Received: 2 August 2019 | Accepted: 19 January 2020 | Published online: 26 March 2020

DOI: 10.1002/tpg2.20013

The Plant Genome

ORIGINAL RESEARCH

Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm

Clara Cruet-Burgos[1] | **Sarah Cox**[3] | **Brian P. Ioerger**[3] | **Ramasamy Perumal**[2] | **Zhenbin Hu**[1] | **Thomas J. Herald**[3] | **Scott R. Bean**[3] | **Davina H. Rhodes**[1]

[1]Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA

[2]Agricultural Research Center, Kansas State University, Hays, KS 67601, USA

[3]Center for Grain and Animal Health Research, USDA-ARS, 1515 College Ave, Manhattan, KS 66502, USA

Correspondence
Davina H. Rhodes, Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA.
Email: davinarhodes@ksu.edu

Funding information
National Institute of Food and Agriculture, Grant/Award Number: 2017-67012-26098

Abstract

Vitamin A deficiency is one of the most prevalent nutritional deficiencies worldwide. Sorghum [*Sorghum bicolor* L. (Moench)] is a major cereal crop consumed by millions of people in regions with high vitamin A deficiency. We quantified carotenoid concentrations in a diverse sorghum panel using high-performance liquid chromatography and conducted a genome-wide association study (GWAS) of grain carotenoids to identify genes underlying carotenoid variation. There was moderate variation for β-carotene (00.8 μg g^{-1}), lutein (0.3–9.4 μg g^{-1}), and zeaxanthin (0.2–9.1 μg g^{-1}), but β-cryptoxanthin and α-carotene were nearly undetectable. Genotype had the largest effect size, at 81% for zeaxanthin, 62% for β-carotene, and 53% for lutein. Using multiple models, GWAS identified several significant associations between carotenoids and single nucleotide polymorphisms (SNPs), some of which colocalized with known carotenoid genes that have not been previously implicated in carotenoid variation. Several of the candidate genes identified have also been identified in maize (*Zea mays* L.) and Arabidopsis (*Arabidopsis thaliana*) carotenoid GWAS studies. Notably, an SNP inside the putative ortholog of maize *zeaxanthin epoxidase* (*ZEP*) had the most significant association with zeaxanthin and with the ratio between lutein and zeaxanthin, suggesting that *ZEP* is a major gene controlling sorghum carotenoid variation. Overall findings suggest there is oligogenic inheritance for sorghum carotenoids and

Abbreviations: DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; FDR, false discovery rate; FPKM, fragments per kilobase of exon per million reads mapped; GAPIT, Genome Association and Prediction Integrated Tool; GGPP, geranylgeranyl diphosphate; GGPPS, geranylgeranyl diphosphate synthase; GWAS, genome-wide association study; HPLC, high-performance liquid chromatography; LD, linkage disequilibrium; MAS, marker-assisted selection; MEP, methylerythritol phosphate; MLM, mixed linear model; MLMM, multi-locus mixed model; NPGS, National Plant Germplasm System; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; SPE, solid phase extraction; ZDS, δ-carotene desaturase; ZEP, zeaxanthin epoxidase.

suitable variation for marker-assisted selection. The high carotenoid germplasm and significant associations identified in this study can be used in biofortification efforts to improve the nutritional quality of sorghum.

1 | INTRODUCTION

Carotenoids are a diverse group of yellow, orange, and red plant pigments, such as the orange β-carotene in carrots and the red lycopene in tomatoes, that are made by all photosynthetic organisms (Cazzonelli & Pogson, 2010). In plants, carotenoids and their derivatives are involved in photosynthesis, seed maturation and dormancy, and protection against oxidation (Esteban, Moran, Becerril, & García-Plazaola, 2015). In humans, dietary carotenoids are thought to act as antioxidants and anti-inflammatories (Eggersdorfer & Wyss, 2018; Kaulmann & Bohn, 2014) and are associated with protection against heart disease (Kulczyński, Gramza-Michałowska, Kobus-Cisowska, & Kmiecik, 2017), cancer (Tanaka, Shnimizu, & Moriwaki, 2012), cataracts, and age-related macular degeneration (Wu, Cho, Willett, Sastry, & Schaumberg, 2015). Additionally, some dietary carotenoids can be converted into vitamin A in the human body (provitamin A carotenoids). Vitamin A is an essential vitamin that cannot be produced by the body and is critical for immune function, vision, reproduction, growth, and development (Tanumihardjo et al., 2016). β-Carotene, α-carotene, and β-cryptoxanthin are the major provitamin A carotenoids.

Vitamin A deficiency is one of the most prevalent nutritional deficiencies worldwide (WHO, 2009). It increases the risk of death from common childhood illnesses such as measles and diarrhea and is the leading cause of preventable blindness in children under 5 years of age. Cereal crops provide the majority of nutrients in many regions where vitamin A deficiency is prevalent; however, most cereal crops are poor sources of provitamin A carotenoids (Mellado-Ortega & Hornero-Méndez, 2015; Zhai, Xia, & He, 2016). Therefore, an increase in cereal provitamin A carotenoids could significantly contribute to alleviating vitamin A deficiency in these regions.

Sorghum[*Sorghum bicolor* L. (Moench)] is a staple food for millions of people in semi-arid regions of sub-Saharan Africa and Asia (FAO, 1995), where the highest rates of vitamin A deficiency are found (Stevens et al., 2015). Sorghum is a genetically diverse cereal crop grown for food, feed, and fiber (National Research Council, 1996). Early research in sorghum established that carotenoids are present in the grain of some varieties (Blessin, Vanetten, & Wiebe, 1958; Gorbet, 1971; Suryanarayana Rao, Rukmini, & Mohan, 1968), and later work established that quantitative genetic variation for grain carotenoid levels exists among sorghum varieties (Cardoso et al., 2015; Fernandez et al., 2009; Kean, Ejeta, Hamaker, & Ferruzzi, 2007). Although many carotenoid-containing sorghums have been identified, none has contained levels high enough to be biologically relevant to human nutrition. Therefore, biofortification is needed to increase the provitamin A carotenoid concentrations.

Biofortification through either breeding or biotechnology can be used to increase carotenoid concentrations in sorghum grain. A transgenic sorghum with high levels of provitamin A carotenoids has been developed by the Africa Biofortified Sorghum project (Africa Harvest Biotech Foundation International, 2010, 2012); however, regulatory restrictions on genetically modified crops have prevented public dissemination and production of the transgenic sorghum in any country (Arvidsson, 2015; Azadi et al., 2015). Effective breeding strategies are therefore needed for biofortification of sorghum grain carotenoids, but developing a biofortified variety with acceptable agronomic and end-use traits can take years using traditional plant breeding methods. Genomics-enabled breeding, however, has the potential to accelerate the development of nutritionally enhanced food crops (Diepenbrock & Gore, 2015). A major goal in carotenoid biofortification efforts, therefore, is to characterize the genetic architecture of carotenoid traits in important food crops in order to identify appropriate breeding strategies and develop molecular breeding resources.

Marker-assisted selection (MAS), which is a tool used in genomics-enabled breeding, uses molecular markers linked to a causal variant for a trait of interest (Varshney, Terauchi, & McCouch, 2014) and is an appropriate breeding strategy for oligogenic traits (traits controlled by a moderate number of genes of moderate effect). In this study we hypothesize (a) that sorghum grain carotenoids are oligogenic traits and (b) that the genetic controls of sorghum grain carotenoid variation are homologous with those identified in other plants. To test these hypotheses, we identified loci underlying sorghum grain carotenoid variation in a sorghum diversity panel using GWAS and identified candidate genes based on a list of a priori genes homologous to maize (*Zea mays* L.) carotenoid genes. We identified between eight to twelve regions of association for each carotenoid trait and determined that *ZEP* is likely a major-effect gene underlying sorghum carotenoid variation.

Core Ideas

- We quantified carotenoid variation in 403 diverse global sorghum accessions using HPLC.
- GWAS identified SNPs near previously unknown genes controlling carotenoid variation.
- GWAS identified SNPs in or near genes identified in maize and Arabidopsis GWAS.
- *Zeaxanthin epoxidase (ZEP)* was identified as a major gene underlying carotenoid variation.

2 | MATERIALS AND METHODS

2.1 | Germplasm

We investigated 403 sorghum accessions from the USDA National Plant Germplasm System (NPGS) (NPGS, 2015). The panel included 321 accessions from the sorghum association panel (Casa et al., 2008) and 82 accessions that were selected based on the presence of yellow endosperm or yellow grain color and photoperiod insensitivity using the NPGS Germplasm Resources Information Network descriptor data. Seeds were obtained through the NPGS and planted in early June 2015 at Kansas State University's Department of Agronomy North Farm in Manhattan, KS. A single replicate of a complete randomized block design was used. To compare among environments, 23 accessions with yellow endosperm and suitable agronomic traits (short stature, bold grain, minimal tillers) were grown in three locations, one of which was in a separate year: Manhattan, KS, a temperate humid environment, in 2015; Hays, KS, a semi-arid environment, in 2016; and Guayanilla, Puerto Rico, a tropical humid environment, in 2016. Samples from two accessions were identified as outliers using Tukey's method, and those accessions were removed from analysis. Panicles from each plot were collected at physiological maturity. Due to differences in maturity among these accessions, harvest occurred between September and November. The harvested panicles were oven-dried, mechanically threshed, and stored for ∼1 yr at −7°C until ready for use. Carotenoids were extracted and quantified using high-performance liquid chromatography (HPLC) as described below.

