigmentation, pollination, and defense against pathogens, predators, and physical factors.<sup>3</sup> In humans, dietary flavonoids are thought to act as antioxidants and signaling molecules, and their consumption is correlated with lower incidence of cardiovascular disease, cancer, type II diabetes, neurodegenerative disease, and other chronic illnesses.<sup>4</sup> Most plant-based foods contain flavonoids, making them some of the most ubiquitous polyphenols in the human diet. Polymerization of flavonoids yields complex compounds including proanthocyanidins, flavonoid polymers predominantly composed of flavan-3-ols, which are abundant in food plants. Proanthocyanidins contribute to the astringency and bitterness found in foods such as wine, cocoa, beans, and fruits, but they are not present in most commonly consumed vegetables and cereals.<sup>5</sup> They are also often considered antinutrients due to their nutrient binding capacity, especially to proteins and iron.<sup>6</sup> In the past decade, however, potential health protective effects of proanthocyanid-

dins have been studied extensively, with particular focus on their contributions to observed health benefits of grape and cranberry.<sup>7</sup>

Sorghum is one of the world's major cereal crops and a dietary staple for more than 500 million people in sub-Saharan Africa and Asia.<sup>8</sup> In the United States, it is primarily used as animal feed, but it is becoming more popular in food products due to a rise in demand for specialty grains, especially those that are gluten-free.<sup>9–12</sup> Domesticated sorghum has been classified into five major races (bicolor, guinea, caudatum, kafir, and durra) and 10 intermediate races (all combinations of the major races), based on morphological differences.<sup>13</sup> Two of the major polyphenol compounds in sorghum grain are proanthocyanidin and 3-deoxyanthocyanidin. Consumption of these two polyphenols has been correlated with several health benefits including protection against oxidative damage, inflammation, obesity, and diabetes.<sup>14</sup> Proanthocyanidins are constitutively expressed, while 3-deoxyanthocyanidins are phytoalexins, expressed only in response to fungal infection.<sup>15,16</sup> Sorghum grain is the only known dietary source of 3-

Received: April 7, 2014

Revised: September 30, 2014

Accepted: October 1, 2014

Published: October 1, 2014

deoxyanthocyanidins, which otherwise have only been found in the flowers of *sinningia* (*Sinningia cardinalis*), the silk tissues of maize (*Zea mays*), and the stalks of sugar cane (*Saccharum* sp.).<sup>17–19</sup>

In sorghum grains, polyphenol compounds can be found in the pericarp (outer seed coat) and the testa (inner layer of tissue between the pericarp and the endosperm). A number of classical loci identified by their effects on grain color and testa presence control the presence or absence of polyphenol compounds in sorghum.<sup>20</sup> Genotypes with dominant alleles at the B1 and B2 loci have proanthocyanidins in the testa. Genotypes with a dominant allele at the spreader (S) locus, as well as dominant alleles at the B1 and B2 loci, have proanthocyanidins in both the pericarp and the testa, often, but not always, resulting in a brown appearance to the grain. The base pericarp color is red, yellow, or white, and these colors are controlled by the R and Y loci. The S locus and additional loci, such as intensifier (I) and mesocarp thickness (Z), modify the base pericarp color, resulting in a range of colors from brilliant white to black with various shades of red, yellow, pink, orange, and brown among sorghum genotypes (see Figure 1). Using mutants for seed color traits, the



**Figure 1.** Natural variation in sorghum grain color. Three accessions (with three seeds of each accession) of grain with the appearance of (A) brown (PI597965, PI533927, PI35038), (B) white (PI533755, PI533845, PI534028), (C) yellow (PI659691, PI656011, PI533776), and (D) red (PI576418, PI534047, PI564165) pericarps. The outer coat has been scraped off of some samples, revealing the presence or absence of a pigmented testa.

biochemical and regulatory pathways underlying flavonoids and flavonoid products have been almost completely elucidated in *Arabidopsis* and maize, and extensively studied in other species (Table S1 in the Supporting Information).<sup>21</sup> Therefore, homology can be used as a guide to discover genes involved in the sorghum flavonoid pathway. The gene underlying the B2 locus was recently cloned and designated *Tannin1*, along with two nonfunctional alleles of *Tannin1*, *tan1-a* and *tan1-b*.<sup>22</sup> *Tannin1* encodes a WD40 protein homologous to the *Arabidopsis* proanthocyanidin regulator *transparent testa glabra1* (TTG1). The gene underlying the Y locus has also been cloned and designated *Yellow seed1*. *Yellow seed1* encodes a MYB protein, orthologous to the maize 3-deoxyanthocyanidin regulator *P1*, that is needed for accumulation of 3-deoxyanthocyanidins in the sorghum pericarp.<sup>23</sup> The R locus has been mapped to chromosome 3 between 57 and 59 Mb, and the Z locus has been mapped to chromosome 2 between 56 and 57 Mb,<sup>24</sup> but the underlying genes have not been identified.

While the genetic controls of polyphenol presence/absence have been well-studied using mutant lines and nonfunctional

polymorphisms, there has been little study of quantitative natural variation in polyphenols.<sup>25</sup> Polyphenol nonfunctional mutations were strongly selected during cereal domestication, when bitter tasting and/or dark compounds were partly or completely lost in most cereals, including wheat, rice, and maize.<sup>26</sup> However, sorghum provides a valuable resource for polyphenol diversity, as adaptation to different environments has led to extensive phenotypic and genetic diversity in the crop.<sup>13,27</sup> This diversity can be useful for biofortification and crop improvement (e.g., desirable traits can be bred into existing elite varieties), but quantitative phenotyping is needed to identify alleles responsible for quantitative trait variation in grain polyphenols (reviewed by Flint-Garcia<sup>28</sup>). The goals of this study were to quantify the natural variation of two of the major sorghum grain polyphenols (proanthocyanidins and 3-deoxyanthocyanidins) and to identify single-nucleotide polymorphisms (SNPs) that are associated with low or high polyphenol concentrations using genome-wide association studies (GWAS). GWAS are used to map the genomic regions underlying phenotypic variation (known as quantitative trait loci) by scanning the genome for statistical associations between genetic variation and phenotypic variation.<sup>29</sup> In contrast to the biparental linkage mapping approach, GWAS takes advantage of historical recombinations in a diverse panel and linkage disequilibrium between causal variants and nearby SNP markers. Although it has been used extensively to identify putative genetic controls of human disease,<sup>30</sup> it is a relatively new but promising tool in plant genomics.<sup>27,31,32</sup> Here we present a survey of the quantitative natural variation of polyphenols in a diverse worldwide panel of sorghum and a catalog of flavonoid-associated loci across the sorghum genome.

## MATERIALS AND METHODS

**Plant Materials.** We investigated a total of 381 sorghum accessions, comprising 308 accessions from the Sorghum Association Panel (SAP)<sup>33</sup> and an additional 73 accessions selected based on presence of a pigmented testa using the U.S. National Plant Germplasm System's Germplasm Resources Information Network (GRIN).<sup>34</sup> The SAP includes accessions from all major cultivated races and geographic centers of diversity in sub-Saharan Africa and Asia, as well as important breeding lines from the United States. The 73 additional accessions were included to increase the proportion of accessions with high proanthocyanidins.

Seeds were obtained through GRIN and planted in late April 2012 at Clemson University Pee Dee Research and Education Center in Florence, SC. A 2-fold replicated complete randomized block design was used. Panicles from each plot were collected at physiological maturity (signified by a black layer at the base of the seed that normally forms about 35 days after anthesis). Due to differences in maturity among these accessions, harvest occurred between September and October. Once harvested, panicles were air-dried in a greenhouse and then mechanically threshed, and any remaining glumes were removed with a wheat head thresher (Precision Machine Company, Lincoln, NE).

**Phenotyping.** Twenty grams of cleaned whole grain from one replicate was scanned with a FOSS XDS spectrometer (FOSS North America, Eden Prairie, MN, USA) at a wavelength range of 400–2500 nm. To determine reproducibility, duplicates on a subset of 218 accessions available from replicate plots were also scanned. The NIR reflectance spectra were recorded using the ISIscan software (Version 3.10.05933) and converted to estimates of total phenol, proanthocyanidin, and 3-deoxyanthocyanidin concentrations. The spectrometer, software, and calibration curves used in this study were recently described.<sup>35</sup> Samples with unusual reflectance were visually inspected, and near-infrared spectroscopy (NIRS) was repeated. Seventeen samples were removed from further analysis either because they

contained mixed grain (mixed size, shape, or color) or because their readings were outside the range of the available NIRS calibration curve. Total phenol, proanthocyanidin, and 3-deoxyanthocyanidin data are expressed as mg gallic acid equivalents (GAE)/g, mg catechin equivalents (CE)/g, and absorbance (abs)/mL/g, respectively. These were the units used in creating the calibration curves, which measured total phenols with the Folin–Ciocalteu method, 3-deoxyanthocyanidins with the colorimetric method of Fuleki and Francis, and proanthocyanidins with the modified vanillin/HCl assay.<sup>35</sup> For the purposes of this study, we use a cutoff of greater than 10.00 mg CE/g to define proanthocyanidin-containing varieties and greater than 50.00 abs/mL/g to define 3-deoxyanthocyanidin-containing varieties.

