

RESEARCH ARTICLE

Open Access



Genetic architecture of kernel composition in global sorghum germplasm

Davina H. Rhodes^{1*}, Leo Hoffmann Jr.², William L. Rooney², Thomas J. Herald¹, Scott Bean¹, Richard Boyles³, Zachary W. Brenton³ and Stephen Kresovich³

Abstract

Background: Sorghum [*Sorghum bicolor* (L.) Moench] is an important cereal crop for dryland areas in the United States and for small-holder farmers in Africa. Natural variation of sorghum grain composition (protein, fat, and starch) between accessions can be used for crop improvement, but the genetic controls are still unresolved. The goals of this study were to quantify natural variation of sorghum grain composition and to identify single-nucleotide polymorphisms (SNPs) associated with variation in grain composition concentrations.

Results: In this study, we quantified protein, fat, and starch in a global sorghum diversity panel using near-infrared spectroscopy (NIRS). Protein content ranged from 8.1 to 18.8%, fat content ranged from 1.0 to 4.3%, and starch content ranged from 61.7 to 71.1%. Durra and bicolor-durra sorghum from Ethiopia and India had the highest protein and fat and the lowest starch content, while kafir sorghum from USA, India, and South Africa had the lowest protein and the highest starch content. Genome-wide association studies (GWAS) identified quantitative trait loci (QTL) for sorghum protein, fat, and starch. Previously published RNAseq data was used to identify candidate genes within a GWAS QTL region. A putative alpha-amylase 3 gene, which has previously been shown to be associated with grain composition traits, was identified as a strong candidate for protein and fat variation.

Conclusions: We identified promising sources of genetic material for manipulation of grain composition traits, and several loci and candidate genes that may control sorghum grain composition. This survey of grain composition in sorghum germplasm and identification of protein, fat, and starch QTL contributes to our understanding of the genetic basis of natural variation in sorghum grain nutritional traits.

Keywords: Sorghum, GWAS, QTL, Natural variation, Grain composition, Cereal crop, Nutrition

Background

Chronic hunger can be alleviated by improving the nutrition of staple cereal crops, which provide the majority of nutrients to the world's population [1]. Grain composition varies within and among cereal crops, but, generally, grain contains 79–83% starch, 7–14% protein and 1–7% fat. Crop yields in the arid and semi-arid regions of the world are challenged by low precipitation, leaving populations in these regions particularly vulnerable to chronic hunger and malnutrition. Sorghum is a cereal crop that is well adapted to regions of low precipitation, and thus, has become a staple crop that feeds millions of people in sub-Saharan Africa [2], where the highest

prevalence of undernourishment in the world is found [3]. Understanding the natural variation of protein, fat, and starch, and identifying QTL associated with their natural variation in sorghum grain can help improve its nutritional quality through crop improvement programs and marker-assisted selection.

Until the seed is self-sustaining, protein, fat, and starch stores are used to support the developing seedling. Since these nutrient stores are also critical components of the human diet, many researchers have focused on improving the nutrient composition of seeds [4]. For instance, the Illinois long-term selection experiment, which began in 1896, has increased the oil and protein content of maize inbred lines to 20 and 27%, respectively, compared to ~6 and ~12%, in an average maize line [5–8]. The chemical composition of grain is controlled by complex regulations that takes place during the seed filling

* Correspondence: davina.rhodes@ars.usda.gov

¹Center for Grain and Animal Health Research, USDA-ARS, 1515 College Ave, Manhattan, Kansas 66502, USA

Full list of author information is available at the end of the article



stage of seed maturation, when protein, fat, and starch storage compounds accumulate [9–11]. Key insights into the genetic controls of grain composition have been discovered through several rice and maize mutations with altered grain composition, including *opaque-2* and *floury-2*, which affect protein content [12–15]; *linoleic1* and *fad2*, which affect fat content [16–18]; and *shrunk1* and *amylose extender1*, which affect starch content [19–21]. Sorghum mutations have also contributed to our knowledge of the genetic controls of grain composition. These mutations include *waxy*, which has little to no amylose, increased protein, and improved starch digestibility [22–24]; *sugary*, which has increased sucrose content [25, 26]; and *high-lysine*, which has increased lysine content and protein digestibility [27].

GWAS have identified allelic polymorphisms for important agronomic traits in cereal crops [28–32], including alleles responsible for variation in grain composition of rice [30, 32], maize [33–36], and barley [37, 38]. Linkage and association studies have identified several loci controlling sorghum grain composition [39–43], and the identification of the gene underlying the *waxy* mutation has been fine mapped to 1.8 Mb on chromosome 10 [44], but more work needs to be done to precisely identify genes responsible for natural variation of grain composition. GWAS for sorghum grain composition have identified QTL for polyphenol [45] and mineral traits [46], but no GWAS have been conducted for protein, fat, and starch composition.