2.2 | Carotenoid extraction and quantification

Dry whole-grain sorghum was ground in a Udy Mill, vacuum sealed in plastic pouches, and stored in a container flushed with liquid nitrogen at −80°C until ready for use within 1 wk. Extraction was based on a method modified from Irakli, Samanidou, and Papadoyannis (2011) and Gupta, Sreelakshmi, and Sharma (2015). To reduce carotenoid degradation, extraction was carried out under yellow light. For extraction, 100 mg of ground sorghum, 100 mg ascorbic acid, and 2 ml ethanol were vortexed for 1 min and then placed in a water bath (80°C) for 5 min. For saponification, 100 µl potassium hydroxide (80% w/v in H_2O) was added, and the sample was vortexed for 20 s and then returned to the water bath for 15 min, with mixing every 5 min. Samples were then cooled to room temperature and centrifuged for 5 min at 1810 *g*, and the supernatant was aspirated to a clean test tube. Ethanol (2 ml) was added to the remaining residue, sonicated for 1 min, and centrifuged for 5 min at 1810 *g*, and the supernatant was added to the clean test tube. Ethanol addition and sonication of residue was repeated one time, but with 1 ml ethanol. To clean up and concentrate the carotenoid extract, solid phase extraction (SPE) cartridges were used. Cartridges were conditioned with 3 ml methanol and 3 ml H_2O. We mixed 2 ml H_2O with the extract and passed through the SPE cartridge, followed by a 2 ml H_2O wash. To elute carotenoids, 2 ml dichloromethane were added to the SPE cartridge, and carotenoid extracts were collected into 3-ml V-vials. Extracts were then dried using a heat block (40°C) under a gentle stream of N_2 and resuspended in 25 µl methanol/ethyl acetate (50:50, v/v) by vortexing. After resuspension, extracts were transferred to glass inserts housed in amber HPLC vials.

For HPLC analysis, 2 µl of extract were injected into a Carotenoid C30 column (3 µm, 2 × 150 mm; YMC American, Inc.). Column temperature was maintained at 20°C. Mobile phase A was methanol/$H_2$0 (98:2, v/v), mobile phase B was methanol/H_2O (95:5, v/v), and mobile phase C was methyl tert-butyl ether. Carotenoids were resolved using the following linear gradient: 0.0 min: 80% A, 0% B, 20% C; 0.5 ml min^{-1}; 2.0 min: 60% A, 0% B, 40% C; 0.5 ml min^{-1}; 2.01 min: 0% A, 60% B, 40% C; 0.36 ml min^{-1}; 12.0 min: 0% A, 0% B, 100% C; 0.36 ml min^{-1}; 13.0 min: 80% A, 0% B, 20% C; 0.5 ml min^{-1}; 20.0 min: 80% A, 0% B, 20% C; 0.5 ml min^{-1}. Carotenoids were detected at 450 nm using a diode array detector on a 1260 HPLC (Agilent Technologies) and quantified relative to six-point standard curves for β-carotene, α-carotene, β-cryptoxanthin, zeaxanthin, and lutein.

2.3 | Analysis across environments

Eta squared was used to determine effect size of genotype, environment, and genotype × environment interactions. This was calculated by dividing the ratio of the sum of squares for each variable (calculated with ANOVA) by the total sum of squares and reported as a percentage. Both genotype and environment were considered fixed effects. The Fligner-Killeen

test of homogeneity of variances was used to determine homogeneity of variance across locations for each trait. Pearson's product moment correlation coefficient was used to determine the association between environments for each trait.

2.4 | Genome-wide analysis of candidate genes

Prior to conducting genome-wide association studies (GWAS), an extensive literature search was carried out to identify and compile a list of potential candidate genes. Genes associated with biosynthesis and degradation of seed carotenoids and carotenoid precursors in maize were identified through a literature search, and then sorghum homologs were identified using Phytozome's homolog data. All genes identified by Phytozome as homologs were added to the a priori list, regardless of percent similarity, resulting in a list of 106 a priori gene candidates (Supplemental Table S1). Sorghum reference genome v3.1 from Phytozome was used (McCormick et al., 2018). To identify candidate genes within a significantly associated region, linkage disequilibrium (LD) and gene expression were considered. Linkage disequilibrium between single nucleotide polymorphisms (SNPs) was calculated using Tassel 5.2 (Glaubitz et al., 2014). Linkage disequilibrium decay varies considerably depending on the chromosome, region of the chromosome, or population, and in the sorghum association panel the range of LD decay is 175–600 kb (Morris et al., 2013). Based on this LD range, we considered candidates up to ∼600 kb but focused on candidates within 200 kb. For gene expression, we used multiple published RNAseq datasets generated as community resources (Davidson et al., 2012; McCormick et al., 2018). Expression levels are reported in fragments per kilobase of exon per million reads mapped (FPKM).

2.5 | Genome-wide association studies

The sorghum association panel was previously genotyped (Hu, Olatoye, Marla, & Morris, 2019). The 93 additional yellow accessions were sequenced as previously described with genotyping-by-sequencing using ApeKI to construct the libraries and Illumina technology for sequencing (Hu et al., 2019). Sequences were aligned to the sorghum reference genome v3.1. The SNP datasets are available for download from the Dryad Data Repository (doi:10.5061/dryad.63h8fd4). After filtering out the SNPs with a minor allele frequency <5%, 341,514 out of 949,378 SNPs remained. There were both phenotype and genotype data, for a total of 311 accessions. The statistical models fitted in the GWAS assume a normal distribution of data, so all phenotypes were tested for normality and had to be corrected by log transformation. Broad-sense heritability was estimated using the statistical genetics package Genome Association and Prediction Integrated Tool (GAPIT) using a mixed linear model (MLM) that uses the genetic marker–based kinship matrix to estimate additive genetic effects (Lipka et al., 2012).

Genome-wide association studies using three different statistical models were implemented. Genome-wide association studies using a unified MLM (Yu et al., 2006) with kinship were carried out in R using GAPIT (Lipka et al., 2012). Multiple testing was controlled for using the procedure described by Benjamini and Hochberg (1995) implemented in GAPIT at a false discovery rate (FDR) of 5% unless otherwise noted. Genome-wide association studies using a multi-locus mixed model (MLMM) were also carried out in R (Segura et al., 2012). Statistical comparisons between thousands of SNPs will by chance lead to the identification of many false positives. To reduce false positives when conducting GWAS, the FDR can be controlled, but this can lead to false negatives (Segura et al., 2012). The MLMM is stepwise and iteratively accounts for the previously identified quantitative trait loci (QTL), improving the power to detect true positives. We conducted GWAS using both models in order to compare results to the MLM and to identify SNPs that may not have been identified by MLM. A biosynthetic pathway–level GWAS analysis was also carried out in R. Based on the previously mentioned LD decay range of 175–600 kb in the sorghum association panel, SNPs that were located within ±200 kb of candidate genes were identified with PLINK v1.07 software (Purcell et al. 2007). An R script was implemented to filter SNPs contained in each of these candidate gene windows. Associations between carotenoid concentrations and SNPs within each candidate gene window were tested by an MLM with kinship using GAPIT. For each trait and candidate gene window, adjusted *P* values were considered significant at an FDR of 5%.

3 | RESULTS

3.1 | Quantitative variation of grain carotenoids

We quantified β-carotene, β-cryptoxanthin, zeaxanthin, α-carotene, and lutein in the grain of 403 sorghum accessions using HPLC. The sorghum association panel contains all major morphological races and geographic centers of diversity in sub-Saharan Africa and Asia as well as important breeding lines from the United States. We first sought to determine the range of carotenoid concentrations and their covariation with each other. There was a wide range of variation for lutein, zeaxanthin, and β-carotene (Figure 1; Supplemental Table S2). Lutein was the most abundant carotenoid (range,