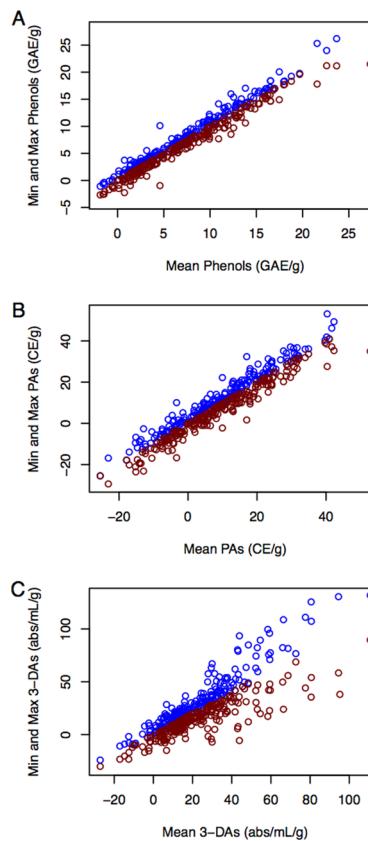
Visual appearance of grain was classified independently by two people by visually scoring three seeds per accession as white, yellow, red, or brown. Testa presence was identified with three seeds per accession by cutting a thin layer off the pericarp and examining under a dissecting microscope. The total grain weight of 100 seeds per accession was recorded.

**Genomic Analysis.** Genotypes were available for the 324 accessions that were part of the SAP.<sup>27</sup> Genotyping-by-sequencing (GBS) was performed for the 73 additional accessions by the Institute for Genomic Diversity using the methods by Elshire et al.<sup>36</sup> Briefly, we provided seeds of the 73 additional accessions (the same seeds obtained from GRIN that we used to grow our panel) to the Institute for Genomic Diversity, where the following work was performed: Seedlings were grown to obtain tissue, DNA was isolated using the Qiagen DNeasy Plant kit, genomic DNA was digested individually using ApeKI, 96X multiplexed GBS libraries were constructed, and DNA sequencing was performed on the Illumina Genome Analyzer IIx. To extract SNP genotypes from sequence data, the GBS pipeline 3.0 in the TASSEL software package<sup>37</sup> was used, with mapping to the BTx623 sorghum reference genome.<sup>37</sup> Missing genotype calls were imputed using the FastImputationBitFixedWindow plugin in TASSEL 4.0.<sup>38</sup>

GWAS was carried out on 404,628 SNP markers, using the statistical genetics package Genome Association and Prediction Integrated Tool (GAPIT)<sup>39</sup> with both a general linear model (GLM) and a mixed linear model (MLM) with kinship. In a previous study we found that an MLM<sup>40</sup> with kinship (K), which controls for relatedness among the accessions in the panel, performs well to identify causative loci for sorghum polyphenols.<sup>41</sup> Bonferroni correction (family wise *P*-value of 0.01,  $P < 10^{-6}$ ) was used to identify significant associations. Pseudoheritability (proportion of phenotypic variation explained by genotype) was estimated from the kinship (K) model in GAPIT<sup>42</sup> as the *R*-squared of a model with no SNP affects. A previously developed a priori candidate gene list was used, and 35 additional candidate genes were added (see Supporting Information).<sup>41</sup>

## RESULTS

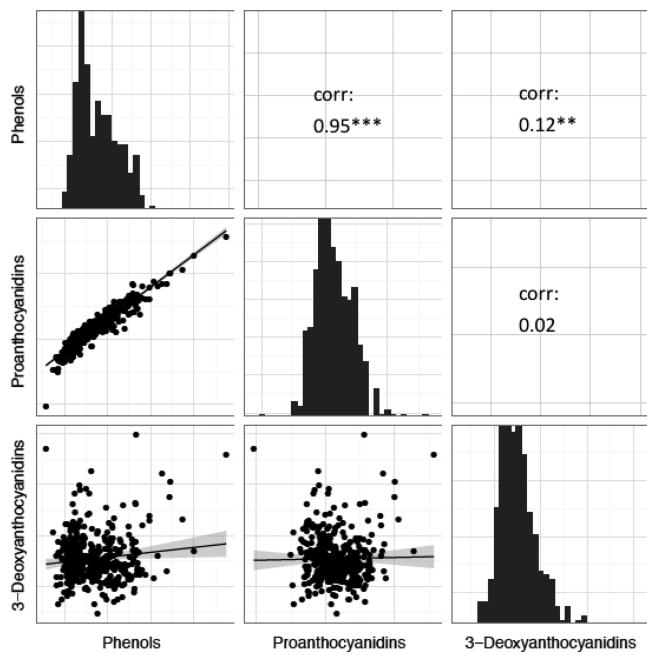
**Quantitative Variation in Grain Polyphenols.** We first sought to determine the reliability of the NIRS estimates across the diverse material in the panel. Phenotypic variation for grain polyphenol concentrations was determined using a diverse association panel with 381 accessions (Figure 2). The standard deviation between the duplicates was similar across all concentrations of polyphenols ( $r^2 = 0.06$ ,  $P = 0.0001$ ) and proanthocyanidins ( $r^2 = 0.01$ ,  $P = 0.12$ ), with an average difference of 47% and 4%, respectively. However, the standard deviation between the 3-deoxyanthocyanidin duplicates becomes much larger for samples with higher 3-deoxyanthocyanidin concentrations ( $r^2 = 0.32$ ,  $P = 10^{-17}$ ), with an average difference of 72% (Figure 2C). To determine if the NIRS measurements of proanthocyanidin concentration were concordant with the known distribution of testa and *tan1-a* nonfunctional allele,<sup>22</sup> we plotted proanthocyanidin concentration of accessions with or without a pigmented testa (Figure



**Figure 2.** Phenotypic variation of grain polyphenol concentrations in 381 sorghum varieties. Samples are ordered on the *x*-axis according to their mean value for the accession. The observed value for each replicate is given on the *y*-axis, with the higher value of the duplicates in red and the lower value of the duplicates in blue. (A) Total polyphenols, (B) proanthocyanidins (PAs), and (C) 3-deoxyanthocyanidins (3-DAs).

S1A in the Supporting Information), and accessions with the wild-type *Tannin1* allele or the *tan1-a* allele (Figure S1B in the Supporting Information). As expected, the absence of a testa and presence of *tan1-a* were primarily found in accessions containing less than 10 mg CE/g of proanthocyanidins. The mean proanthocyanidin concentrations in accessions with a pigmented testa were significantly higher than in accessions without a pigmented testa (18.17 versus 1.45 mg CE/g;  $P = 10^{-17}$ ), and the mean proanthocyanidin concentrations in accessions with the wild-type *Tannin1* were significantly higher than in accessions with *tan1-a* (12.28 versus 0.86 mg CE/g;  $P = 10^{-11}$ ).

Next we investigated the range of total phenol, proanthocyanidin, and 3-deoxyanthocyanidin concentrations and their covariation with each other and grain weight (Figure 3). Overall, proanthocyanidins were detected in 55% of the samples, while only 13% contained 3-deoxyanthocyanidins, and only 6% contained both polyphenols. The mean total polyphenol concentration was 7.00 mg (GAE)/g, the mean proanthocyanidin concentration was 7.73 mg CE/g, and the mean 3-deoxyanthocyanidin concentration was 27.40 abs/mL/g (Table 1 and Figure 3). Pearson's correlations were calculated between total phenols, proanthocyanidins, and 3-deoxyanthocyanidins. There was no significant correlation between proanthocyanidins and 3-deoxyanthocyanidins ( $0.02$ ,  $P = 0.7$ ), consistent with independent genetic control. In contrast,



**Figure 3.** Relationship within and between grain polyphenol traits in a global sorghum germplasm collection. The center diagonal presents histograms of the mean concentrations of each trait ( $n = 381$ ). The lower corner contains scatter plots with regression lines showing the relationships between the traits. The upper corner shows Pearson's correlations between the traits. Units are mg GAE/g for total phenols, mg CE/g for proanthocyanidins, and abs/mL/g for 3-deoxyanthocyanidins.

**Table 1. Polyphenol Concentrations in 373 Sorghum Varieties**

constituent	mean	range	SD
total phenols (mg GAE/g)	7.00	ND–37.46	5.92
proanthocyanidins (mg CE/g)	7.73	ND–78.51	15.45
3-deoxyanthocyanidins (abs/mL/g)	27.40	ND–149.21	24.05

there were a strong positive correlation between total phenols and proanthocyanidins ( $0.95, P < 10^{-15}$ ) and a weak positive correlation between total phenols and 3-deoxyanthocyanidins ( $0.12, P = 0.02$ ). Variance in proanthocyanidins accounted for 90% of all the variance in total phenols (Figure 3). Since the seed coat (pericarp and testa) contains most of the polyphenols in the grain, and the ratio of seed coat (surface area) to endosperm is generally greater in smaller grains, we wondered

if differences in grain size might be underlying variation in polyphenol concentrations. In other words, are high grain polyphenol concentrations limited to small-grain varieties, which have a high proportion of seed coat to endosperm? No significant correlation was found between grain weight and either proanthocyanidins ( $-0.02, P = 0.7$ ) or 3-deoxyanthocyanidins ( $-0.02, P = 0.7$ ), and a small negative correlation was found between grain weight and total polyphenols ( $-0.10, P = 0.04$ ). Pseudoheritability was 81.7% for proanthocyanidins and 66.5% for 3-deoxyanthocyanidins.