Surveying the natural variation of grain composition in the sorghum germplasm and finding loci underlying the variation can aid efforts to improve the nutritional value of sorghum. New sources of genetic variation can be used for crop improvement, especially in developing countries where technologies that exist for improving the nutritional value of grain, such as commercial fortification, are not accessible or affordable [47–49]. The goals of this study were to quantify natural variation of sorghum grain protein, fat, and starch and to identify associated SNPs. Here, we characterize the natural variation of sorghum grain composition in a global sorghum diversity panel and use GWAS to identify allelic variation underlying variation in grain composition.

Methods

Plant materials

We grew 390 sorghum accessions from the Sorghum Association Panel (SAP) [50]. The panel includes important breeding lines from the United States and traditional varieties from all five major races (bicolor, guinea, caudatum, kafir, and durra) and 10 intermediate races (all combinations of the major races) [51]. Seeds were originally obtained from the U.S. National Plant Germplasm System's Germplasm Resources Information Network (GRIN) [52]

and planted in late April to May 2012, 2013, and 2014 at Clemson University Pee Dee Research and Education Center in Florence, SC. The field design has been described previously [28]. Briefly, a two-fold replicated complete randomized block design was used. Panicles from each plot were collected at physiological maturity (black layer), which occurs once grain filling is complete. Due to differences in maturity among these accessions, harvest occurred between September and October. Once harvested, panicles were air dried in a greenhouse and mechanically threshed. In the following analyses we consider 265 accessions for which we obtained replicated data in all 3 years.

Phenotyping

Protein, fat, and starch content were measured using NIRS at Texas A&M University's sorghum breeding and genetics lab. Twenty grams of cleaned whole grain were scanned with a FOSS XDS spectrometer (FOSS North America, Eden Prairie, MN, USA). The NIR reflectance spectra were recorded using the ISIscan software (Version 3.10.05933) and converted to estimates using in-house developed models for protein, fat, and starch concentrations (expressed as a percentage of dry weight). The total grain weight in grams of 100 grains per accession was recorded. Analysis was conducted on the mean trait values across years.

Genomic analysis

Genotypes were available for all of the accessions [53]. GWAS was carried out on 404,627 SNP markers, using the statistical genetics package Genome Association and Prediction Integrated Tool (GAPIT) [54]. SNPs with a minor allele frequency (MAF) less than 0.05 and with more than 20% missing data were removed from analysis, leaving 141,310 SNPs. A unified mixed linear model (MLM) [55] with kinship, which accounts for relatedness among the accessions in the panel, was performed [56]. Multiple testing was controlled with a false discovery rate (FDR) of 5% using the Benjamini and Hochberg procedure [57] implemented in GAPIT. Narrow-sense heritability was calculated in GAPIT using a compressed mixed linear model that uses the genetic marker-based kinship matrix to estimate additive genetic effects [54]. Linkage disequilibrium (LD) was calculated using Tassel 5.2 [58]. Prior to conducting GWAS, we carried out an extensive literature search to identify potential candidate genes, and used *Sorghum bicolor* genome v1.4 from Phytozome [59] to compile a list of previously identified candidate genes associated with grain composition [35, 36, 60], as well as genes known to be involved in grain maturation and grain filling [9, 11, 61, 62] in Arabidopsis, rice, and maize, resulting in a list of 430 *a priori* gene candidates (Additional file 1). To analyze

population structure of the SAP, we used previously published genetic groupings that were determined through Bayesian hierarchical clustering analysis [29]. Five genetic groupings were used and we designated them group A through E (Additional file 2).

Expression data

To identify candidate genes within a significantly associated region, we used RNAseq data that was generated as a community resource for transcriptomic analyses [63]. Genes in a QTL region that were expressed during grain maturation were considered strong candidates. Expression levels were reported in fragments per kilobase of exon per million reads mapped (FPKM). We used the definitions of Davidson et al [63], as follows: $FPKM \leq 1$ = “not expressed”; $FPKM \leq 4$ = “low-expressed”; FPKM between 4 and 24 = “intermediate-expressed”; and $FPKM \geq 24$ = “high-expressed”.

Results

Phenotypic variation and heritability of sorghum grain composition

Overall, grain composition was similar across years, with protein, fat, and starch all having a strong correlation across years. Protein was the most consistent at 73–82% correlation between the 3 years, whereas fat (57–69%) and starch (51–65%) had slightly lower year to year correlations (Additional file 3). Similarly, protein had the highest narrow-sense heritability ($H^2 = 0.90$), followed by fat ($H^2 = 0.85$) and starch ($H^2 = 0.80$). Next, we investigated the range of sorghum grain protein, fat, and starch content and their covariation with each other using the mean of the 3 years (Additional file 4). The germplasm showed a wide range of diversity in grain composition. Protein content ranged from 8.3 to 18.8%, fat content ranged from 1.0 to 4.4%, and starch content ranged from 61.7 to 70.8% (Fig. 1). Pearson’s correlations were calculated between protein, fat, and starch (Fig. 1). There was a strong negative correlation between starch and both protein ($r = -0.90$, $p < 10^{-16}$) and fat ($r = -0.70$, $p < 10^{-16}$), and a strong positive correlation between protein and fat ($r = 0.77$, $p < 10^{-16}$). When grain composition concentrations are expressed as percentage by total seed weight, an increase in one component decreases the percentage of other components. Therefore, the percent concentration was multiplied by the seed weight of each accession to get absolute estimates of the mass of each constituent per grain, and Pearson’s correlations were recalculated. Using these estimates, there was a positive correlation between starch and both protein ($r = 0.66$, $p < 10^{-16}$) and fat ($r = 0.56$, $p < 10^{-16}$), and a strong positive correlation between protein and fat ($r = 0.85$, $p < 10^{-16}$). In contrast to correlations when using the percent concentration, the positive correlations between the mass of the traits