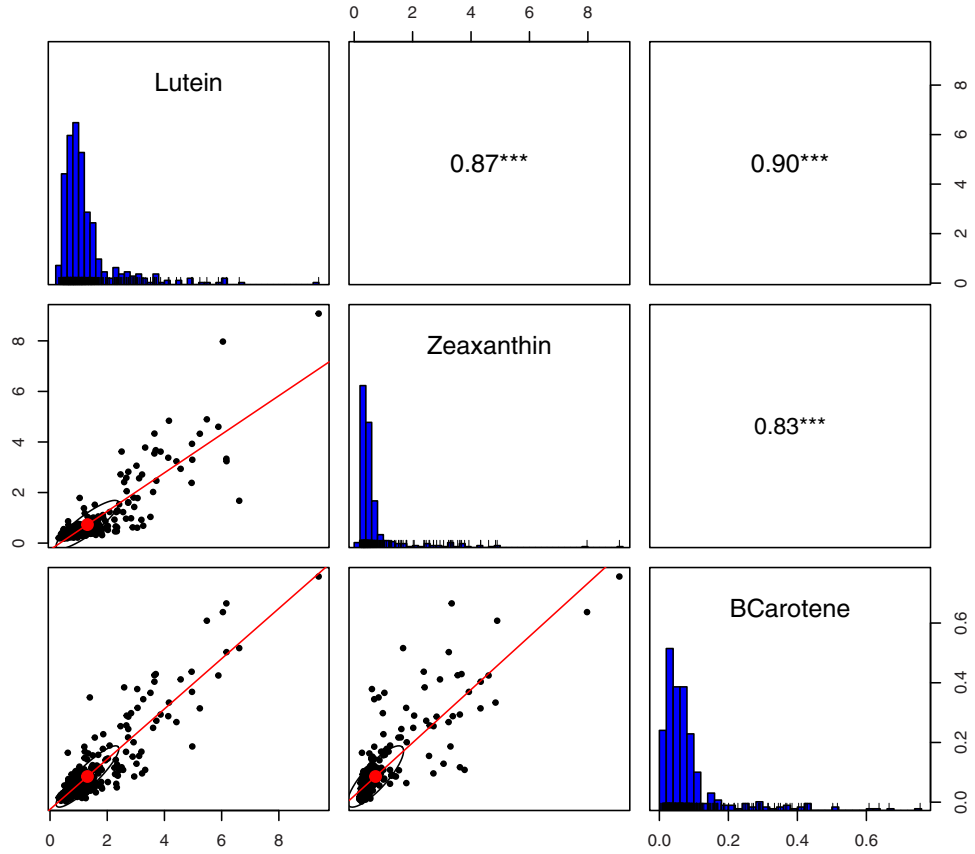

FIGURE 1 Relationship among grain carotenoid traits in a global sorghum germplasm collection. The center diagonal contains histograms of the concentrations of each trait ($n = 403$). The lower corner contains scatter plots with regression lines showing the relationships between the traits. The upper corner shows Pearson's correlations between the traits. ***Significance at the .001 level. Units are μg g^{-1}

0.3–9.4 μg g^{-1}; mean, 1.3 μg g^{-1}), followed by zeaxanthin (range, 0.2–9.1 μg g^{-1}; mean, 0.7 μg g^{-1}) and then β-carotene (range, 0–0.8 μg g^{-1}; mean, 0.1 μg g^{-1}). After measuring 120 accessions, β-cryptoxanthin was nearly undetectable (range, 0–0.07 μg g^{-1}; mean, 0 μg g^{-1}), and α-carotene was undetectable, so β-cryptoxanthin and α-carotene were not used in subsequent analyses. There were strong positive correlations between lutein and zeaxanthin ($r = .87$; $P < 2.2 \times 10^{-16}$), lutein and β-carotene ($r = .9$; $P < 2.2 \times 10^{-16}$), and zeaxanthin and β-carotene ($r = .83$; $P < 2.2 \times 10^{-16}$). Broad-sense heritability estimated from the genetic marker–based kinship matrix was 54.4% for β-carotene, 61.2% for zeaxanthin, and 56.9% for lutein.

3.2 | Environmental effect

To determine the effect of genotype and environment on carotenoid traits, eta-squared was calculated for 23 accessions grown in Manhattan, KS, a temperate humid environment; Hays, KS, a semi-arid environment; and Guayanilla, Puerto Rico, a tropical humid environment. Genotype had the largest effect size for each carotenoid trait, at 81% for zeaxanthin, 62% for β-carotene, and 53% for lutein, although there was also a significant genotype × environment interaction (Table 1). The homogeneity of variance across locations was not statistically significant for zeaxanthin, lutein, or β-carotene (Table 2). There was a strong positive correlation between locations for zeaxanthin and weaker correlations between locations for β-carotene and lutein (Table 2; Supplemental Figure S1). Our data were also compared with lutein and zeaxanthin data from Cardoso et al. (2015) on 72 sorghum association accessions grown in Brazil in 2010. For zeaxanthin, there was a significant correlation between locations ($r = .63$; $P = 3.7 \times 10^{-9}$), whereas for lutein, there was no significant correlation between locations ($r = .19$; $P = .1$) (Supplemental Figure S2).

TABLE 1 Effect size of genotype and environment on carotenoid traits for 21 sorghum accessions grown in three environments

Factors	df	Zeaxanthin	β-carotene	Lutein
Genotype (G)	20	80.7***	62.1***	52.6***
Environment (E)	2	1.6***	4.7***	6.7***
G × E	39	16.6***	31.8***	37.2***
Residuals	66	1.1***	1.4***	3.5***

***Significant at the .001 probability level.

TABLE 2 Range of carotenoid concentrations within and correlations between three environments for 21 sorghum accessions

	Range of concentrations[a]				Correlation between environments		
	Man	Hays	PR	Fligner-Killeen[b]	Man-Hays	Man-PR	Hays-PR
	μg g^{-1}						
Lutein	2.1–5.9	1.7–7.9	2.4–6.8	$P = .1$	0.42 ($P = .06$)	0.47 ($P = .03$)	0.34 ($P = .14$)
Zeaxanthin	1.0–7.1	0.4–8.8	0.8–8.1	$P = .7$	0.72 ($P < .001$)	0.68 ($P < .001$)	0.85 ($P < .001$)
β-Carotene	0.1–0.7	0.1–0.8	0.1–0.6	$P = .4$	0.59 ($P = .005$)	0.53 ($P = .01$)	0.42 ($P = .2$)

[a]Man, Manhattan; PR, Puerto Rico.

[b]Fligner-Killeen test of homogeneity of variances across the three locations. Nonsignificance indicates no statistical difference across the three locations.

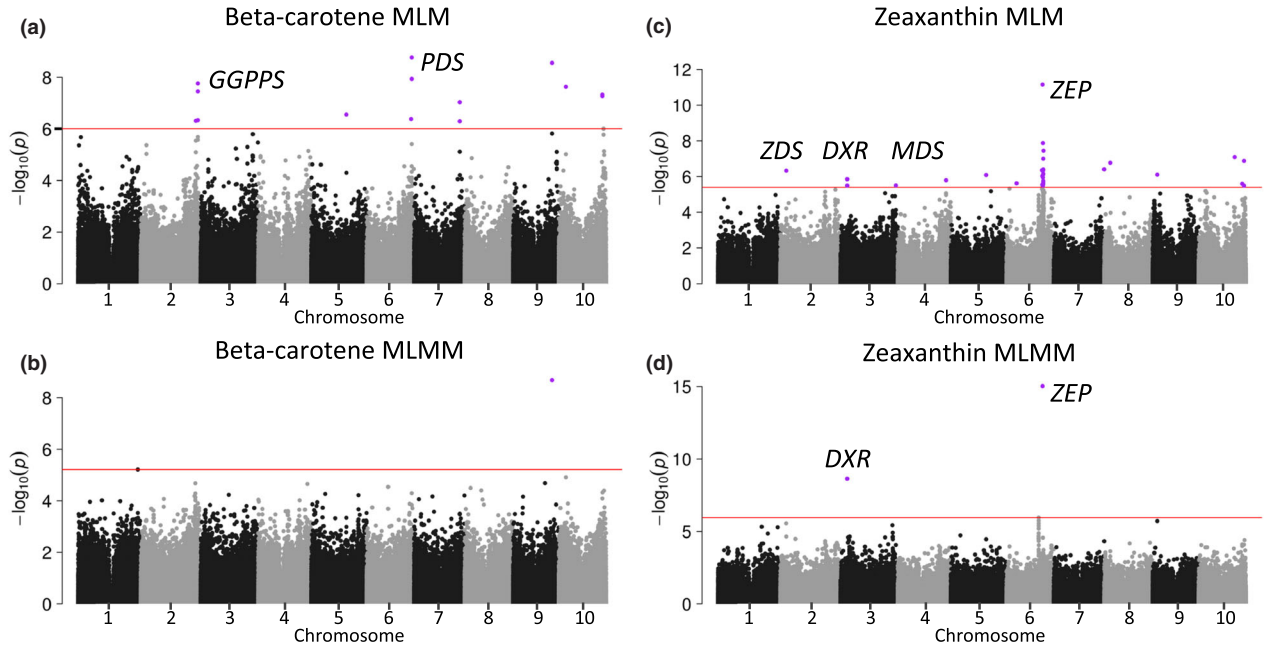

FIGURE 2 Genome-wide association study for carotenoid concentrations in sorghum grain. Manhattan plot of association results for (A) β-carotene mixed linear model (MLM), (B) β-carotene multi-locus mixed model (MLMM), (C) zeaxanthin MLM, and (D) zeaxanthin MLMM. Axes: The $-\log_{10} p$ values (y-axis) plotted against the position on each chromosome (x-axis). Each circle represents a single nucleotide polymorphism. The horizontal line represents the genome-wide significance threshold at 5% false discovery rate

3.3 | Genome-wide association studies

3.3.1 | β-Carotene

To investigate the genetic basis of carotenoid variation in sorghum grain, we conducted GWAS with 311 globally diverse sorghum accessions and 341,514 SNPs with a minor allele frequency >0.05. We first sought to identify loci associated with quantitative variation of β-carotene. An MLM identified 14 SNPs significantly associated with β-carotene, encompassing nine regions of association (Figure 2a; Supplemental Table S3). Only three of the regions were near a priori gene candidates. A significant association was identified on chromosome 2 at 74.1 Mb, 75 kb from a putative *CYP97A* hydroxylase gene (Sobic.002G389000). This gene is considered a secondary a priori gene candidate because it has 35% genetic similarity to the sorghum ortholog of the maize gene but does not itself have high genetic similarity to the maize gene. The *CYP97A* gene catalyzes the hydroxylation of the β-ring of α- and β-carotene. Another significant association was identified on chromosome 6 at 57.4 Mb, 95 kb from a phytoene desaturase (PDS; Sobic.006G232600). Phytoene desaturase is one of the desaturases that carries out a dehydrogenation reaction in the beginning of the carotenoid pathway. There was also a significant association identified on chromosome 2 at 71.5 Mb, 98 kb from an a priori candidate *geranylgeranyl diphosphate synthase* (*GGPPS*; Sobic.002G353300). Geranylgeranyl diphosphate synthase (GGPPS) catalyzes the condensation of two geranylgeranyl phosphates to produce geranylgeranyl diphosphate (GGPP) in the final reaction in the methylerythritol phosphate (MEP) pathway. The MEP pathway produces the five-carbon precursors for carotenoid biosynthesis.