**Population Structuring of Polyphenol Concentrations.** To determine the distribution of polyphenol traits with respect to global genetic diversity, we conducted a principal component analysis and highlighted the variation in polyphenol concentration (Figures S2A and S2B in the Supporting Information), as well as morphological races (Figure S2C in the Supporting Information). At least some high proanthocyanidin accessions were found in most subpopulations, whereas high 3-deoxyanthocyanidin accessions were more restricted (Table 2). Bicolor (21.18 mg CE/g) and guinea-caudatum (17.89 mg CE/g) had the highest mean concentration of proanthocyanidins. Caudatum had moderate concentrations (13.20 mg CE/g), and the other botanical races and intermediate groups showed an average less than 10.00 mg CE/g. The highest mean concentrations of 3-deoxyanthocyanidins were found in bicolor-durra (36.95 abs/mL/g) and guinea (35.63 abs/mL/g) accessions (Table 2). We also determined the mean concentrations by country to better understand the geographic patterns for sorghum polyphenols (Table 3). Accessions from Uganda (19.03 mg CE/g) had the highest mean proanthocyanidin concentrations, accessions from South Africa (12.23 mg CE/g) and Sudan (10.33 mg CE/g) had moderate concentrations, while accessions from the other countries showed an average less than 10.00 mg CE/g. The highest mean concentrations of 3-deoxyanthocyanidins were found in accessions from Nigeria (36.39 abs/mL/g) and Ethiopia (32.87 abs/mL/g).

**Genome-Wide Association Studies.** To investigate the genetic basis of natural variation in sorghum grain polyphenols, we conducted GWAS using 404,628 SNP markers. We were able to obtain genotype data for 373 out of the 381 phenotyped accessions. As a data quality check, we first collapsed the quantitative proanthocyanidin data to qualitative (presence or absence) data, and were able to repeat findings from previous GWAS and linkage studies (Figures S3 and S4 and Tables S2 and S3 in the Supporting Information). Next, to identify novel alleles associated with quantitative variation of proanthocyanidins,

**Table 2. Polyphenol Concentrations by Race**

race <sup>a</sup>	n	total phenols (mg GAE/g)		PA (mg CE/g)		3-DA (abs/mL/g)	
		mean	range	mean	range	mean	range
bicolor	15	$13.68 \pm 6.69$	0.74–24.49	$21.18 \pm 17.68$	ND–50.16	$26.91 \pm 33.65$	ND–102.96
bicolor-durra	19	$6.59 \pm 4.28$	ND–13.38	$3.89 \pm 12.06$	ND–23.35	$36.95 \pm 28.24$	1.30–113.42
caudatum	86	$9.08 \pm 5.86$	ND–27.32	$13.20 \pm 14.15$	ND–52.83	$28.22 \pm 21.06$	ND–110.73
caudatum-kafir	20	$6.27 \pm 5.41$	ND–15.68	$7.00 \pm 15.13$	ND–31.98	$26.65 \pm 16.87$	6.70–58.25
durra	15	$2.17 \pm 3.61$	ND–11.68	ND	ND–17.64	$22.17 \pm 21.33$	ND–71.10
guinea	11	$1.95 \pm 5.25$	ND–15.44	ND	ND–33.45	$35.63 \pm 36.88$	0.93–135.34
guinea-caudatum	15	$10.01 \pm 3.13$	2.54–15.87	$17.89 \pm 9.76$	ND–34.92	$19.72 \pm 15.69$	0.40–60.10
kafir	29	$6.02 \pm 4.05$	1.32–14.71	$6.50 \pm 10.20$	ND–28.72	$17.59 \pm 20.65$	ND–94.49

<sup>a</sup>If a race contained a small sample size (less than 10 accessions), it was not included in this analysis. PA, proanthocyanidins; 3-DA, 3-deoxyanthocyanidins; ND, not detected (absorbance was less than 0.001).

Table 3. Polyphenol Concentrations by Geographic Origin

country <sup>a</sup>	n	total phenols (mg GAE/g)		PA (mg CE/g)		3-DA (abs/mL/g)	
		mean	range	mean	range	mean	range
Uganda	44	10.99 ± 5.17	1.17–27.32	19.03 ± 12.02	ND–52.83	27.37 ± 20.8	1.30–110.73
South Africa	31	9.11 ± 5.21	1.11–20.63	12.23 ± 12.37	ND–43.75	13.52 ± 14.1	ND–38.82
Sudan	31	7.50 ± 3.34	ND–14.67	10.33 ± 8.93	ND–25.26	27.15 ± 15.1	4.13–60.10
Nigeria	21	5.0 ± 6.46	ND–24.49	1.21 ± 21.36	ND–50.16	36.39 ± 35.8	ND–135.34
Ethiopia	29	5.71 ± 5.43	ND–15.94	1.53 ± 13.13	ND–23.53	32.87 ± 21.1	ND–77.59
India	21	3.90 ± 5.09	ND–16.98	ND	ND–32.13	28.74 ± 28.7	ND–113.42
USA	71	5.09 ± 5.25	ND–29.93	3.6 ± 12.55	ND–63.80	27.50 ± 24.2	ND–95.20

<sup>a</sup>If a country contained a small sample size (less than 10 accessions), it was not included in this analysis. PA, proanthocyanidins; 3-DA, 3-deoxyanthocyanidins; ND, not detected (absorbance was less than 0.001).

dins, we conducted a GWAS on the 373 accessions (Figure 4; Table S4 in the Supporting Information). A GLM identified 3,272 significant SNPs (Figure 4A), while the MLM identified 24 significant SNPs after accounting for population structure (Figure 4B). The genomic locations of the association peaks were generally similar between methods. A peak on chromosome 4 at ~61 Mb colocalized with *Tannin1* (Sb04g031730), as well as three a priori candidate genes in the region: a putative *Zm1* homologue (Sb04g031110), a putative *TTG1* homologue (Sb04g030840), and a putative *TT16* homologue (Sb04g031750) (Figure 4C). The GLM identified a peak at 58.6 Mb on chromosome 7 (S7\_58603858;  $P < 10^{-15}$ ), which was not present in the MLM.

In order to reduce the effects of known *Tannin1* nonfunctional alleles and identify additional quantitative loci, samples with the *tan1-a* and *tan1-b* alleles were removed and a GWAS was conducted on the remaining samples (Figure 5 and Table S5 in the Supporting Information). The GLM identified 2,641 significant SNPs (Figure 5A). The association peak on chromosome 7 was again identified in the GLM and not in the MLM (Figure 5B). Additionally, there was a peak on chromosome 2 around 8 Mb (S2\_8258226;  $P < 10^{-11}$ ) identified in the GLM, near a putative *TT8* homologue (Sb02g006390). Both the GLM and the MLM identified a peak on chromosome 4, again around 61 Mb, and another peak on chromosome 4 between 53 Mb and 55 Mb, close to an F3'H *Pr1* coorthologue.

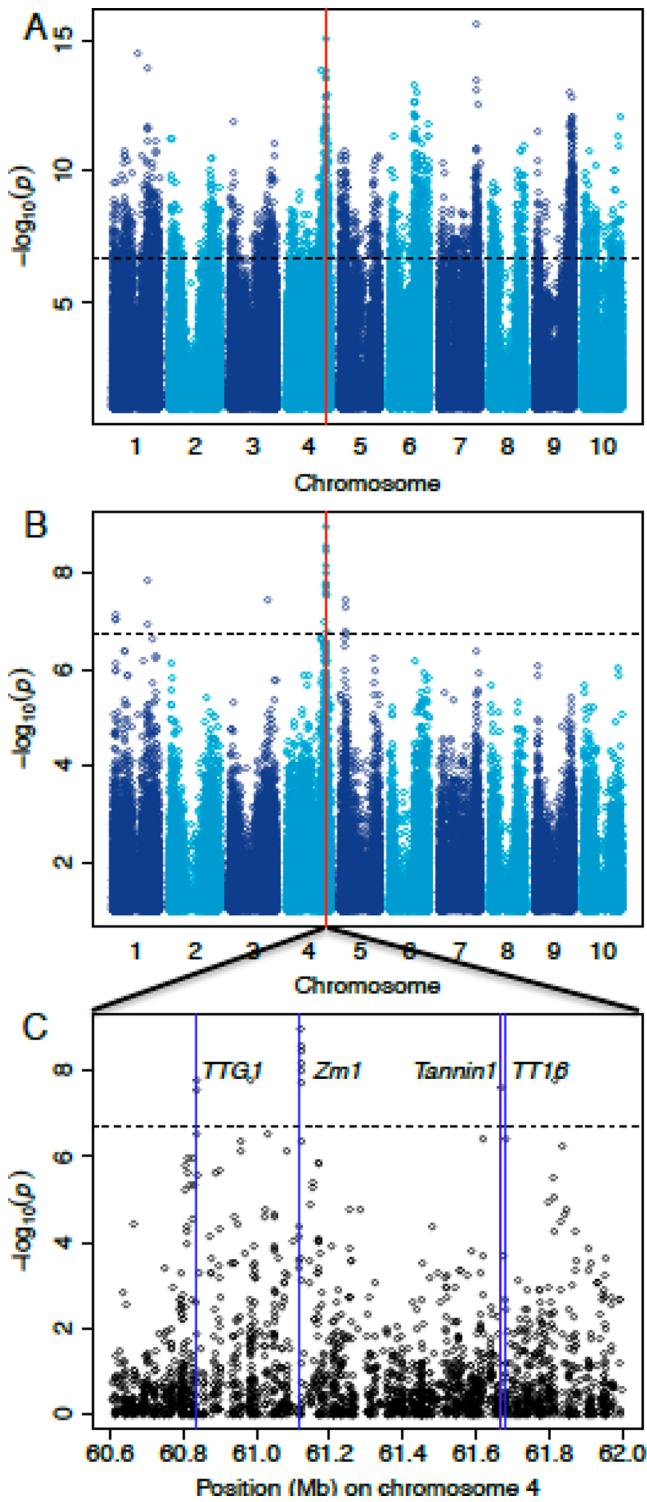
To further map loci controlling quantitative proanthocyanidin variation, we ran a GWAS only on samples that contained proanthocyanidins (greater than 10.00 mg CE/g) and/or had a visible pigmented testa (Figure 6 and Table S6 in the Supporting Information). With this subset, there were 676 significant SNPs identified in the GLM, but association peaks were more diffuse (Figure 6A). The most significant SNP was on chromosome 6 (S6\_56992521,  $P < 3 \times 10^{-10}$ ) near a *TT16* a priori candidate (Sb06g028420). The MLM identified two significant SNPs, with a peak on chromosome 4, again around 61 Mb, and another peak on chromosome 4 between 53 Mb and 55 Mb (Figure 6B). Both the GLM and the MLM identified significant SNPs around 61.1 Mb on chromosome 1, which is near *yellow seed1*.