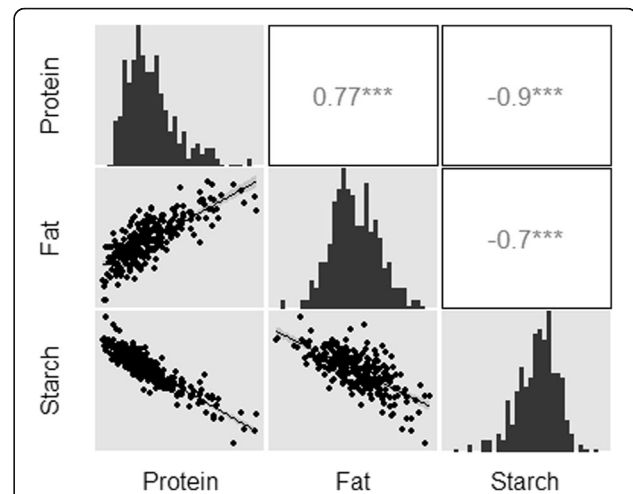


Fig. 1 Relationship of grain composition traits in a sorghum germplasm collection averaged over 3 years. The center diagonal presents histograms of each trait. The scatter plots with regression lines show the relationships between the traits. ($n = 265$)

reflect that total amounts of protein, fat, and starch increase with increases in total seed weight.

Next we investigated grain protein, fat, and starch covariation with factors that could reduce their biological availability for human consumption. Since the digestibility of protein and starch can be decreased by proanthocyanidins, and possibly other polyphenols [64], it is useful to know if there is a pattern of covariation between grain composition traits and polyphenol content. To this end, we used previously generated polyphenol data that was measured in the same samples as the current study [45] to calculate Pearson’s correlations with starch, protein, and fat concentrations (Additional file 5). Starch was negatively correlated with total polyphenols ($r = -0.34$, $p < 10^{-9}$), proanthocyanidins ($r = -0.29$, $p < 10^{-6}$), and 3-deoxyanthocyanidins ($r = -0.29$, $p < 10^{-6}$); protein was positively correlated with total polyphenols ($r = 0.42$, $p < 10^{-12}$), proanthocyanidins ($r = 0.34$, $p < 10^{-8}$), and 3-deoxyanthocyanidins ($r = 0.28$, 10^{-6}); while fat was only positively correlated with total polyphenols ($r = 0.25$, $p < 10^{-5}$) and proanthocyanidins ($r = 0.20$, $p = 0.001$).

Population structure of grain composition traits

Knowledge of variation in grain composition across genetically similar sorghum groups can be applied to germplasm utilization. Using five genetic groupings, population differences in grain composition were determined (Table 1). Group A consisted of 46 accessions that were primarily durras and bicolor-durras from Ethiopia and India; group C consisted of 55 accessions that were primarily kafirs from USA, India, and South Africa; group D consisted of 23 accessions that were primarily caudatum

Table 1 Population structure of grain composition traits in a global sorghum germplasm collection^a

Genetic group	Number of accessions	Protein (%) ^c	Fat (%) ^c	Starch (%) ^c	Predominant race	Predominant country
A	46	12.6 ± 2.3 ^a	3.0 ± 0.6 ^a	66.1 ± 1.6 ^a	durra, bicolor-durra	Ethiopia, India
B ^b	2	13.1 ± 3.4	3.5 ± 1.4	65.9 ± 1.0	bicolor-guinea	Ethiopia
C	55	10.9 ± 2.0 ^b	2.7 ± 0.6 ^{a,b}	67.6 ± 1.2 ^b	kafir	USA, India, S. Africa
D	23	11.6 ± 1.8 ^a	3.0 ± 0.5 ^a	67.2 ± 1.3 ^b	caudatum, guinea	Nigeria, USA
E	117	11.2 ± 1.8 ^b	2.6 ± 0.6 ^b	67.1 ± 1.2 ^b	caudatum	Sudan, USA, Ethiopia, Uganda

^aData represent mean grain composition concentrations across 3 years

^bGenetic grouping B was not considered in analysis due to the small sample size

^cValues in the same column with different letters are significantly different from each other, based on a post hoc Tukey HSD test

and guineas from Nigeria and USA; and group E consisted of 117 accessions that were primarily caudatums from Sudan and USA. Group B consisted of only 2 accessions, so was not included in analysis. Group A had the highest protein (12.6%) and fat (3.0%) and the lowest starch (66.1%), while group C had the lowest protein (10.9%) and the highest starch (67.6%).