To compare results with the MLM and to identify possible SNPs not identified by MLM, we conducted GWAS using an MLMM. The MLMM identified one significant SNP on

chromosome 9 at 50.5 Mb, which was also identified by MLM (Figure 2b; Supplemental Table S4). There were no nearby a priori gene candidates.

3.3.2 | Zeaxanthin

Next, a GWAS was conducted to identify genetic controls of zeaxanthin variation among the 311 accessions. An MLM identified 38 SNPs with an FDR adjusted *P* value <.05 that were significantly associated with zeaxanthin, encompassing 12 regions of association (Figure 2c; Supplemental Table S5). Of the 12 regions, only four colocalized with a priori gene candidates. One region of association with zeaxanthin was on chromosome 4 at 62.4 Mb (S4_62459432), 46 kb from a priori gene candidate *2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase* (*MDS*; Sobic.004G281900), an MEP pathway gene. A second region of association was near an a priori gene candidate on chromosome 2 at 7.5 Mb, 212 kb from δ-*carotene desaturase* (*ZDS*; Sobic.002G072400), which is one of the desaturases that carries out a dehydrogenation reaction in the early part of the carotenoid pathway. A third region of association was near an a priori gene candidate on chromosome 3 at 8.7 Mb, 476 kb from a *deoxyxylulose reductoisomerase* (*DXR*; Sobic.003G103300), another MEP pathway gene.

The strongest region of association in the GWAS was detected on chromosome 6 with the most significant SNP (S6_46717975) located inside a *zeaxanthin epoxidase (ZEP)* a priori gene candidate (Sobic.006G097500). Zeaxanthin epoxidase catalyzes epoxidation of zeaxanthin's β-ionone rings to produce antheraxanthin and violaxanthin, the first steps leading to synthesis of abscisic acid (Moise, Al-Babili, & Wurtzel, 2014). There were several other associations with SNPs between 46.0 and 47.9 Mb on chromosome 6, some of which were in partial LD with the peak SNP in the *ZEP* gene (R^2 = .43). There are two possibilities for the SNPs in this region. The GWAS model could be identifying two genes near each other that are both responsible for zeaxanthin variation, or the *ZEP* haplotype exhibits an extended LD stretching nearly 2 Mb.

To better resolve the region of association on chromosome 6, an MLMM was conducted (Figure 2d; Supplemental Table S6). If there are multiple zeaxanthin QTL in this region, we would predict that these models would also identify SNPs at 47 Mb. The MLMM identified the significant SNP inside the *ZEP* gene (S6_46717975) but did not identify any other nearby SNPs, meaning the significant associations with the SNPs at 47 Mb were more likely due to LD with a SNP in the 46.7-Mb region rather than an additional causal gene near 47 Mb. The MLMM also identified a significant association with an SNP on chromosome 3 near the *DXR* homolog (Sobic.003G103300) identified in the MLM.

3.3.3 | Lutein

A GWAS using an MLM (Figure 3a) and an MLMM (Figure 3b) did not identify any SNPs significantly associated with lutein at 5 or 10% FDR.

3.3.4 | Zeaxanthin/lutein ratios

Loci controlling the ratios between carotenoids in the α-branch and β-branch of the pathway could be used to shunt the pathway toward the higher provitamin A carotenoid β-branch. Therefore, we conducted a GWAS with MLM using the ratio of zeaxanthin to lutein (Figure 3c; Supplemental Table S7). An SNP on chromosome 6 at 47.2 Mb was the only significant SNP at 5% FDR, and the SNP at 46.7 Mb inside the *ZEP* gene was significant at 10% FDR.

To better resolve the region on chromosome 6, an MLMM was conducted (Figure 3d; Supplemental Table S8). In contrast to the zeaxanthin MLMM, only the SNP at 47.2 Mb was significant. The SNP on chromosome 6 at 46.7 Mb was identified as one of the top SNPs in the MLMM, although it did not reach the significance threshold.

3.3.5 | Biosynthesis pathway-targeted genome-wide association studies

Another way to control for possible false negatives is to reduce the multiple comparisons by conducting a GWAS using only SNPs in proximity to a priori candidate genes (Lipka et al., 2014; Owens et al., 2014). We conducted a pathway-level analysis using 62,123 SNPs within ±200 kb of a priori candidate genes. Zeaxanthin was the only trait with significant SNP associations (Supplemental Table S9). One was the SNP within the *ZEP* gene (Sobic.006G097500), and the other was the SNP near the *MDS* gene (Sobic.004G281900); both identified in the zeaxanthin GWAS using MLM.

4 | DISCUSSION

4.1 | Characterization of carotenoids in sorghum

Given the high incidence of vitamin A deficiencies in many regions of the world where sorghum is a staple crop, our primary research goal is to develop sorghum varieties with biologically relevant concentrations of provitamin A carotenoids.

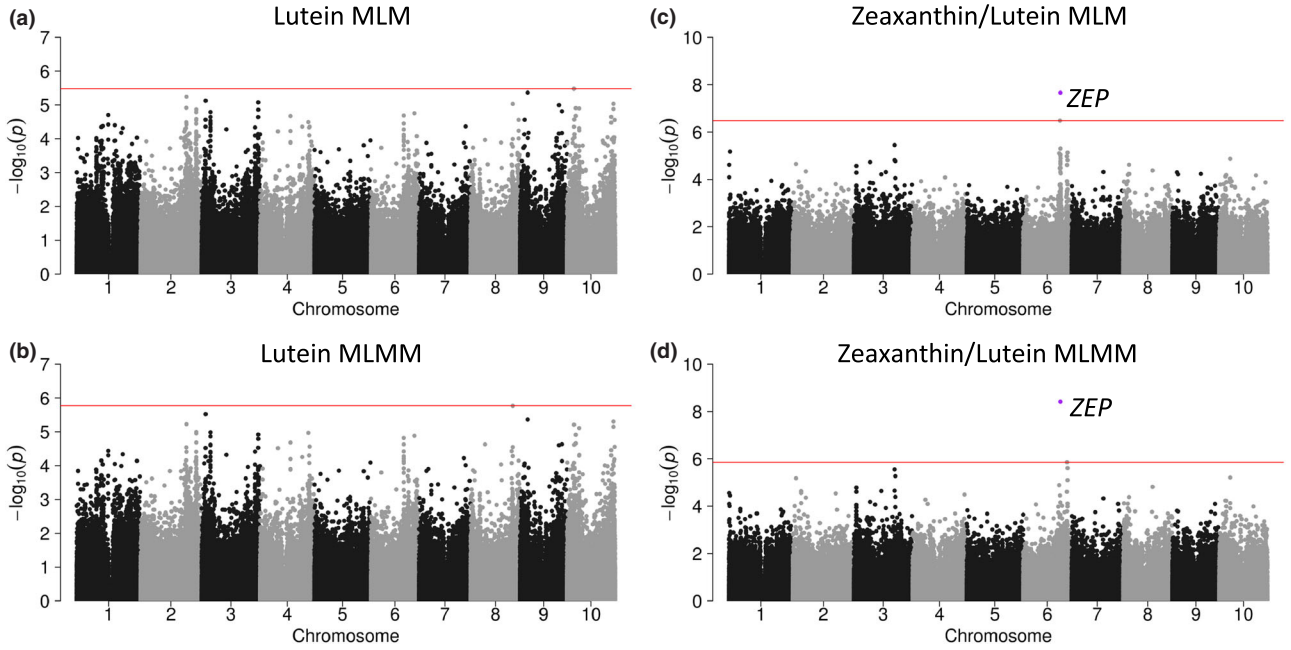

FIGURE 3 Genome-wide association study for carotenoid concentrations in sorghum grain. Manhattan plot of association results for (a) lutein mixed linear model (MLM), (b) lutein multi-locus mixed model (MLMM), (c) zeaxanthin/lutein ratio MLM, and (d) zeaxanthin/lutein ratio MLMM. Axes: The $-\log_{10} p$ values (y-axis) plotted against the position on each chromosome (x-axis). Each circle represents a single nucleotide polymorphism. The horizontal line represents the genome-wide significance threshold at 5% false discovery rate

Of the three provitamin A carotenoids expected to be present in sorghum grain, β-carotene was measured in the highest concentrations. β-Cryptoxanthin was found in low concentrations in some of the samples, whereas α-carotene was not detectable in any of the samples. This is in contrast to maize grain, in which both α-carotene and β-cryptoxanthin have been identified in higher concentrations. However, the pattern in sorghum grain is similar to maize, which has high concentrations of zeaxanthin and lutein and much lower concentrations of β-carotene, β-cryptoxanthin, and particularly α-carotene (Owens et al., 2014). We predict that, as we expand the amount of germplasm that we phenotype, we will identify sorghum varieties with higher concentrations of these carotenoids. Another consideration is that the whole grain samples were stored in the freezer for ∼1 yr before analysis, so there was likely some degradation of carotenoids. Experiments are being conducted to determine degradation of sorghum grain carotenoids over time in multiple storage conditions.