Next, a GWAS was conducted to identify genetic controls of 3-deoxyanthocyanidin variation among the 373 accessions (Figure 7 and Table S7 in the Supporting Information). The GLM identified 233 significant SNPs, with distinct association peaks on chromosomes 3 and 4 (Figure 7A). The peak on chromosome 3 was between 71 and 72 Mb and colocalized with a gene (Sb03g045170) homologous to both *TT18* (ANS) and *TT6* (F3H). The peak on chromosome 4 was between 53

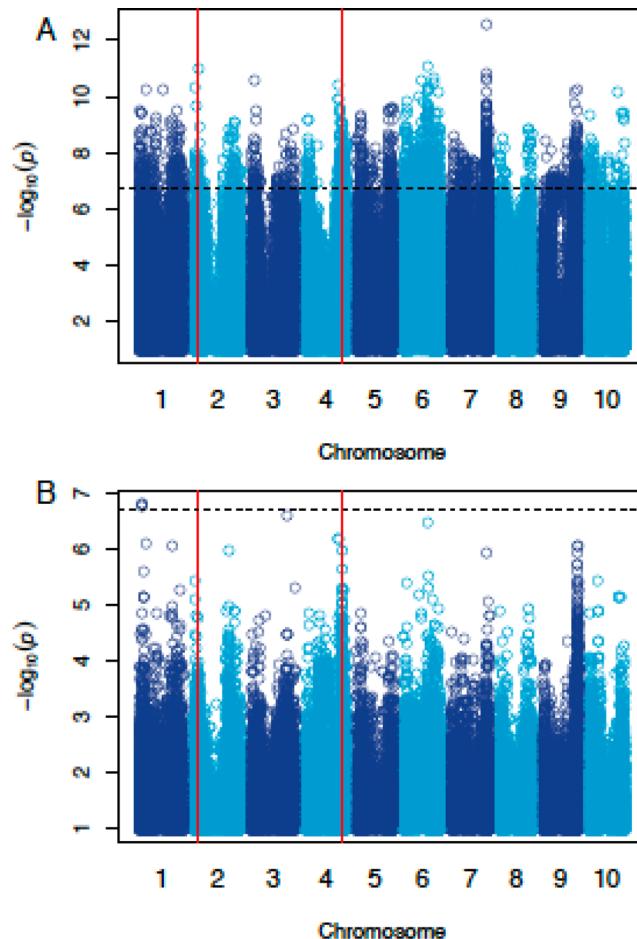
Mb and 55 Mb, close to *TT1* and *TT2* homologues, and an F3'H *Pr1* coorthologue. While there was not a distinct peak on chromosome 1, the strongest association signal in the GWAS was found in a diffuse peak on chromosome 1 around 55 Mb ( $P < 10^{-9}$ ). The closest a priori candidates were putative *TTG2* (Sb01g032120) and *TT2* (Sb01g032770) homologues. There were no distinct peaks or significant associations identified in the MLM (Figure 7B).

**Grain Color.** Since grain color is commonly used as a visual marker for sorghum polyphenol content, we used our data set to better understand both the correlation between visually scored grain color and polyphenol concentration, and the potential shared genetic basis for these traits. Based on visual assessment of grain appearance, we designated 142 white, 35 yellow, 48 red, and 152 brown grain accessions. An analysis of variance (ANOVA) showed significant variation among the grain color groups, so we conducted a post hoc Tukey test. Grain classified as red contained significantly more 3-deoxyanthocyanidins than brown ( $P < 10^{-5}$ ) or white grain ( $P < 10^{-5}$ ) accessions, but no significant difference was found between red and yellow accessions (Figure 8A and Table 4). Brown grain accessions contained significantly more proanthocyanidins than accessions with red ( $P = 0.0001$ ), white ( $P = 0.001$ ), or yellow ( $P = 0.001$ ) grain (Figure 8B and Table 4). This was expected as most of the sorghums with testa layers were classified as brown (57%). We also compared proanthocyanidin concentrations between grain color in proanthocyanidin-containing (greater than 10.00 mg CE/g or presence of pigmented testa) accessions. Brown grain color classes contained significantly more proanthocyanidins than nonbrown (brown  $n = 120$ , nonbrown  $n = 85$ ,  $P < 10^{-13}$ ). However, when brown grain color classes were compared to each color class individually, they only contained significantly more proanthocyanidins than white color classes ( $P < 10^{-4}$ ). Red and lemon-yellow grain color classes also contained significantly more proanthocyanidins than white in the proanthocyanidin-containing accessions ( $P = 0.001$  and  $P = 0.02$ ).

To identify genes associated with brown grain, we conducted a presence/absence (brown versus nonbrown) GWAS on all 373 of the accessions (Figures S5A and S5B and Table S8 in the Supporting Information) and another presence/absence (brown versus nonbrown) GWAS on the 203 proanthocyanidin-containing accessions (Figures S5C and S5D and Table S9 in the Supporting Information). A distinct association peak on chromosome 8 at 52.9 Mb was observed in both GWAS. The nearest a priori candidate was a putative *TT12* homologue within 400 Kb (Sb08g021640). The GWAS conducted on all 373 accessions identified a peak on chromosome 3 at 63.6 Mb,



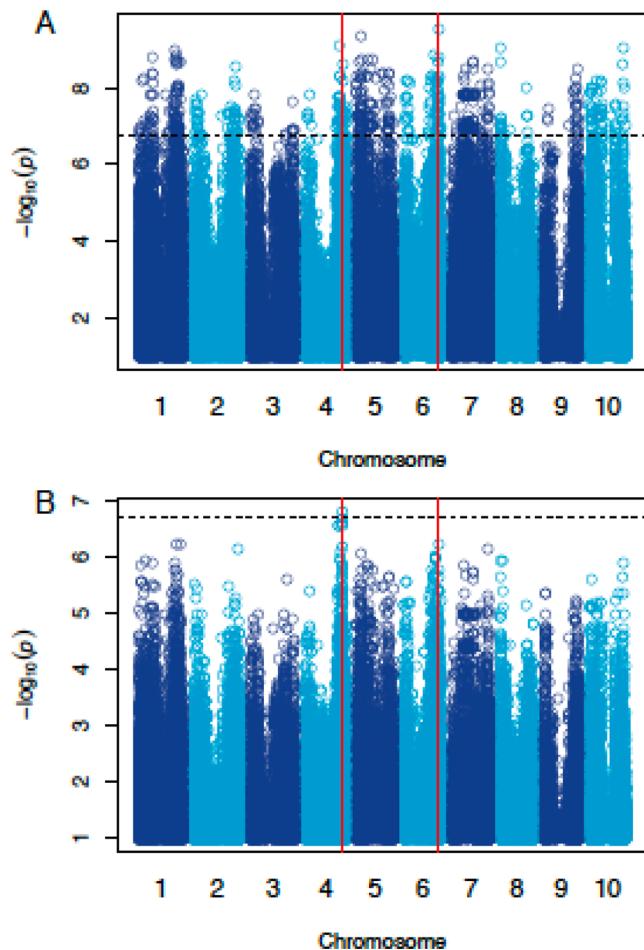
**Figure 4.** GWAS for proanthocyanidin concentration in sorghum grain. Manhattan plot of association results from (A) a GLM analysis, (B) an MLM analysis, and (C) a closeup of the peak on chromosome 4 showing *Tannin1* and other candidate genes in the region, using 404,628 SNP markers and 373 accessions. Axes: the  $-\log_{10} p$ -values (y axis) plotted against the position on each chromosome (x axis). Each circle represents a SNP. The dashed horizontal line represents the genome-wide significance threshold as determined by Bonferroni correction. Regions with  $-\log_{10} p$ -values above the threshold are candidates. The vertical lines indicate the location of *Tannin1* and a priori candidate genes in the *Tannin1* region (~61 Mb).



**Figure 5.** GWAS for proanthocyanidin concentration in sorghum grain with accessions containing *tan1-a* and *tan1-b* nonfunctional alleles removed. Manhattan plot of association results from (A) a GLM analysis and (B) an MLM analysis, using 404,628 SNP markers and 312 accessions. Axes: the  $-\log_{10} p$ -values (y axis) plotted against the position on each chromosome (x axis). Each circle represents a SNP. The dashed horizontal line represents the genome-wide significance threshold as determined by Bonferroni correction. Regions with  $-\log_{10} p$ -values above the threshold are candidates. The red vertical lines highlight the location of candidate genes. (TT8 on chrm. 2 and *TTG1*, *Zm1*, and *TT16* on chrm. 4)

within 100 Kb of another putative *TT12* homologue (Sb03g035610), and also a peak on chromosome 6 (S6\_56992521,  $P < 3 \times 10^{-10}$ ) near a *TT16* a priori candidate (Sb06g028420) (Figures S4A and S4B in the Supporting Information). The GWAS conducted on the proanthocyanidin-containing accessions identified a peak on chromosome 2 around 69.6 Mb, very near another *TT12* homologue (Sb02g034720) (Figures SSC and SSD in the Supporting Information). This peak was also identified in the GWAS conducted on all 373 accessions, but was more diffuse. There were no peaks on chromosome 4 around *Tannin1* or on chromosome 2 around the *Z* locus.