Genome wide association study

To investigate the genetic basis of natural variation of protein, fat, and starch in sorghum grain, we conducted a GWAS using 265 accessions from the diverse association panel. Control experiments to support the validity of the GWAS results are described in the supplemental material (Additional file 6). Using the estimated mass of protein, fat, and starch, the MLM identified 4, 41, and 0 significant SNPs, respectively, at a genome-wide FDR of 5% and MAF ≥ 0.05 (Fig. 2a-c; Additional file 7). For both protein and fat, there was an association peak on chromosome 4 between 57.6 and 57.7 Mb and on chromosome 2 between 57.6 and 57.7 Mb. All of the significant SNPs on chromosome 2 (with the exception of S2_57592740) are in partial to strong LD with each other ($r^2 = 0.5-1.0$). Most of the SNPs in the chromosome 2 peak are in partial LD ($r^2 = 0.5$) with an *a priori* candidate gene that is a putative homolog of alpha-amylase 3 (AMY3, Sb02g023790; 57,701,214–57,703,517 bp). This gene is expressed in the day 10 seeds (2.1 FPKM) and in the endosperm (5.8 FPKM; Additional file 8).

Since starch makes up the majority of the grain, it is possible that some variation in protein and fat content are driven by variation in starch content. We hypothesized that natural variation in starch pathways might be affecting protein and fat content in the grain due to a limited pool of carbon. To determine if starch could be influencing the values, we ran two linear models in which we fit either protein or fat as the dependent variable and starch as the independent variable, using their estimated mass. If we assume that patterns in protein and fat are driven by starch, then starch could account for a significant proportion of the variance—45% of the variance in protein ($p < 10^{-16}$) and 32% of the variance

in fat ($p < 10^{-16}$)—but there is a large portion of variance still unexplained. Therefore, we conducted GWAS on the residuals (the amount of variation in protein and fat that could not be explained by starch) from the linear models to determine if accounting for this variation allowed for more accurate mapping results. The GWAS for protein and fat residuals identified 40 and 45 significant SNPs, respectively, at the FDR adjusted significance threshold, all within the peak on chromosome 2 at ~57.6 Mb and the peak on chromosome 4 at ~57.6 Mb (Fig. 3a-b; Additional file 7).

Discussion

QTL for kernel composition

GWAS for protein and fat in the sorghum global diversity panel identified two major peaks in common, one on chromosome 2 at 57.7 Mb and the other on chromosome 4 at 57.7 Mb. The peak on chromosome 2 at 57.7 Mb remained when GWAS was performed on the individual biological replicates in each year (Additional file 1). The peak is near a grain fat QTL from a sorghum linkage study that used a biparental population (Rio X BTx623) grown in Texas [41]. The previously identified grain fat QTL on chromosome 2 is near the genetic marker txp298 at ~57 Mb [65]. A promising *a priori* candidate near this peak is an AMY3 homolog. AMY3 is an alpha-amylase debranching enzyme that hydrolyzes the glucosidic bonds that make up starch. AMY1 was previously identified as a candidate gene in a maize grain composition GWAS study [35]. A recent study using AMY3 overexpression lines found that the increased levels of AMY3 did not significantly affect starch content, but fat content was increased in the mature endosperm where starch had been partially degraded [66]. The authors suggested that starch degradation during grain maturation led to the release of sucrose that was then shunted into the Kennedy pathway for fat synthesis.

Improvement of sorghum grain composition for human nutrition

The range of protein, fat, and starch content found in our diverse association panel may be useful for sorghum