Although additional phenotypic data from subsequent years and environments, as well as different populations, will help to validate significant associations identified in GWAS, HPLC is expensive and time consuming and requires specialized training. High-throughput methods of quantification, such as near-infrared spectroscopy or ultraviolet spectroscopy, and the use of MAS to reduce reliance on phenotypic selection could substantially decrease the time and cost of carotenoid biofortification. This would be particularly useful for sorghum breeders in middle- to low-income countries. We are currently developing high-throughput quantification methods for sorghum grain carotenoids.

4.2 | Genetic controls of sorghum carotenoids

The inheritance of carotenoids in food crops has received intense focus for nearly a century due to the critical role of carotenoids in human nutrition and health (Buishand & Gabelman, 1979; Giuliano, Pollock, & Scolnik, 1986; Iorizzo et al., 2016; LeRosen, Went, & Zechmeister, 1941; Lichtenthaler, Schwender, Disch, & Rohmer, 1997; Mangelsdorf & Fraps, 1931; Porter & Lincoln, 1950; Ye et al., 2000). The carotenoid biosynthetic pathway (Figure 4) is highly conserved across plant species and has been extensively studied in Arabidopsis and maize. Although the structural genes are well described, the regulatory pathway controls are largely unknown. This is reflected in the many association studies, including ours, in which only a small percentage of the identified loci colocalize with known pathway genes (Azmach, Menkir, Spillane, & Gedil, 2018; Gonzalez-Jorge et al., 2016; Owens et al., 2014; Schulz et al., 2016; Suwarno, Pixley, Palacios-Rojas, Kaeppler, & Babu, 2015). Studies focused on identifying the causal genes underlying significant associations identified with GWAS are needed to improve our understanding of carotenoid pathway controls.

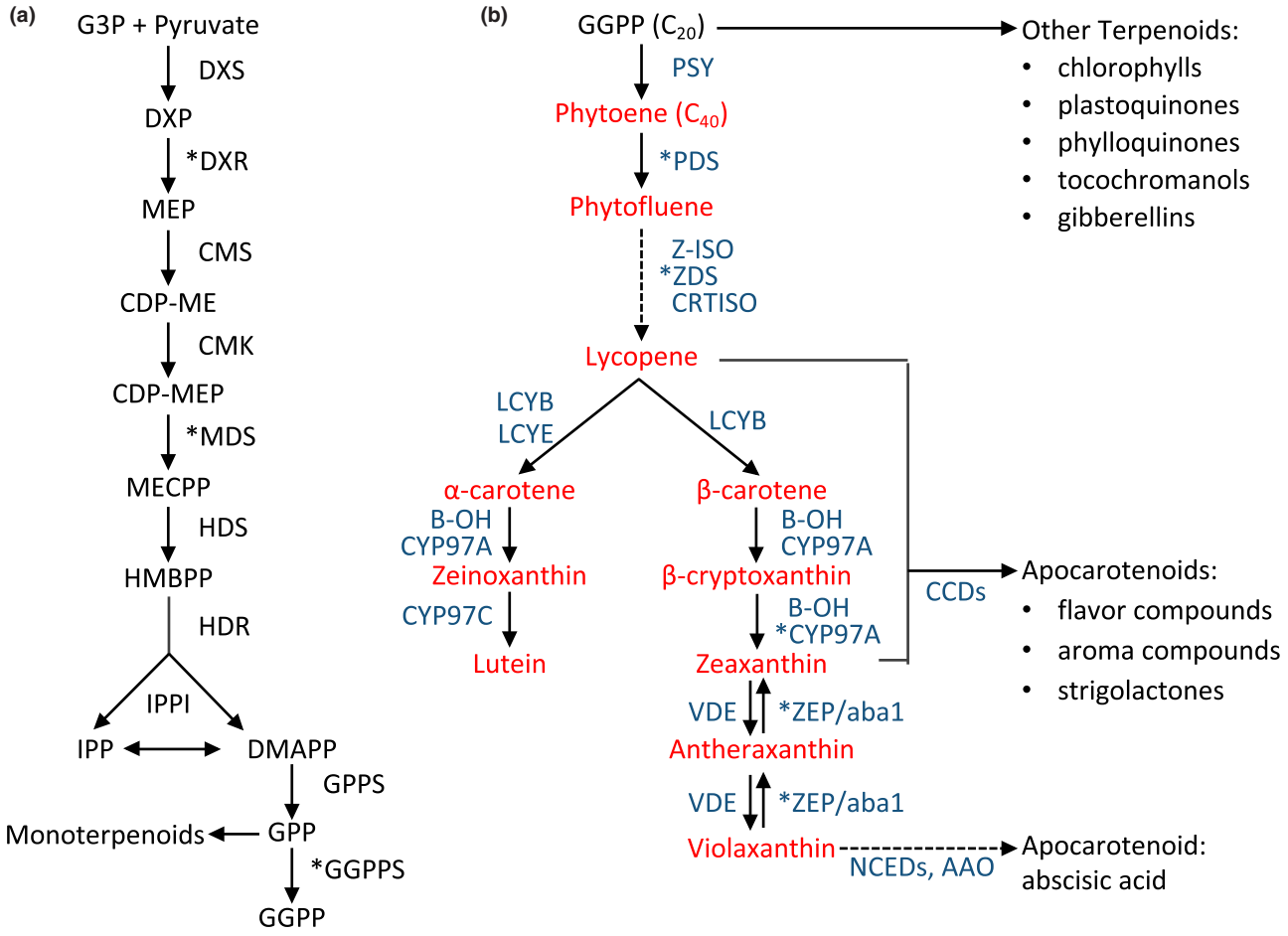

FIGURE 4 Simplified scheme of carotenoid biosynthesis. (a) Methylerythritol phosphate (MEP) pathway, producing carotenoid precursors, and (b) carotenoid pathway, with carotenoids in red and biosynthetic enzymes in blue. Compounds denoted with an asterisk are candidate genes identified in this study. AAO, abscisic aldehyde oxidase; B-OH, β-ring hydroxylase; CCD, carotenoid cleavage dioxygenase; CDP-ME, methylerythritol cytidyl diphosphate; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate; CMS, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CRTISO, carotenoid isomerase; CYP97A, cytochrome P450 97A; CYP97C, cytochrome P450 97C; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-d-zylulose 5-phosphate; DXR, 1-deoxy-d-zylulose 5-phosphate reductoisomerase; DXS, 1-deoxy-d-zylulose 5-phosphate synthase; G3P, glyceraldehyde 3-phosphate; GGPP, geranylgeranyl diphosphate; GGPPS, geranylgeranyl diphosphate synthase; GPP, geranyl diphosphate; GPPS, geranyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HMBPP, 4-hydroxy-3-methyl-butenyl 1-diphosphate; IPP, isopentenyl diphosphate; IPPI, isopentenyl diphosphate isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; MDS, ME-cPP synthase; ME-cPP, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; MEP, methylerythritol phosphate; NCED, 9-cis-epoxycarotenoid dioxygenase; PDS, phytoene desaturase; PSY, phytoene synthase; VDE, violaxanthin de-epoxidase; ZDS, δ-carotene desaturase; ZEP, zeaxanthin epoxidase (also known as aba1, abscisic acid deficient 1); Z-ISO, 15-cis-δ-carotene isomerase

4.2.1 | Zeaxanthin epoxidase

A gene that is highly conserved across species can be hypothesized to have similar functions across species. The sorghum *ZEP* gene has 95% similarity to the maize ortholog (GRMZM2G127139) and 75% similarity to the Arabidopsis ortholog (At5g67030). The SNP identified inside the putative sorghum *ZEP* gene (Sobic.006G097500) had the strongest marker association in this study and is particularly interesting because *ZEP* has been identified not only in a grain carotenoid GWAS for sorghum's close monocot relative maize (Owens et al., 2014; Suwarno et al., 2015) but also in the much more distantly related dicot Arabidopsis (Gonzalez-Jorge et al., 2016). This conservation across species is not surprising given the essential function of carotenoids in photosynthesis and production of key plant hormones such as abscisic acid and strigolactone. Identifying the underlying mechanism for the role of ZEP in carotenoid variation can help inform molecular breeding approaches. In Arabidopsis, differential expression of *ZEP* during seed development was found to be responsible for variation in carotenoid content, and a *ZEP* homozygous null mutant had increased zeaxanthin, β-carotene, and lutein (Gonzalez-Jorge et al., 2016). The authors of the study hypothesized that pleiotropy, rather than a direct

effect, is responsible for the increased β-carotene and lutein concentrations in the absence of *ZEP* expression, speculating that the loss of ZEP activity leads to abscisic acid deficiency and drastically decreased expression of *CCD4*, a major carotenoid degradation enzyme. Gene expression experiments will help determine if there are similar mechanisms in sorghum. To further assess the potential of the *ZEP* gene in sorghum carotenoid biofortification, Kompetitive Allele Specific PCR (KASP) markers are being developed to use for MAS.