To identify genes associated with red grain, we conducted a presence/absence (red versus nonred) GWAS on all of the samples (Figure S6 and Table S10 in the Supporting Information). Two association peaks on chromosome 4 were identified by both the GLM and MLM, in the same region as the peak in the 3-deoxyanthocyanidin GWAS. The first peak, at 54.5 Mb, colocalized with a priori candidate Sb04g024710, the

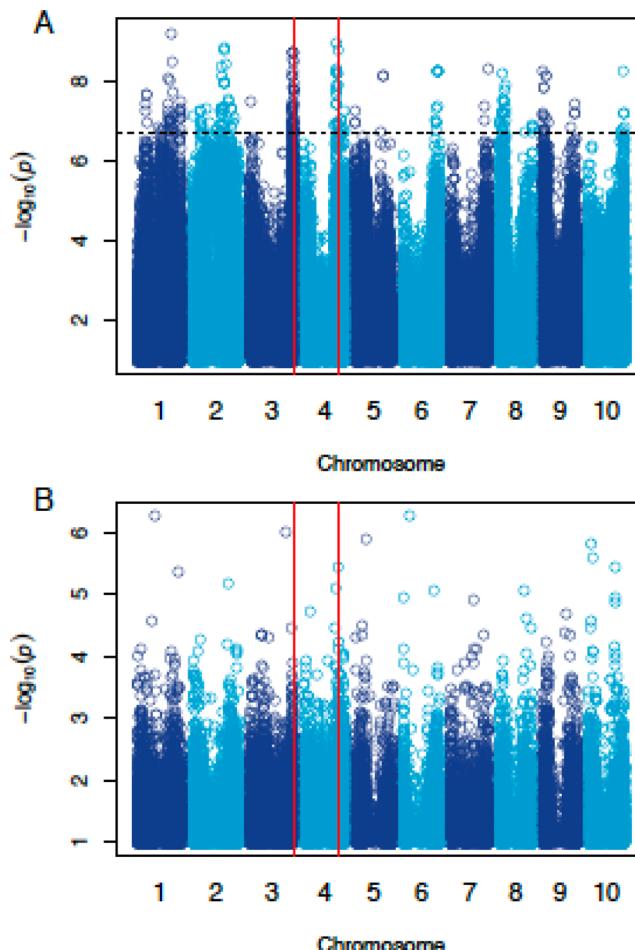


**Figure 6.** GWAS for proanthocyanidin concentration in proanthocyanidin-containing sorghum grain (greater than 10.00 mg CE/g or pigmented testa). Manhattan plot of association results from (A) a GLM analysis and (B) an MLM analysis, using 404,628 SNP markers and 208 accessions. Axes: the  $-\log_{10} p$ -values (y axis) plotted against the position on each chromosome (x axis). Each circle represents a SNP. The dashed horizontal line represents the genome-wide significance threshold as determined by Bonferroni correction. Regions with  $-\log_{10} p$ -values above the threshold are candidates. The red vertical lines highlight the location of candidate genes (*TT16*, *Tannin1* region, *Pr1*/*TT7*).

F3'H *Pr1* coorthologue that was also in one of the 3-deoxyanthocyanidin GWAS peaks. The second peak, at 55.9 Mb, was very close to a priori candidate Sb04g026480, a putative MYB homologue. There was also a peak around 72 Mb on chromosome 3, in the same region as the peak in the 3-deoxyanthocyanidin GWAS, near a priori candidate Sb03g044980, a putative *TT19* homologue. A peak was identified on chromosome 6 between 7 and 8 Mb, which was not near any a priori genes, but was near a putative vacuolar sorting protein gene (Sb06g003780). There were no peaks on chromosome 3 around the R locus.

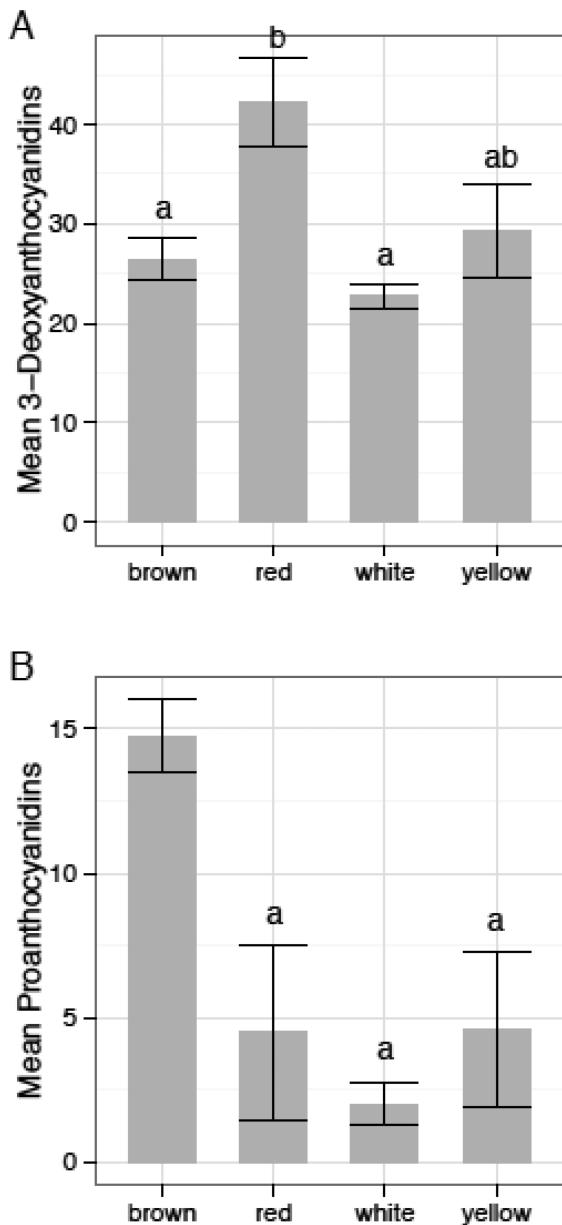
## DISCUSSION

**Genetic Controls of Sorghum Polyphenols.** The genetic controls of the flavonoid pathway (Figure 9) have been well studied in many economically important food plants, including grape (*Vitis vinifera*), barley (*Hordeum vulgare*), maize (*Zea mays*), rice (*Oryza sativa*), and wheat (*Triticum spp.*).<sup>43</sup> Much of our understanding of flavonoid genetics, including



**Figure 7.** GWAS for 3-deoxyanthocyanidin concentration in sorghum grain. Manhattan plot of association results from (A) a GLM analysis and (B) an MLM analysis, using 404,628 SNP markers and 373 accessions. Axes: the  $-\log_{10} p$ -values (y axis) plotted against the position on each chromosome (x axis). Each circle represents a SNP. The dashed horizontal line represents the genome-wide significance threshold as determined by Bonferroni correction. Regions with  $-\log_{10} p$ -values above the threshold are candidates. The red vertical lines highlight the location of candidate genes (*TT18*/*ANS*, *TT6*/*F3H*, *Pr1*/*TT7*).

biosynthetic enzymes, transporters, and regulatory proteins, come from analysis of Transparent Testa (TT) mutants in *Arabidopsis*.<sup>44</sup> Transcriptional regulation occurs through a ternary complex made up of *TT2*, *TT8*, and *TTG1*, which encode for MYB, bHLH, and WD40 proteins (MBW complex), respectively.<sup>44</sup> This ternary complex is highly conserved among plant species.<sup>45</sup> In the sorghum proanthocyanidin pathway, the WD40 (*Tannin1*) component of the MBW complex has been identified, as well as a likely candidate for the bHLH; several studies have found a significant linkage and association on sorghum chromosome 2 around 8 Mb, near a putative bHLH transcription factor orthologous to *Arabidopsis TT8*.<sup>22,24,41,46,47</sup> The MYB transcription factor that would complete the ternary complex has not been found in sorghum. The *Zm1* homologue on chromosome 4 at 61.1 Mb (Sb04g031110, 66.8% similarity), which was mapped in all of our proanthocyanidin GWAS, is a possible candidate for the missing MYB. The maize *Zm1* gene is a MYB transcription factor, homologous to classical maize grain pigmentation gene *C1* that can induce transcription of DFR, an essential structural enzyme in the flavonoid pathway.<sup>48</sup>



**Figure 8.** Polyphenol differences between grain colors. Mean concentrations of (A) proanthocyanidins and (B) 3-deoxyanthocyanidins in accessions of each grain color. Color categories share the same letter if they are not significantly different from each other, based on a post hoc Tukey HSD test (brown,  $n = 152$ ; red,  $n = 48$ ; white,  $n = 142$ ; yellow,  $n = 35$ ).

Another possible explanation for the significant SNPs at this location is an indirect association with an undescribed allele at *Tannin-1*.