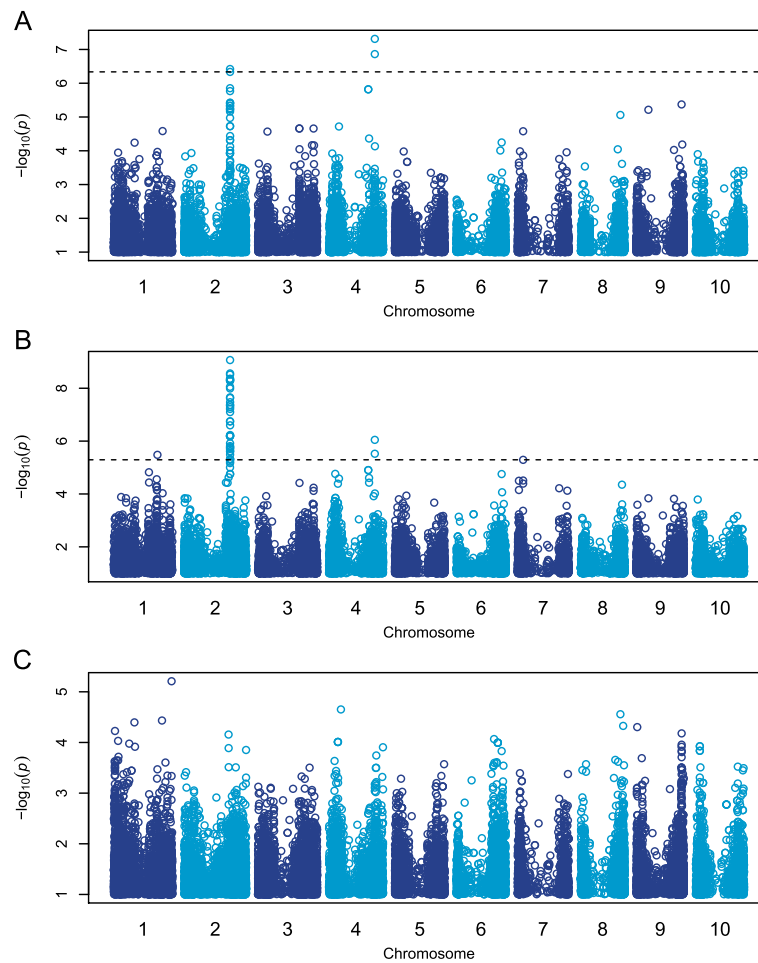


Fig. 2 GWAS for protein, fat, and starch content in sorghum grain. Manhattan plots of association results from a MLM analysis using 265 accessions. Each point represents a SNP, with the $-\log_{10}(p)$ values plotted against the position on each chromosome. SNPs with MAF < 0.05 were removed. The horizontal dashed line represents the genome-wide significance threshold at 5% FDR. **a** protein; **b** fat; **c** starch

improvement. Genetic group A, consisting of durra and bicolor-durra sorghums had significantly higher mean protein levels than the other groups, and are promising sources of genetic material for high protein sorghums. Durra sorghums are genetically similar to bicolor sorghums [29], which is the least derived race (i.e., retains most similarity to wild ancestors among the races), and high protein varieties may have been inadvertently counter-selected during cereal domestication when high starch varieties were selected. It may be that human selection for different food uses influenced the patterns of grain composition distribution among genetic groups (e.g., thick porridge from one region requires a certain grain composition, whereas flat bread from another region requires a different grain composition). It could also be that adaptation to environmental factors is driving some of the grain composition differences between genetic groups. Evidence of this adaptation was recently found for tannins in sorghum grain, when a variant of

the *Tannin1* gene, which controls the presence of tannins in the grain, was found to be correlated with several bioclimatic factors [53].

This study provides genetic trait association loci that can be explored further for their potential use in molecular breeding to modify the composition of grain sorghum. The high heritability of each trait suggests the genetic contribution to variation is strong. However, a GWAS with the SAP grown in Kansas (Additional file 6) did not identify the same large association peaks identified in the GWAS in the current study, suggesting a genotype-by-environment interaction. Several previous studies have found grain composition variation between environments, indicating that at least some genes may only be influential in a particular environment [67]. For example, in the biparental population (Rio X BTx623) grown in Texas genotype-by-environment effects explained a significant proportion of phenotypic variability in grain protein, fat, and starch [41]. This suggests that a

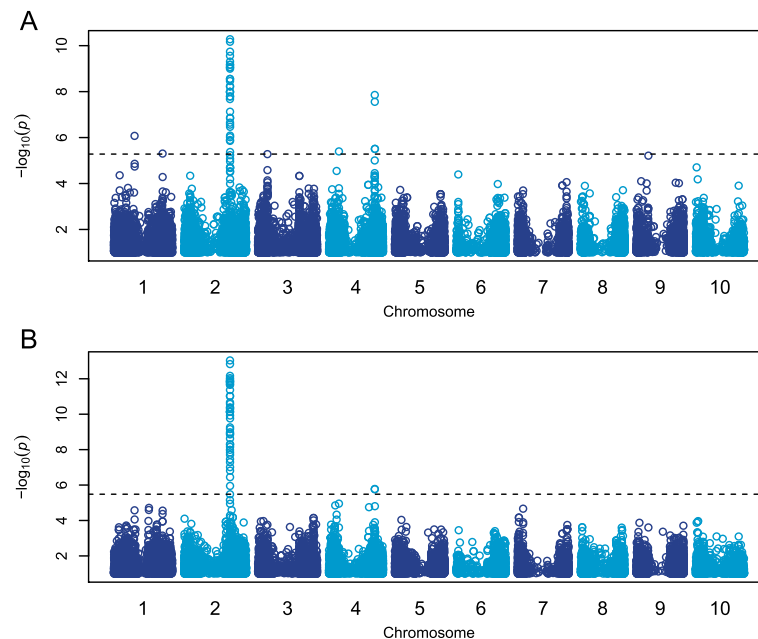


Fig. 3 Residuals GWAS for protein and fat content in sorghum grain. Manhattan plots of association results from a MLM analysis using 265 accessions. Each point represents a SNP, with the $-\log_{10}(p)$ values plotted against the position on each chromosome. SNPs with MAF < 0.05 were removed. The horizontal dashed line represents the genome-wide significance threshold at 5% FDR. **a** protein; **b** fat

more systematic investigation of genotype-by-environment interaction on grain composition may be needed to guide breeding efforts.