The extent of LD in a genomic region affects the mapping resolution and contributes to interpretation of regions of association identified in a GWAS. The distance between the *ZEP* gene and the 24 SNPs identified in the 46–47 Mb region on chromosome 6 ranged from 0 to 1244 kb. According to Morris et al. (2013), the average LD decay on chromosome 6 in the diversity panel used in this study is 500–600 kb, so it is not clear if the significant SNPs identified around the 47 Mb region are in LD with a separate gene than the SNP identified at 46.7 Mb inside the *ZEP* gene. The MLMM for zeaxanthin only identified the SNP at 46.7 Mb, but the MLMM for the zeaxanthin/lutein ratio only identified a SNP at 47.2 Mb. However, the SNP at 46.7 Mb was near the significance threshold in the MLMM for the zeaxanthin/lutein ratio. A potential second candidate is a gene 97 kb from the SNP at 47.2 Mb (Sobic.006G103400). This gene is a putative homolog of the Arabidopsis *orange* (*OR*) gene (AT5G61670) and a putative ortholog of the Arabidopsis *low quantum yield of photosystem II1* (*LQY1*) gene (AT1G75690), both of which code for DnaJ-like zinc finger domain proteins known to modulate carotenoids through plastid regulation (Wang et al., 2016; Yuan, Zhang, Nageswaran, & Li, 2015). Further analysis with a biparental mapping population or nested association mapping population could help to resolve this region.

4.2.2 | Methylerythritol phosphate pathway bottlenecks

Carotenoids are terpenoid compounds derived from the plastid-localized MEP pathway (Figure 4a). The MEP pathway produces the five-carbon building blocks isopentyl diphosphate and dimethylallyl diphosphate that are needed for the synthesis of carotenoids as well as synthesis of several other important terpenoid compounds, such as chlorophylls, tocochromanols, and gibberellins. Geranylgeranyl diphosphate, the final product of the MEP pathway, is the precursor for all of these terpenoid compounds, and substrate competition may limit carotenoid biosynthesis. We identified SNPs near three MEP pathway genes: *MDS*, *DXR*, and *GGPPS*. Availability of carotenoid pathway precursors represents bottlenecks for flux through the pathway, so it is not surprising that loci near MEP pathway genes were identified by GWAS. The *MDS* gene had not been identified by GWAS before this study. The *DXR* gene, previously identified in an Arabidopsis carotenoid GWAS (Gonzalez-Jorge et al., 2016), is considered a key rate-limiting enzyme in the MEP pathway, and its overexpression leads to increases in carotenoids (Li, Nian, Xian, & Zhang, 2018; Mahmoud & Croteau, 2001; Matthews & Wurtzel, 2000; Yuan, Rouvière, LaRossa, & Suh, 2006). Furthermore, GGPPS, also identified in the Arabidopsis GWAS (Gonzalez-Jorge et al., 2016), functions at the branch point in the MEP pathway where several terpenoid pathways are competing for substrate. Interestingly, in a previous sorghum carotenoid QTL study in a biparental population, *PSY3* was hypothesized to be the causal gene within a QTL region identified on chromosome 2 between ∼65 and 75 Mb (Fernandez et al., 2008). However, *GGPPS* is also within this QTL region, and SNPs near *GGPPS* were identified in our GWAS but not near the *PSY3* gene. This raises the possibility that the QTL identified in the Fernandez paper was actually in LD with *GGPPS* rather than *PSY3*. Also of interest, the SNP identified in our GWAS, which is 98 kb from *GGPPS*, is only 13 kb from a putative *homogentisate solanesyl transferase* (Sobic.002G351600) gene, with 79% similarity to the gene recently identified as underlying the maize white seedling 3 mutant that significantly reduces colored carotenoids (Hunter et al., 2018). Analysis of a biparental mapping population may help to clarify the causal gene in this region.

4.2.3 | Lycopene ε-cyclase

The carotenoid pathway bifurcates into two branches after lycopene synthesis (Figure 4b). Lycopene ε-cyclase (LYCE) and lycopene β-cyclase (LYCB) catalyze the cyclization of lycopene, leading to α-carotene with a β- and ε-ionone ring and β-carotene with two β-ionone rings. The *LYCE* gene has been identified by GWAS studies in Arabidopsis and maize. Harjes et al. (2008) used association analysis, linkage mapping, expression analysis, and mutagenesis to show that LYCE plays a key role in controlling the ratio between the α- and β-carotene branches of the maize carotenoid pathway and is highly associated with the α/β ratio, as well as β-carotene and β-cryptoxanthin. A maize GWAS identified an association between *LYCE* and ratios of α- and β-carotene branch carotenoids, as well as lutein and zeaxanthin (Owens et al., 2014), and an Arabidopsis GWAS identified an association between *LYCE* and ratios of lutein and zeaxanthin (Gonzalez-Jorge et al., 2016). Although it did not reach the significance threshold, one of the top associations identified by the zeaxanthin/lutein ratio GWAS using an MLM and MLMM was with an SNP 150 kb from a putative *LYCE* homolog (Sobic.003G197500), with 94% similarity to the maize *LYCE*

gene (GRMZM2G012966). Given the importance of this gene in other species, we were surprised that it did not reach the significance threshold in our GWAS analysis, but a larger sample size or biparental mapping population may help to resolve this region and clarify the importance of *LYCE* in sorghum carotenoid variation.

4.2.4 | β-Carotene hydroxylase

Hydroxylation of the β-rings on α- and β-carotene is catalyzed by β-carotene hydroxylase (β-OH; Figure 4b), leading to the formation of lutein and zeaxanthin. The importance of β-*OH* in carotenoid variation was first identified through Arabidopsis mutants (Tian, Magallanes-Lundback, Musetti, & DellaPenna, 2003) and later through identification of variants in a maize ortholog associated with carotenoid variation (Yan et al., 2010). The gene has subsequently also been identified by GWAS in maize (GRMZM2G152135) (Owens et al., 2014; Suwarno et al., 2015) and Arabidopsis (Gonzalez-Jorge et al., 2016). Given the importance of β-*OH* in maize and Arabidopsis carotenoid variation (Gonzalez-Jorge et al., 2016; Owens et al., 2014; Suwarno et al., 2015; Venado et al., 2017; Yan et al., 2010), we predicted this gene would also be important in sorghum. However, GWAS did not identify any SNPs near the sorghum β-*OH* gene (Sobic.006G188200). It may be that there are allelic variants of the sorghum β-*OH* that control variation in sorghum carotenoids but that the allele frequency is too low in our population for detection. To get a sense of the effect genetic variants may have on carotenoid concentration, we compared concentrations between six alleles within the sorghum β-*OH* gene in the enhanced sorghum association panel. Depending on the SNP, there was a 1.2-fold up to 2.8-fold difference in carotenoid concentrations between the high and low alleles. A larger sample size may help to clarify the importance of this gene in sorghum.

An alternative hypothesis is that variation in the sorghum β-*OH* gene is constrained due to low gene copy number. A search for homologs in Phytozome identifies only one copy of the gene in sorghum, whereas two homologs in Arabidopsis and eight homologs in maize are identified. A single copy gene in an essential pathway is more constrained than duplicated genes because a change or loss in function of a single copy gene could have deleterious consequences if there is not a duplicate gene with redundant function. This constraint could be the reason there is not greater variation of the sorghum β-*OH* gene in the sorghum association panel.

4.2.5 | Post hoc gene candidates

Genome-wide association studies identified several SNP associations that were not near any a priori gene candidates, but we can speculate on a few interesting genes that were identified after conducting GWAS. The most significant SNP association with β-carotene identified by MLM was on chromosome 6 at 58.2 Mb. There were no a priori candidates in this region, but a putative *farnesyl diphosphate synthase* gene (*FPPS*; Sobic.006G248000) was 534 kb away. In maize endosperm, FPPS has the ability to produce both FPP and GGPP and can influence the ratio of FPP (not a carotenoid precursor) to GGPP (a carotenoid precursor) (Cervantes-Cervantes, Gallagher, Zhu, & Wurtzel, 2006). Interestingly, this *FPPS* homolog only showed expression in the seed, with the highest expression in the embryo (FPKM = 36). There was also a significant association on chromosome 10 at 8.9 Mb that was not near any a priori candidates but was 290 kb from a putative *golden2-like* MYB transcription factor (*GLK2*; Sobic.010G096300). The overexpression of *GLK2*, a regulator of chloroplast differentiation, leads to increased concentrations of carotenoids in ripe tomato (*Solanum lycopersicum* L.) (Nguyen et al., 2014; Powell et al., 2012). Genome-wide association studies identified another SNP on chromosome 7 at 59.0 Mb that was not near any a priori candidate genes but was 40 kb away from *abscisic acid 8′-hydroxylase 2* (*CYP707A*; Sobic.007G156300), which catalyzes the first step in abscisic acid catabolism and is known to be important in seed dormancy (Vallabhaneni & Wurtzel, 2010). Expression data indicate this gene has little to no expression in all tissue except the seed, in which it is highly expressed (97 FPKM in Day 10 seed and 407 FPKM in the endosperm). Further studies are needed to identify the role of these genes in modulating cereal grain carotenoid content.