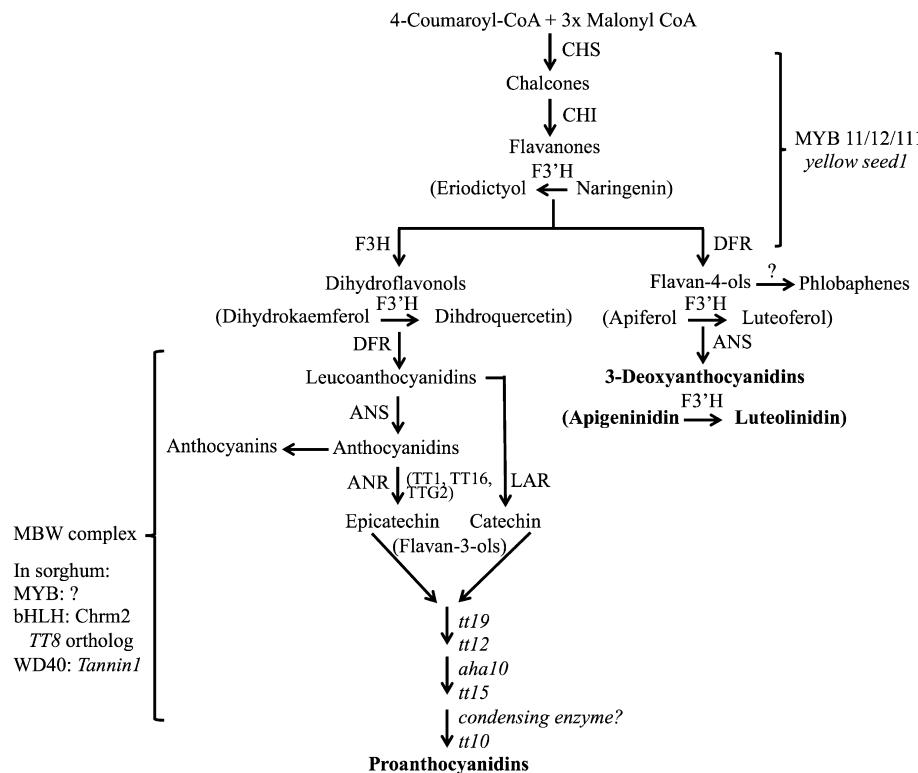
**Table 4. Polyphenol Concentrations by Color**

color	n	total phenols (mg GAE/g)		PA (mg CE/g)		3-DA (abs/mL/g)	
		mean	range	mean	range	mean	range
white	142	4.0 ± 3.10	ND–14.67	2.00 ± 8.84	ND–25.26	22.74 ± 14.03	ND–58.41
yellow	35	6.03 ± 6.18	ND–23.69	4.60 ± 15.98	ND–42.30	29.30 ± 27.89	ND–98.90
red	48	6.97 ± 7.30	ND–27.32	4.48 ± 21.10	ND–52.83	42.21 ± 30.43	ND–135.34
brown	152	10.01 ± 6.01	ND–37.46	14.74 ± 15.63	ND–78.51	26.46 ± 26.64	ND–149.21

About two-thirds of the SAP accessions we studied were “converted” tropical accessions, meaning that alleles for reduced height and early flowering have been introgressed so they can be grown in temperate regions.<sup>49</sup> Surprisingly, the proanthocyanidin GWAS association peak on chromosome 7 (~58.6 Kb) precisely colocalizes with *dw3* (Sb07g023730), a dwarfing loci used in the conversion, in conjunction with *dw1*, *dw2*, and *dw4*.<sup>27</sup> Smaller peaks on chromosomes 6 (~39 Kb) and 9 (~57 Kb) were near the *dw2* and *dw1* loci. The association peaks on chromosomes 6, 7, and 9 may be artifacts arising from a lower mean proanthocyanidin concentration in the converted lines (4.4 mg CE/g) which all shared the same *dw* alleles, compared to the unconverted lines (11.0 mg CE/g). Accordingly, when we conducted a proanthocyanidin GWAS using only converted accessions to control for this spurious phenotypic covariation between proanthocyanidin and height, the peaks near *dw1*, *dw2*, and *dw3* disappeared, while the *Tannin1* peak remained (Figure S7 in the Supporting Information).

As a phytoalexin,<sup>15,16</sup> the effect of the environment may make it more difficult to map the genetic basis of 3-deoxyanthocyanidins than the genetic basis of proanthocyanidins. Although the GLM was able to identify significant SNP associations for 3-deoxyanthocyanidins, there were few peaks, and the MLM did not identify any significant associations. Detection of alleles contributing to variance of 3-deoxyanthocyanidins may require a larger sample size, additional replication, a biparental mapping population, or controlled fungal inoculations to induce biosynthesis of polyphenol compounds.<sup>23</sup> However, our results did provide a promising candidate for followup. A *Pr1* orthologue (Sb04g024750) lies within a distinct peak on chromosome 4, about 400 Kb from the top SNP identified in the 3-deoxyanthocyanidin GWAS (S4\_54975391;  $P < 10^{-8}$ ), and 100 Kb from the top SNP in the red grain GWAS (S4\_54555458,  $P < 10^{-13}$ ). *Pr1* is a maize F3'H enzyme, homologous to *TT7* in *Arabidopsis*. The F3'H enzyme is essential for production of 3-deoxyanthocyanidins, as well as the red phlobaphene pigments visible in maize,<sup>18</sup> and has been implicated in production of these compounds in sorghum (Figure 9).<sup>50</sup> Overall, we observe a 1.6-fold difference in 3-deoxyanthocyanidin concentrations between accessions carrying the high concentration alleles and low concentration alleles for the top red grain-associated SNP ( $P = 0.001$ ). F3'H is necessary for proanthocyanidin production as well, and, indeed, significant associations with SNPs in the ~54 Mb region on chromosome 4 were also identified in the GWAS with *tan1-a* and *tan1-b* samples removed, as well as the GWAS with only proanthocyanidin-containing samples.

Our study identified many peaks and SNPs significantly associated with proanthocyanidins and 3-deoxyanthocyanidins, hence there appear to be many small effect genes controlling natural variation of these traits. Consequently, a larger association panel, or a targeted biparental mapping population,



**Figure 9.** Simplified scheme of flavonoid biosynthetic pathway in sorghum grain with candidate genes noted. Enzyme abbreviations are in uppercase letters, while gene abbreviations are in italicics. Question marks depict unknown steps. Chalcone synthase (CHS), chalcone-flavanone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavanone 3'-hydroxylase (F3'H), dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), anthocyanidin reductase (ANR), leucoanthocyanidin reductase (LAR); MYB-bHLH-WD40 (MBW).

may be more effective in precisely identifying causal alleles. Moving forward, sequence analysis and expression analysis of the candidate genes are needed to identify causal polymorphisms and lay the groundwork for the use of polyphenol genetic variation in crop improvement.

**Crop Improvement for Sorghum Polyphenols.** Efforts to characterize polyphenols, with the goal of producing high polyphenol specialty varieties, have been undertaken in several grain crops, including purple wheat,<sup>51</sup> black rice,<sup>52</sup> multicolored maize,<sup>53</sup> multicolored barley,<sup>54</sup> and black sorghum.<sup>55</sup> Our diverse association panel contained a wide range of proanthocyanidin and 3-deoxyanthocyanidin concentrations, and this genetic variation may be useful in breeding programs to produce high polyphenol specialty varieties. Bicolor sorghums had the highest mean proanthocyanidin concentrations, but their grain weight is significantly less (20% less) than that of non-bicolor sorghums ( $P < 10^{-9}$ ). Combined with low yield potential, the small grain size makes it difficult to use bicolor race sorghums in a grain sorghum breeding program, but may still be of interest to breeders wanting to produce specialty varieties. In addition to bicolor sorghums, caudatum and guinea-caudatum sorghums also had high mean proanthocyanidin concentrations, and are promising sources for increasing proanthocyanidin concentrations in sorghum. In particular, among the caudatum and guinea-caudatum sorghums, caudatum sorghums from tropical climates such as Uganda had the highest mean proanthocyanidin concentrations, so they may be good material for breeding high polyphenol sorghums. While bicolor-durra and guinea sorghums had the highest mean 3-deoxyanthocyanidin concentrations, the difference among all the races was not significant,

so it may be more important to simply identify unique genotypes across the sorghum collection. Chemical analysis is underway on the samples that were outside of the NIRS calibration curves, and true biological outliers may open up new avenues for future work on sorghum varieties with extreme polyphenol concentrations.