Genetic correlations among traits can complicate improvement of any single trait. The shared QTL for protein and fat in sorghum grain raises the question of whether protein and fat levels can be selected independently. Several other studies have found strong correlations and shared QTLs between protein, fat, and starch, as well as between these traits and grain yield [6, 35, 39, 41, 43, 68–71]. Shared genetic controls or developmental mechanisms of the grain components may be the cause of the correlations, however, some of the correlations may be due to evolutionary correlations rather than a shared genetic or developmental basis. Further studies to identify genes that control each grain composition trait could be useful. Since biparental mapping populations break up the evolutionary correlations present in association panels, they can be used to determine if the associations are due to a shared genetic basis or to evolutionary history.

Conclusions

Promising sources of genetic material for manipulation of grain composition traits have been identified, as well as several loci and candidate genes that may control sorghum grain composition. The starch GWAS did not identify any significant SNP associations, implying that, given the high heritability of starch and the lack of significant QTL, starch variation is likely controlled by

many small effect genes. Biparental mapping or nested association mapping may be helpful in identifying starch gene candidates. Identification of a highly significant peak on chromosome 2 associated with protein and fat provides a good starting point for marker-assisted breeding of sorghum grain composition traits. This survey of grain composition in sorghum germplasm and identification of QTL significantly associated with protein and fat contributes to our understanding of the genetic basis of natural variation in sorghum grain composition.

Additional files

Additional file 1: *a priori* gene candidate list. (XLSX 49 kb)

Additional file 2: Five genetic groupings of sorghum association panel. (XLSX 18 kb)

Additional file 3: Relationship of three separate years of grain composition traits in a sorghum germplasm collection. (PDF 138 kb)

Additional file 4: Protein, fat, and starch content of sorghum accessions. (XLSX 21 kb)

Additional file 5: Relationship within and between grain composition traits and polyphenol content. (PDF 90 kb)

Additional file 6: Table S6.1. Correlations between traits in Kansas panel⁹ and South Carolina panel. **Figure S6.1.** QQ plots. **Figure S6.2.** GWAS for protein, fat, and starch content in sorghum grain grown in 2012. **Figure S6.3.** GWAS for protein, fat, and starch content in sorghum grain grown in 2013. **Figure S6.4** GWAS for protein, fat, and starch content in sorghum grain grown in 2014. **Figure S6.5.** GWAS for protein, fat, and starch content in sorghum grain grown in Kansas. **Figure S6.6.** GWAS for flowering time in sorghum grain. (PDF 1087 kb)

Additional file 7: Significant SNPs associated with protein, fat, and starch. (XLSX 29 kb)

Additional file 8: Expression data for genes under major peaks. (XLSX 25 kb)

Abbreviations

AMY3: Alpha-amylase 3; FPKM: Fragments Per Kilobase of exon per Million fragments mapped; GRIN: Germplasm Resources Information Network; GWAS: Genome-wide association study; Mb: Mega basepairs; MLM: Mixed linear model; NIRS: Near-infrared spectroscopy; QTL: Quantitative trait loci; SNP: Single-nucleotide polymorphism

Acknowledgements

Not applicable.

Funding

This research was funded by the United Sorghum Checkoff Program (HVM06-14).

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors' contributions

DR helped conceive the study, participated in its design and coordination, carried out data analysis and drafted the manuscript. WR and LH critically revised the manuscript and generated the NIRS data; TH and SB critically revised the manuscript; RB and ZB assisted in field design and phenotyping, and critically revised the manuscript; SK helped conceive the study, participated in its design and coordination, and critically revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Disclaimer

Names are necessary to report factually on available data; however, the U.S. Department of Agriculture neither guarantees nor warrants the standard of the product, and use of the name by the U.S. Department of Agriculture implies no approval of the product to the exclusion of others that may also be suitable.

Author details

¹Center for Grain and Animal Health Research, USDA-ARS, 1515 College Ave, Manhattan, Kansas 66502, USA. ²Department of Soil & Crop Sciences, Texas A&M University, College Station, TX 77843, USA. ³Institute of Translational Genomics, Clemson University, Clemson, SC 29634, USA.