5 | CONCLUSIONS

We hypothesized that sorghum carotenoids are oligogenic traits and that the genes underlying their variation are homologous to genes controlling carotenoid variation in other species. Our results support these hypotheses and point to new avenues of research. Genome-wide association studies identified up to 12 regions of association depending on the carotenoid trait and model used for analysis, consistent with oligogenic traits. Two regions identified by GWAS were near *MDS* and *PDS*, carotenoid genes that have not been previously identified in carotenoid linkage or association studies. The other regions were in or near carotenoid pathway genes identified in other species, including *ZEP*, *DXR*, *GGPPS*, and *ZDS*. Strong evidence from prior studies in maize and Arabidopsis suggest that the *ZEP* gene identified in this study is the causal gene underlying variation in sorghum zeaxanthin and zeaxanthin/lutein ratio. Recombinant inbred lines are being developed to confirm that the markers identified in this GWAS underlie carotenoid variation in sorghum grain, and breeding experiments are underway to test the hypothesis

that MAS will accelerate sorghum carotenoid biofortification. The genetic markers identified in this study can potentially be used to increase provitamin A carotenoids in locally adapted sorghum varieties in regions of the world with high rates of vitamin A deficiency.

DATA AVAILABILITY

All data are publicly available. The SNP dataset is available in the Dryad Data Repository (doi:10.5061/dryad.63h8fd4).

AUTHOR CONTRIBUTIONS

D.H.R. designed the experiments, conducted GWAS, and wrote the manuscript. C.C.B. performed the pathway-level analysis. S.C., B.I., and S.R.B. designed the methods for and conducted the HPLC analysis. R.P. planted the multi-environment germplasm. D.H.R. conducted genotyping-by-sequencing, and Z.H. conducted sequence alignment and imputation. C.C.B., S.R.B., and D.H.R. contributed to data an and interpretation. All authors reviewed and edited the manuscript.

ACKNOWLEDGMENTS

This work was supported by USDA NIFA AFRI Postdoctoral Fellowship grant no. 2017-67012-26098 from the USDA National Institute of Food and Agriculture. This is publication 20-069-J of the Kansas Agricultural Experiment Station. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable. USDA is an equal opportunity provider and employer.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Davina H. Rhodes
https://orcid.org/0000-0003-2140-1519

REFERENCES

Africa Harvest Biotech Foundation International. (2010). *Africa Biofortified Sorghum Project: Five-year progress report*. Nairobi: Africa Harvest Biotech Foundation International.

Africa Harvest Biotech Foundation International. (2012). *Africa harvest strategic plan 2012–2022*. Nairobi: Africa Harvest Biotech Foundation International.

Arvidsson, T. (2015). *Food security now or wait for research to assess risks?* Uppsala, Sweden: Nordic Africa Institute.

Azadi, H., Samiee, A., Mahmoudi, H., Jouzi, Z., Rafiaani Khachak, P., De Maeyer, P., & Witlox, F. (2015). Genetically modified crops and small-scale farmers: Main opportunities and challenges. *Critical Reviews in Biotechnology*, *36*(3). https://doi.org/10.3109/07388551.2014.990413

Azmach, G., Menkir, A., Spillane, C., & Gedil, M. (2018). Genetic loci controlling carotenoid biosynthesis in diverse tropical maize lines. *G3: Genes, Genomes, Genetics*, *8*, 1049–1065. https://doi.org/10.1534/g3.117.300511

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Blessin, C. W., Vanetten, C. H., & Wiebe, R. (1958). Carotenoid content of the grain from yellow endosperm-type sorghums. *Cereal Chemistry*, *35*, 359–365.

Buishand, J. G., & Gabelman, W. H. (1979). Investigations on the inheritance of color and carotenoid content in phloem and xylem of carrot roots (*Daucus carota* L.). *Euphytica*, *28*, 611–632. https://doi.org/10.1007/BF00038928

Cardoso, L. de M., Pinheiro, S. S., da Silva, L. L., de Menezes, C. B., de Carvalho, C. W. P., Tardin, F. D., … Pinheiro-Sant'Ana, H. M. (2015). Tocochromanols and carotenoids in sorghum (*Sorghum bicolor* L.): Diversity and stability to the heat treatment. *Food Chemistry*, *172*, 900–908. https://doi.org/10.1016/j.foodchem.2014.09.117

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., … Kresovich, S. (2008). Community resources and strategies for association mapping in sorghum. *Crop Science*, *48*, 30–40. https://doi.org/10.2135/cropsci2007.02.0080

Cazzonelli, C. I., & Pogson, B. J. (2010). Source to sink: Regulation of carotenoid biosynthesis in plants. *Trends in Plant Science*, *15*, 266–274. https://doi.org/10.1016/j.tplants.2010.02.003

Cervantes-Cervantes, M., Gallagher, C. E., Zhu, C., & Wurtzel, E. T. (2006). Maize cDNAs expressed in endosperm encode functional farnesyl diphosphate synthase with geranylgeranyl diphosphate synthase activity. *Plant Physiology*, *141*, 220–231. https://doi.org/10.1104/pp.106.077008

Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.-H., … Robin Buell, C. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *The Plant Journal*, *71*, 492–502. https://doi.org/10.1111/j.1365-313X.2012.05005.x

Diepenbrock, C. H., & Gore, M. A. (2015). Closing the divide between human nutrition and plant breeding. *Crop Science*, *55*(4), 1437–1448. https://doi.org/10.2135/cropsci2014.08.0555

Eggersdorfer, M., & Wyss, A. (2018). Carotenoids in human nutrition and health. *Archives of Biochemistry and Biophysics*, *652*, 18–26. https://doi.org/10.1016/j.abb.2018.06.001

Esteban, R., Moran, J. F., Becerril, J. M., & García-Plazaola, J. I. (2015). Versatility of carotenoids: An integrated view on diversity, evolution, functional roles and environmental interactions. *Environmental and Experimental Botany*, *119*, 63–75. https://doi.org/10.1016/j.envexpbot.2015.04.009

FAO. (1995). Sorghum and millets in human nutrition. Retrieved from http://www.fao.org/docrep/T0818E/T0818E04.htm

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., & Kresovich, S. (2008). Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. *Crop Science*, *48*, 1732. https://doi.org/10.2135/cropsci2007.12.0684

Fernandez, M. G. S., Kapran, I., Souley, S., Abdou, M., Maiga, I. H., Acharya, C. B., … Kresovich, S. (2009). Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. *Genetic Resources and Crop Evolution*, *56*, 991–1000. https://doi.org/10.1007/s10722-009-9417-3

Giuliano, G., Pollock, D., & Scolnik, P. A. (1986). The gene crtI mediates the conversion of phytoene into colored carotenoids in *Rhodopseudomonas capsulata*. *Journal of Biological Chemistry*, *261*, 12925–12929.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLOS ONE*, *9*, e90346. https://doi.org/10.1371/journal.pone.0090346

Gonzalez-Jorge, S., Mehrshahi, P., Magallanes-Lundback, M., Lipka, A. E., Angelovici, R., Gore, M. A., & DellaPenna, D. (2016). ZEAXANTHIN EPOXIDASE activity potentiates carotenoid degradation in maturing seed. *Plant Physiology*, *171*, 1837–1851. https://doi.org/10.1104/pp.16.00604

Gorbet, D. W. (1971). *Inheritance of some endosperm types in sorghum (Sorghum bicolor (L.) Moench* (Doctoral dissertation, Oklahoma State University). Retrieved from https://shareok.org/bitstream/handle/11244/27496/Thesis-1971D-G661I.pdf?sequence=1

Gupta, P., Sreelakshmi, Y., & Sharma, R. (2015). A rapid and sensitive method for determination of carotenoids in plant tissues by high performance liquid chromatography. *Plant Methods*, *11*, 5. https://doi.org/10.1186/s13007-015-0051-0

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., … Buckler, E. S. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, *319*, 330–333. https://doi.org/10.1126/science.1150255

Hu, Z., Olatoye, M. O., Marla, S., & Morris, G. P. (2019). An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. *Plant Genomes*, *12*(1). https://doi.org/10.3835/plantgenome2018.06.0044

Hunter, C. T., Saunders, J. W., Magallanes-Lundback, M., Christensen, S. A., Willett, D., Stinard, P. S., … Koch, K. E. (2018). Maize w3 disrupts homogentisate solanesyl transferase (ZmHst) and reveals a plastoquinone-9 independent path for phytoene desaturation and tocopherol accumulation in kernels. *The Plant Journal*, *93*, 799–813. https://doi.org/10.1111/tpj.13821

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., … Simon, P. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature Genetics*, *48*, 657–666. https://doi.org/10.1038/ng.3565

Irakli, Maria N., Samanidou, Victoria F., & Papadoyannis, Ioannis N. (2011). Development and validation of an HPLC method for the simultaneous determination of tocopherols, tocotrienols and carotenoids in cereals after solid-phase extraction. *Journal of Separation Science*, *12*, 1375–1382. https://doi.org/10.1002/jssc.201100077.