Increasing 3-deoxyanthocyanidin production may be challenging, since, as phytoalexins, they are not constitutively expressed, but rather synthesized by plants under pathogen attack.<sup>15,16</sup> We note in our comparison of 3-deoxyanthocyanidin concentrations from duplicate samples that the difference between duplicates becomes larger for accessions with higher 3-deoxyanthocyanidin concentrations. One possibility is that there is greater technical variation in the 3-deoxyanthocyanidin NIRS estimates, but Dykes et al.<sup>35</sup> demonstrated the same correlation coefficient between the NIRS-predicted values and the values in the validation set for proanthocyanidins ( $r = 0.81$ ) and 3-deoxyanthocyanidins ( $r = 0.82$ ). Therefore, we would not expect to see differences in accuracy of the NIRS predictions for proanthocyanidins and 3-deoxyanthocyanidins in our study. As this was a field study, another possibility is that uncontrolled environmental variation may have contributed to the difference between the duplicate samples. Accessions with the genetic capability to produce grain 3-deoxyanthocyanidins may be producing low or high 3-deoxyanthocyanidin concentrations depending on the exposure to inducing agents on a given panicle. Controlled inoculation studies are needed to further explore this possibility.<sup>23</sup>

The spreader gene is a promising target for increasing grain proanthocyanidin concentrations, and a previous report using a small number of varieties has shown higher proanthocyanidin

concentrations in varieties with a functional spreader.<sup>56</sup> Given that three peak SNP associations in the brown grain GWAS were near putative MATE transporter *TT12* homologues, we propose that the spreader gene may be a *MATE* transporter. A biparental mapping population segregating the spreader gene would be needed to confirm this hypothesis. To get a sense of the effect these loci may have on proanthocyanidin concentrations, we compared concentrations of each allele in proanthocyanidin-containing accessions. There was a 1.8-fold (*S3\_63633634*,  $P = 0.04$ ), a 1.5-fold (*S2\_69656067*,  $P = 0.0003$ ), and a 1.7-fold (*S8\_52906014*,  $P = 0.0002$ ) difference between accessions carrying the high concentration alleles and low concentration alleles. When the three polymorphisms are considered together, accessions with all three high-alleles (*S2\_69656067* = "A", *S3\_63633634* = "A", *S8\_52906014* = "G") have 1.7- to 2.7-fold higher proanthocyanidin concentrations ( $P = 10^{-8}$ ), consistent with an additive effect more than doubling the concentration of proanthocyanidins in sorghum grain.

Appearance of grain color is predominantly due to polyphenols, but can also be influenced by endosperm color and grain weathering. Taken in total, the color classes used for our analysis represent general groups and are not definitive descriptors of any specific trait. For example, it is possible to have a sorghum classified as brown that does not have a testa layer, as well to have a sorghum classified as white that has a testa layer (see Figure 1). However, our results support the use of visual categorization of grain color as a simple assessment of polyphenol concentrations in crop improvement programs; brown grain has significantly higher proanthocyanidin concentrations than nonbrown, red grain has significantly higher 3-deoxyanthocyanidin concentrations than nonred, and white grain has significantly lower concentrations of these polyphenols than nonwhite. Additionally, the genetic architecture of grain color reflects, to an extent, that of the polyphenols with which they are associated. For instance, the red grain GWAS and the 3-deoxyanthocyanidin GWAS produced similar association peaks on chromosome 4 (~54 Mb), which may map to the sorghum *Pr1* orthologue, and chromosome 3 (~72 Mb), which colocalizes with putative homologues of *ANS*, *F3H*, and *TT19*. The brown grain GWAS and the proanthocyanidin-containing GWAS produced similar association peaks on chromosome 6 (~57 Mb) near a priori candidate *TT16*, a key regulatory protein in the proanthocyanidin branch of the flavonoid pathway. Overall, to increase sorghum proanthocyanidin and 3-deoxyanthocyanidin concentrations quantitatively, there are many associated alleles available, but none of them have large effect. This survey of grain polyphenol variation in sorghum germplasm and catalog of flavonoid pathway-associated loci contributes toward the goal of producing sorghum crops that will contribute to marker-assisted breeding of sorghum crops that will benefit human health.

## ASSOCIATED CONTENT

### Supporting Information

Proanthocyanidin concentration in accessions with *Tannin1* and *tan1-a* (Figure S1), population structure of polyphenols (Figure S2), GWAS for proanthocyanidin presence/absence (Figures S3, S4), GWAS for grain color (Figures S5, S6), GWAS for converted lines (Figure S7), flavonoid pathway-related genes (Table S1), significant SNPs identified in each GWAS (Tables S2–S10), and flavonoid pathway a priori

candidate gene list. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Tel: 773-603-2897. Fax: 785-532-6094. E-mail: rhodesdh@email.sc.edu.

### Present Address

<sup>II</sup>Institute for Genomic Diversity, Cornell University, Ithaca, NY, 14853, USA.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Zach Brenton for considerable help with sample preparation, and Scott Bean, Prini Gadgil, Tom Herald, Amy Murphy, and the reviewers for their thoughtful comments.

## ABBREVIATIONS USED

abs, absorbance; ANOVA, analysis of variance; CE, catechin equivalents; GAE, gallic acid equivalent; GBS, genotyping-by-sequencing; GAPIT, Genome Association and Prediction Integrated Tool; GLM, general linear model; GRIN, Germplasm Resources Information Network; GWAS, genome-wide association study; K, kinship; MLM, mixed linear model; NIRS, near-infrared spectroscopy; SAP, Sorghum Association Panel; SNP, single-nucleotide polymorphism

## REFERENCES

- (1) Tsao, R. Chemistry and biochemistry of dietary polyphenols. *Nutrients* **2010**, *2*, 1231–1246.
- (2) Hichri, I.; Barrieu, F.; Bogs, J.; Kappel, C.; Delrot, S.; Lauvergeat, V. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* **2011**, *62*, 2465–2483.
- (3) Buer, C. S.; Imin, N.; Djordjevic, M. A. Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* **2010**, *52*, 98–111.
- (4) Del Rio, D.; Rodriguez-Mateos, A.; Spencer, J. P. E.; Tognolini, M.; Borges, G.; Crozier, A. Dietary (poly)phenolics in human health: structures, bioavailability, and evidence of protective effects against chronic diseases. *Antioxid. Redox Signaling* **2013**, *18*, 1818–1892.
- (5) Hellstrom, J. K.; Törrönen, A. R.; Mattila, P. H. Proanthocyanidins in common food products of plant origin. *J. Agric. Food Chem.* **2009**, *57*, 7899–7906.
- (6) Santos-Buelga, C.; Scalbert, A. Proanthocyanidins and tannin-like compounds—nature, occurrence, dietary intake and effects on nutrition and health. *J. Sci. Food Agric.* **2000**, *80*, 1094–1117.
- (7) Dixon, R. A.; Xie, D.-Y.; Sharma, S. B. Proanthocyanidins—a final frontier in flavonoid research? *New Phytol.* **2005**, *165*, 9–28.
- (8) FAO. Sorghum and millets in human nutrition. <http://www.fao.org/docrep/T0818E/T0818E04.htm> (accessed Feb 18, 2014).
- (9) Janzen, E. L.; Wilson, W. W. *Cooperative marketing in specialty grains and identity preserved grain markets*; Agribusiness & Applied Economics Report No. 500; North Dakota State University, Department of Agribusiness and Applied Economics: Fargo, ND, September 2002.
- (10) Taylor, J. R. N.; Schober, T. J.; Bean, S. R. Novel food and non-food uses for sorghum and millets. *J. Cereal Sci.* **2006**, *44*, 252–271.
- (11) Elbehri, A. *The changing face of the U.S. grain system: differentiation and identity preservation trends*; Economic Research Report 7185; United States Department of Agriculture, Economic Research Service: 2007.
- (12) Cureton, P.; Fasano, A. The increasing incidence of celiac disease and the range of gluten-free products in the marketplace. In *Gluten-Free Food Science and Technology*; Gallagher, E., Ed.; Wiley-Blackwell: Oxford, U.K., 2009; pp 1–15.