Received: 29 July 2016 Accepted: 9 December 2016

Published online: 05 January 2017

References

- FAO. FAO Statistical Pocketbook. Food and Agriculture Organization of the United Nations; 2015. <http://www.fao.org/documents/card/en/c/383d384a-28e6-47b3-a1a2-2496a9e017b2/>. Accessed 06 Feb 2016.
- National Research Council (U.S.), editor. Lost crops of Africa, Volume I Grains. Washington: National Academy Press; 1996.
- FAO. The State of Food Insecurity in the World: The multiple dimensions of food security. Food and Agriculture Organization. 2013. <http://www.fao.org/docrep/018/i3434e/i3434e00.htm>. Accessed 17 April 2016.
- Grusak MA, DellaPenna D. Improving the nutrient composition of plants to enhance human nutrition and health. *Annu Rev Plant Physiol Plant Mol Biol*. 1999;50:133–61.
- Hopkins CG. Improvement in the chemical composition of the corn kernel. *J Am Chem Soc*. 1899;21:1039–57.
- Goldman IL, Rocheford TR, Dudley JW. Quantitative trait loci influencing protein and starch concentration in the Illinois Long Term Selection maize strains. *Theor Appl Genet*. 1993;87:217–24.
- Dudley JW, RJ Lambert. 100 Generations of Selection for Oil and Protein in Corn. In *Plant Breeding Reviews*. edited by Jules Janick. Oxford: John Wiley & Sons, Inc.; 2010. pp. 79–110. <http://doi.wiley.com/10.1002/9780470650240.ch5>.
- Moose SP, Dudley JW, Rocheford TR. Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends Plant Sci*. 2004;9:358–64.
- Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. Combined networks regulating seed maturation. *Trends Plant Sci*. 2007;12:294–300.
- Baud S, Dubreucq B, Miquel M, Rochat C, Lepiniec L. Storage reserve accumulation in arabidopsis: metabolic and developmental control of seed filling. *Arab Book*. 2008;6:e0113.
- Vicente-Carbajosa J, Carbonero P. Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol*. 2005;49:645–51.
- Mertz ET, Bates LS, Nelson OE. Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*. 1964;145:279–80.
- Nelson OE, Mertz ET, Bates LS. Second mutant gene affecting the amino acid pattern of maize endosperm proteins. *Science*. 1965;150:1469–70.
- Schmidt RJ, Burr FA, Aukerman MJ, Burr B. Maize regulatory gene *opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc Natl Acad Sci U S A*. 1990;87:46–50.
- Coleman CE, Lopes MA, Gillikin JW, Boston RS, Larkins BA. A defective signal peptide in the maize high-lysine mutant *floury 2*. *Proc Natl Acad Sci U S A*. 1995;92:6828–31.
- Poneleit CG, Alexander DE. Inheritance of linoleic and oleic acids in maize. *Science*. 1965;147:1585–6.
- Mikkilineni V, Rocheford TR. Sequence variation and genomic organization of *fatty acid desaturase-2 (fad2)* and *fatty acid desaturase-6 (fad6)* cDNAs in maize. *Theor Appl Genet*. 2003;106:1326–32.
- Wassom JJ, Mikkilineni V, Bohn MO, Rocheford TR. QTL for fatty acid composition of maize kernel oil in Illinois High Oil × B73 backcross-derived lines. *Crop Sci*. 2008;48:69–78.
- Chourey PS, Nelson OE. The enzymatic deficiency conditioned by the *shrunk-1* mutations in maize. *Biochem Genet*. 1976;14:1041–55.
- Shure M, Wessler S, Fedoroff N. Molecular identification and isolation of the Waxy locus in maize. *Cell*. 1983;35:225–33.
- Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell Online*. 2004;16:2719–33.
- Karper RE. Inheritance of waxy endosperm in sorghum. *J Hered*. 1933;24:257–62.
- Lichtenwalner RE, Ellis EB, Rooney LW. Effect of incremental dosages of the waxy gene of sorghum on digestibility. *J Anim Sci*. 1978;46:1113–9.
- Rooney LW, Pflugfelder RL. Factors affecting starch digestibility with special emphasis on sorghum and corn. *J Anim Sci*. 1986;63:1607–23.
- Martin JH. Sorghum improvement. In: *Yearbook of Agriculture*. Washington: USDA; 1936. p. 523–60.
- Boyer CD, Liu K-C. Starch and water-soluble polysaccharides from sugary endosperm of sorghum. *Phytochemistry*. 1983;22:2513–5.
- Singh R, Axtell JD. High lysine mutant gene (*hl*) that improves protein quality and biological value of grain sorghum. *Crop Sci*. 1973;13:535–9.
- Boyles RE, Cooper EA, Myers MT, Brenton Z, Rauh BL, Morris GP, et al. Genome-wide association studies of grain yield components in diverse sorghum germplasm. *Plant Genome*. 2016;9:1–17.
- Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci U S A*. 2013;110:453–8.
- Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*. 2012;490:497–501.
- Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, et al. Genome-wide genetic changes during modern breeding of maize. *Nat Genet*. 2012;44:812–5.
- Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun*. 2011;2:467.
- Owens BF, Lipka AE, Magallanes-Lundback M, Tiede T, Diepenbrock CH, Kandianis CB, et al. A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. *Genetics*. 2014;198:1699–716.