Kaulmann, A., & Bohn, T. (2014). Carotenoids, inflammation, and oxidative stress: Implications of cellular signaling pathways and relation to chronic disease prevention. *Nutrition Research*, *34*, 907–929. https://doi.org/10.1016/j.nutres.2014.07.010

Kean, E. G., Ejeta, G., Hamaker, B. R., & Ferruzzi, M. G. (2007). Characterization of carotenoid pigments in mature and developing kernels of selected yellow-endosperm sorghum varieties. *Journal of Agricultural and Food Chemistry*, *55*, 2619–2626. https://doi.org/10.1021/jf062939v

Kulczyński, B., Gramza-Michałowska, A., Kobus-Cisowska, J., & Kmiecik, D. (2017). The role of carotenoids in the prevention and treatment of cardiovascular disease: Current state of knowledge. *Journal of Functional Foods*, *38*, 45–65. https://doi.org/10.1016/j.jff.2017.09.001

LeRosen, A. L., Went, F. W., & Zechmeister, L. (1941). Relation between genes and carotenoids of the tomato. *Proceedings of the National Academy of Sciences of the United States of America*, *27*, 236–242.

Li, M., Nian, R., Xian, M., & Zhang, H. (2018). Metabolic engineering for the production of isoprene and isopentenol by Escherichia coli. *Applied Microbiology and Biotechnology*, *102*, 7725–7738. https://doi.org/10.1007/s00253-018-9200-5

Lichtenthaler, H. K., Schwender, J., Disch, A., & Rohmer, M. (1997). Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Letters*, *400*, 271–274. https://doi.org/10.1016/S0014-5793(96)01404-4

Lipka, A. E., Gore, M. A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., … DellaPenna, D. (2014). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3; Genes|Genomes|Genetics*, *3*, 1287–1299. https://doi.org/10.1534/g3.113.006148

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*, 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Mahmoud, S. S., & Croteau, R. B. (2001). Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 8915–8920. https://doi.org/10.1073/pnas.141237298

Mangelsdorf, P. C., & Fraps, G. S. (1931). A Direct quantitative relationship between vitamin a in corn and the number of genes for yellow pigmentation. *Science*, *73*, 241–242. https://doi.org/10.1126/science.73.1887.241-b

Matthews, P. D., & Wurtzel, E. T. (2000). Metabolic engineering of carotenoid accumulation in Escherichia coli by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Applied Microbiology and Biotechnology*, *53*, 396–400

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., … Mullet, J. E. (2018). The sorghum bicolor reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant Journal* , *93*, 338–354. https://doi.org/10.1111/tpj.13781

Mellado-Ortega, E., & Hornero-Méndez, D. (2015). Carotenoids in cereals: An ancient resource with present and future applications. *Phytochemistry Reviews*, *14*, 873–890. https://doi.org/10.1007/s11101-015-9423-3

Moise, A. R., Al-Babili, S., & Wurtzel, E. T. (2014). Mechanistic aspects of carotenoid biosynthesis. *Chemical Reviews*, *114*, 164–193. https://doi.org/10.1021/cr400106y

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., … Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 453–458. https://doi.org/10.1073/pnas.1215985110

National Research Council. (1996). *Lost crops of Africa: Volume I: Grains*. Washington, DC: The National Academies Press.

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., & Giovannoni, J. J. (2014). Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, *26*, 585–601. https://doi.org/10.1105/tpc.113.118794

NPGS. (2015). *Germplasm Resources Information Network (GRIN)*. Washington, DC: USDA-ARS. https://doi.org/10.15482/USDA.ADC/1212393

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., … Rocheford, T. R. (2014). A foundation for provitamin A biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. *Genetics*, *198*, 1699–1716. https://doi.org/10.1534/genetics.114.169979

Porter, J. W., & Lincoln, R. E. (1950). I. Lycopersicon selections containing a high content of carotenes and colorless polyenes: II. The mechanism of carotene biosynthesis. *Archives of Biochemistry*, *27*, 390–403.

Powell, A. L. T., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., … Bennett, A. B. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science*, *336*, 1711–1715. https://doi.org/10.1126/science.1222218

Purcell, S., Neale B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Schulz, D. F., Schott, R. T., Voorrips, R. E., Smulders, M. J. M., Linde, M., & Debener, T. (2016). Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Frontiers in Plant Science*, *7*, 1798. https://doi.org/10.3389/fpls.2016.01798

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, *44*, 825–830. https://doi.org/10.1038/ng.2314

Stevens, G. A., Bennett, J. E., Hennocq, Q., Lu, Y., De-Regil, L. M., Rogers, L., … Ezzati, M. (2015). Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: A pooled analysis of population-based surveys. *The Lancet Global Health*, *3*, e528–e536. https://doi.org/10.1016/S2214-109X(15)00039-X

Suryanarayana Rao, K., Rukmini, C., & Mohan, V. (1968). Carotene content of some yellow-endosperm varieties of sorghum. *Indian Journal of Agricultural Sciences*, *38*, 368–372

Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, *128*, 851–864. https://doi.org/10.1007/s00122-015-2475-3

Tanaka, T., Shnimizu, M., & Moriwaki, H. (2012). Cancer chemoprevention by carotenoids. *Molecules*, *17*, 3202–3242. https://doi.org/10.3390/molecules17033202

Tanumihardjo, S. A., Russell, R. M., Stephensen, C. B., Gannon, B. M., Craft, N. E., Haskell, M. J., … Raiten, D. J. (2016). Biomarkers of Nutrition for Development (BOND): Vitamin A review. *Journal of Nutrition*, *146*, 1816S–1848S. https://doi.org/10.3945/jn.115.229708

Tian, Li, Magallanes-Lundback, Maria, Musetti, Valeria, & DellaPenna, Dean (2003). Functional analysis of β- and ε-ring carotenoid hydroxylases in *Arabidopsis*. *The Plant Cell*, *6*, 1320–1332. https://doi.org/10.1105/tpc.011403.

Vallabhaneni, R., & Wurtzel, E. T. (2010). From epoxycarotenoids to ABA: The role of ABA 8′-hydroxylases in drought-stressed maize roots. *Archives of Biochemistry and Biophysics*, *504*, 112–117. https://doi.org/10.1016/j.abb.2010.07.005

Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLOS Biology*, *12*, e1001883. https://doi.org/10.1371/journal.pbio.1001883

Venado, R. E., Owens, B. F., Ortiz, D., Lawson, T., Mateos-Hernandez, M., Ferruzzi, M. G., & Rocheford, T. R. (2017). Genetic analysis of provitamin A carotenoid β-cryptoxanthin concentration and relationship with other carotenoids in maize grain (*Zea mays* L.). *Molecular Breeding*, *37*, 127. https://doi.org/10.1007/s11032-017-0723-8

Wang, Y., Chen, S., Wang, W., Huang, X., Zhou, C., Zhuang, Z., & Lu, S. (2016). The DnaJ-like zinc finger domain protein PSA2 affects light acclimation and chloroplast development in *Arabidopsis thaliana*. *Frontiers in Plant Science*, *7*, 360. https://doi.org/10.3389/fpls.2016.00360

WHO. (2009). *Global prevalence of vitamin A deficiency in populations at risk: 1995–2005*. Geneva, Switzerland: WHO Global Database on Vitamin A Deficiency.

Wu, J., Cho, E., Willett, W. C., Sastry, S. M., & Schaumberg, D. A. (2015). Intakes of lutein, zeaxanthin, and other carotenoids and age-related macular degeneration during 2 decades of prospective follow-up. *JAMA Ophthalmology*, *133*, 1415–1424. https://doi.org/10.1001/jamaophthalmol.2015.3590

Yan, J., Kandianis, C. B., Harjes, C. E., Bai, L., Kim, E.-H., Yang, X., … Rocheford, T. (2010). Rare genetic variation at *Zea mays* crtRB1 increases β-carotene in maize grain. *Nature Genetics*, *42*, 322–327. https://doi.org/10.1038/ng.551

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., & Potrykus, I. (2000). Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, *287*, 303–305. https://doi.org/10.1126/science.287.5451.303

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., … Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*, 203–208. https://doi.org/10.1038/ng1702

Yuan, H., Zhang, J., Nageswaran, D., & Li, L. (2015). Carotenoid metabolism and regulation in horticultural crops. *Horticulture Research*, *2*, 15036. https://doi.org/10.1038/hortres.2015.36

Yuan, L. Z., Rouvière, P. E., LaRossa, R. A., & Suh, W. (2006). Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metabolic Engineering*, *8*, 79–90. https://doi.org/10.1016/j.ymben.2005.08.005

Zhai, S., Xia, X., & He, Z. (2016). Carotenoids in staple cereals: Metabolism, regulation, and genetic manipulation. *Frontiers in Plant Science*, *7*, 1197. https://doi.org/10.3389/fpls.2016.01197

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Cruet-Burgos C, Cox S, Ioerger BP, et al. Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *Plant Genome*. 2020;13:e20013. https://doi.org/10.1002/tpg2.20013