- (13) Harlan, J. R.; de Wet, J. M. J. A simplified classification of cultivated sorghum. *Crop Sci.* **1972**, *12*, 172–176.
- (14) Awika, J. M.; Rooney, L. W. Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* **2004**, *65*, 1199–1221.
- (15) Nicholson, R. L.; Kollipara, S. S.; Vincent, J. R.; Lyons, P. C.; Cadena-Gomez, G. Phytoalexin synthesis by the sorghum mesocotyl in response to infection by pathogenic and nonpathogenic fungi. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5520–5524.
- (16) Dixon, R. A. Natural products and plant disease resistance. *Nature* **2001**, *411*, 843–847.
- (17) Winefield, C. S.; Lewis, D. H.; Swinny, E. E.; Zhang, H.; Arathoon, H. S.; Fischer, T. C.; Halbwirth, H.; Stich, K.; Gosch, C.; Forkmann, G.; Davies, K. M. Investigation of the biosynthesis of 3-deoxyanthocyanins in *Sinningia cardinalis*. *Physiol. Plant.* **2005**, *124*, 419–430.
- (18) Sharma, M.; Chai, C.; Morohashi, K.; Grotewold, E.; Snook, M. E.; Chopra, S. Expression of flavonoid 3'-hydroxylase is controlled by *p1*, the regulator of 3-deoxylflavonoid biosynthesis in maize. *BMC Plant Biol.* **2012**, *12*, 196.
- (19) Malathi, P.; Viswanathan, R.; Padmanaban, P.; Mohanraj, D.; Kumar, V. G.; Salin, K. P. Differential accumulation of 3-deoxyanthocyanidin phytoalexins in sugarcane varieties varying in red rot resistance in response to *colletotrichum falcatum* infection. *Sugar Tech* **2008**, *10*, 154–157.
- (20) Rooney, W. L. Genetics and cytogenetics. In *Sorghum: Origin, History, Technology, and Production*, 1st ed.; Smith, C. W., Frederiksen, R. A., Eds.; John Wiley & Sons: New York, NY, 2000; pp 261–307.
- (21) Morohashi, K.; Casas, M. I.; Ferreyra, L. F.; Mejia-Guerra, M. K.; Pourcel, L.; Yilmaz, A.; Feller, A.; Carvalho, B.; Emiliani, J.; Rodriguez, E.; Pellegrinet, S.; McMullen, M.; Casati, P.; Grotewold, E. A genome-wide regulatory framework identifies maize *pericarp color1* controlled genes. *Plant Cell* **2012**, *24* (7), 2745–2764.
- (22) Wu, Y.; Li, X.; Xiang, W.; Zhu, C.; Lin, Z.; Wu, Y.; Li, J.; Pandravada, S.; Ridder, D. D.; Bai, G.; Wang, M. L.; Trick, H. N.; Bean, S. R.; Tuinstra, M. R.; Tesso, T. T.; Yu, J. Presence of tannins in sorghum grains is conditioned by different natural alleles of *Tannin1*. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (26), 10281–10286.
- (23) Ibraheem, F.; Gaffoor, I.; Chopra, S. Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional *yellow seed1* in *Sorghum bicolor*. *Genetics* **2010**, *184*, 915–926.
- (24) Mace, E. S.; Jordan, D. R. Location of major effect genes in sorghum (*Sorghum bicolor* (L.) Moench). *Theor. Appl. Genet.* **2010**, *121*, 1339–1356.
- (25) Routaboul, J.-M.; Dubos, C.; Beck, G.; Marquis, C.; Bidzinski, P.; Loudet, O.; Lepiniec, L. Metabolite profiling and quantitative genetics of natural variation for flavonoids in arabidopsis. *J. Exp. Bot.* **2012**, *63* (10), 3749–3764.
- (26) Olsen, K. M.; Wendel, J. F. Crop plants as models for understanding plant adaptation and diversification. *Front. Plant Sci.* **2013**, *4*.
- (27) Morris, G. P.; Ramu, P.; Deshpande, S. P.; Hash, C. T.; Shah, T.; Upadhyaya, H. D.; Riera-Lizarazu, O.; Brown, P. J.; Acharya, C. B.; Mitchell, S. E.; Harriman, J.; Glaubitz, J. C.; Buckler, E. S.; Kresovich, S. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 453–458.
- (28) Flint-Garcia, S. A. Genetics and consequences of crop domestication. *J. Agric. Food Chem.* **2013**, *61*, 8267–8276.
- (29) Myles, S.; Peiffer, J.; Brown, P. J.; Ersoz, E. S.; Zhang, Z.; Costich, D. E.; Buckler, E. S. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* **2009**, *21*, 2194–2202.
- (30) Hirschhorn, J. N.; Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **2005**, *6*, 95–108.
- (31) Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; Li, M.; Fan, D.; Guo, Y.; Wang, A.; Wang, L.; Deng, L.; Li, W.; Lu, Y.; Weng, Q.; Liu, K.; Huang, T.; Zhou, T.; Jing, Y.; Li, W.; Lin, Z.; Buckler, E. S.; Qian, Q.; Zhang, Q.-F.; Li, J.; Han, B. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967.
- (32) Shu, X.; Backes, G.; Rasmussen, S. K. Genome-wide association study of resistant starch (RS) phenotypes in a barley variety collection. *J. Agric. Food Chem.* **2012**, *60*, 10302–10311.
- (33) Casa, A. M.; Pressoir, G.; Brown, P. J.; Mitchell, S. E.; Rooney, W. L.; Tuinstra, M. R.; Franks, C. D.; Kresovich, S. Community resources and strategies for association mapping in sorghum. *Crop Sci.* **2008**, *48*, 30.
- (34) USDA. GRIN National Genetic Resources Program. <http://www.ars-Grin.gov> (2014-07-25).
- (35) Dykes, L.; Hoffmann, L., Jr.; Portillo-Rodriguez, O.; Rooney, W. L.; Rooney, L. W. Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. *J. Cereal Sci.* **2014**, *60* (1), 138–142.
- (36) Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S.; Mitchell, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **2011**, *6*, e19379.
- (37) Paterson, A. H.; Bowers, J. E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; Schmutz, J.; Spannagl, M.; Tang, H.; Wang, X.; Wicker, T.; Bharti, A. K.; Chapman, J.; Feltus, F. A.; Gowik, U.; Grigoriev, I. V.; Lyons, E.; Maher, C. A.; Martis, M.; Narechania, A.; Otillar, R. P.; Penning, B. W.; Salamov, A. A.; Wang, Y.; Zhang, L.; Carpeita, N. C.; Freeing, M.; Gingle, A. R.; Hash, C. T.; Keller, B.; Klein, P.; Kresovich, S.; McCann, M. C.; Ming, R.; Peterson, D. G.; Mehboobur-Rahman; Ware, D.; Westhoff, P.; Mayer, K. F. X.; Messing, J.; Rokhsar, D. S. The sorghum bicolor genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556.
- (38) Buckler Lab for Maize Genetics and Diversity. TASSEL. <http://sourceforge.net/projects/tassel> (7/25/14).
- (39) Lipka, A. E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P. J.; Gore, M. A.; Buckler, E. S.; Zhang, Z. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399.
- (40) Yu, J.; Pressoir, G.; Briggs, W. H.; Bi, I. V.; Yamasaki, M.; Doebley, J. F.; McMullen, M. D.; Gaut, B. S.; Nielsen, D. M.; Holland, J. B.; Kresovich, S.; Buckler, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208.
- (41) Morris, G. P.; Rhodes, D. H.; Brenton, Z.; Ramu, P.; Thayil, V. M.; Deshpande, S.; Hash, C. T.; Acharya, C.; Mitchell, S. E.; Buckler, E. S.; Yu, J.; Kresovich, S. Dissecting genome-wide association signals for loss-of-function phenotypes in sorghum flavonoid pigmentation traits. *G3: Genes, Genomes, Genet.* **2013**, *3*, 2085–2094.
- (42) Zhang, Z.; Ersoz, E.; Lai, C.-Q.; Todhunter, R. J.; Tiwari, H. K.; Gore, M. A.; Bradbury, P. J.; Yu, J.; Arnett, D. K.; Ordovas, J. M.; Buckler, E. S. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360.
- (43) Liu, Z.; Liu, Y.; Pu, Z.; Wang, J.; Zheng, Y.; Li, Y.; Wei, Y. Regulation, evolution, and functionality of flavonoids in cereal crops. *Biotechnol. Lett.* **2013**, *35*, 1765–1780.
- (44) Lepiniec, L.; Debeaujon, I.; Routaboul, J.-M.; Baudry, A.; Pourcel, L.; Nesi, N.; Caboche, M. Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **2006**, *57*, 405–430.
- (45) Petroni, K.; Tonelli, C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science* **2011**, *181*, 219–229.
- (46) Nesi, N.; Debeaujon, I.; Jond, C.; Pelletier, G.; Caboche, M.; Lepiniec, L. The *tt8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in arabidopsis siliques. *Plant Cell* **2000**, *12*, 1863–1878.
- (47) Furukawa, T.; Maekawa, M.; Oki, T.; Suda, I.; Iida, S.; Shimada, H.; Takamure, I.; Kadokawa, K. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **2007**, *49*, 91–102.
- (48) Franken, P.; Schrell, S.; Peterson, P. A.; Saedler, H.; Wienand, U. Molecular analysis of protein domain function encoded by the *myb*-

homologous maize genes *C1*, *Zm 1* and *Zm 38*. *Plant Journal* **1994**, *6*, 21–30.

(49) Smith, C. W.; Finlayson, S. A. physiology and genetics of maturity and height. In *Sorghum: Origin, History, Technology, and Production*, 1st ed.; Smith, C. W., Frederiksen, R. A., Eds.; John Wiley & Sons: New York, NY, 2000; pp 261–307.

(50) Boddu, J.; Svabek, C.; Sekhon, R.; Gevens, A.; Nicholson, R. L.; Jones, A. D.; Pedersen, J. F.; Gustine, D. L.; Chopra, S. Expression of a putative flavonoid 3'-hydroxylase in sorghum mesocotyls synthesizing 3-deoxyanthocyanidin phytoalexins. *Physiol. Mol. Plant Pathol.* **2004**, *65*, 101–113.

(51) Chen, W.; Müller, D.; Richling, E.; Wink, M. Anthocyanin-rich purple wheat prolongs the life span of *Caenorhabditis elegans* probably by activating the DAF-16/FOXO transcription factor. *J. Agric. Food Chem.* **2013**, *61*, 3047–3053.

(52) Sriseadka, T.; Wongpornchai, S.; Rayanakorn, M. Quantification of flavonoids in black rice by liquid chromatography-negative electrospray ionization tandem mass spectrometry. *J. Agric. Food Chem.* **2012**, *60*, 11723–11732.

(53) Zilic, S.; Serpen, A.; Akıllıoğlu, G.; Gökmən, V.; Vančetović, J. Phenolic compounds, carotenoids, anthocyanins, and antioxidant capacity of colored maize (*Zea mays* L.) Kernels. *J. Agric. Food Chem.* **2012**, *60*, 1224–1231.

(54) Kim, M.-J.; Hyun, J.-N.; Kim, J.-A.; Park, J.-C.; Kim, M.-Y.; Kim, J.-G.; Lee, S.-J.; Chun, S.-C.; Chung, I.-M. Relationship between phenolic compounds, anthocyanins content and antioxidant activity in colored barley germplasm. *J. Agric. Food Chem.* **2007**, *55*, 4802–4809.

(55) Dykes, L.; Rooney, W. L.; Rooney, L. W. Evaluation of phenolics and antioxidant activity of black sorghum hybrids. *J. Cereal Sci.* **2013**, *58*, 278–283.

(56) Dykes, L.; Rooney, L. W.; Waniska, R. D.; Rooney, W. L. Phenolic compounds and antioxidant activity of sorghum grains of varying genotypes. *J. Agric. Food Chem.* **2005**, *53*, 6813–6818.

(57) Glaubitz, J. C.; Casstevens, T. M.; Lu, F.; Harriman, J.; Elshire, R. J.; Sun, Q.; Buckler, E. S. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **2014**, *9*, e90346.