34. Lipka AE, Gore MA, Magallanes-Lundback M, Mesberg A, Lin H, Tiede T, et al. Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3:GenesGenomesGenetics*. 2014;3:1287–99.
35. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, et al. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol*. 2012;158:824–34.
36. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*. 2013;45:43–50.
37. Rasmussen SK, Shu X. Quantification of amylose, amylopectin, and β -glucan in search for genes controlling the three major quality traits in barley by genome-wide association studies. *Crop Sci Hortic*. 2014;5:197.
38. Shu X, Backes G, Rasmussen SK. Genome-wide association study of resistant starch (RS) phenotypes in a barley variety collection. *J Agric Food Chem*. 2012;60:10302–11.
39. Sukumaran S, Xiang W, Bean SR, Pedersen JF, Kresovich S, Tuinstra MR, et al. Association mapping for grain quality in a diverse sorghum collection. *Plant Genome J*. 2012;5:126–35.
40. de Alencar Figueiredo LF, Sine B, Chantereau J, Mestres C, Fliedel G, Rami J-F, et al. Variability of grain quality in sorghum: association with polymorphism in *Sh2*, *Bt2*, *Sss1*, *Ae1*, *Wx* and *O2*. *Theor Appl Genet*. 2010;121:1171–85.
41. Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, et al. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Sci*. 2008;48:2165–79.
42. Hamblin MT, Salas Fernandez MG, Tuinstra MR, Rooney WL, Kresovich S. Sequence variation at candidate loci in the starch metabolism pathway in sorghum: prospects for linkage disequilibrium mapping. *Crop Sci*. 2007;47:5-125–34.
43. Rami J-F, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, et al. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theor Appl Genet*. 1998;97:605–16.
44. McIntyre CL, Drenth J, Gonzalez N, Henzell RG, Jordan DR. Molecular characterization of the waxy locus in sorghum. *Genome*. 2008;51:524–33.
45. Rhodes DH, Hoffmann L, Rooney WL, Ramu P, Morris GP, Kresovich S. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] Germplasm. *J Agric Food Chem*. 2014;62:10916–27.
46. Shakoob N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, Connolly EL, et al. Integration of experiments across diverse environments identifies the genetic determinants of variation in *Sorghum bicolor* seed element composition. *Plant Physiol*. 2016;170:1989–98.
47. Seleka TB, Jackson JC, Batsetswe L, Kebakile PG. Small-scale milling and the feasibility of mandatory fortification of sorghum and maize flour in Botswana. *Dev South Afr*. 2011;28:461–76.
48. Nestel P, Bouis HE, Meenakshi JV, Pfeiffer W. Biofortification of staple food crops. *J Nutr*. 2006;136:1064–7.
49. Horton S. The economics of food fortification. *J Nutr*. 2006;136:1068–71.
50. Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, et al. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor Appl Genet*. 2005;111:23–30.
51. Harlan JR, Wet D, J JM. A simplified classification of cultivated sorghum. *Crop Sci*. 1972;12:172–6.
52. Germplasm Resources Information Network. USDA National Genetic Resources Program. 2016. <http://www.ars-grin.gov>. Accessed 04 July 2016.
53. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, et al. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv*. 2015;1:e1400218.
54. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;28:2397–9.
55. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006;38:203–8.
56. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. 2010;42:355–60.
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57:289–300.
58. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A High capacity genotyping by sequencing analysis pipeline. *PLoS One*. 2014;9:e90346.
59. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:D1178–86.
60. Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, et al. Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat Genet*. 2008;40:1370–4.
61. Holdsworth MJ, Bentsink L, Soppe WJJ. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol*. 2008;179:33–54.
62. Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J*. 2008;54:608–20.
63. Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu S-H, et al. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J*. 2012;71:492–502.
64. Duodu K, Taylor JR, Belton P, Hamaker B. Factors affecting sorghum protein digestibility. *J Cereal Sci*. 2003;38:117–31.
65. Mace ES, Jordan DR. Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. *Theor Appl Genet*. 2011;123:169–91.
66. Whan A, Dielen A-S, Mieog J, Bowerman AF, Robinson HM, Byrne K, et al. Engineering α -amylase levels in wheat grain suggests a highly sophisticated level of carbohydrate regulation during development. *J Exp Bot*. 2014;65:5443–52.
67. Beta T, Corke H. Genetic and environmental variation in sorghum starch properties. *J Cereal Sci*. 2001;34:261–8.
68. Yang G, Dong Y, Li Y, Wang Q, Shi Q, Zhou Q. Verification of QTL for grain starch content and its genetic correlation with oil content using two connected RIL populations in high-oil maize. *PLoS One*. 2013;8:e53770.
69. Li Y, Wang Y, Wei M, Li X, Fu J. QTL identification of grain protein concentration and its genetic correlation with starch concentration and grain weight using two populations in maize (*Zea mays* L.). *J Genet*. 2009;88:61–7.
70. Zhang J, Lu XQ, Song XF, Yan JB, Song TM, Dai JR, et al. Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. *Euphytica*. 2008;162:335–44.
71. Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci*. 2005;45:2015–22.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

