

RESEARCH

Open Access



# Characterization of grain carotenoids in global sorghum germplasm to guide genomics-assisted breeding strategies

Clara Cruet-Burgos<sup>1</sup>, Geoffrey P. Morris<sup>2\*</sup> and Davina H. Rhodes<sup>1</sup>

## Abstract

**Background** Crop biofortification is a successful strategy to ameliorate Vitamin A deficiency. Sorghum is a good candidate for vitamin A biofortification, as it is a staple food in regions with high prevalence of vitamin A deficiency.  $\beta$ -carotene—the main provitamin A carotenoid—is below the target concentration in sorghum grain, therefore biofortification breeding is required. Previous studies found evidence that sorghum carotenoid variation is oligogenic, suggesting that marker-assisted selection can be an appropriate biofortification method. However, we hypothesize that sorghum carotenoids have both oligogenic and polygenic components of variation. Genomics-assisted breeding could accelerate breeding efforts, but there exists knowledge gaps in the genetics underlying carotenoid variation, as well as appropriate germplasm to serve as donors.

**Results** In this study, we characterized carotenoids in 446 accessions from the sorghum association panel and carotenoid panel using high-performance liquid chromatography, finding high carotenoid accessions not previously identified. Genome-wide association studies conducted with 345 accessions, confirmed that *zeaxanthin epoxidase* is a major gene underlying variation for not only zeaxanthin, but also lutein and  $\beta$ -carotene. High carotenoid lines were found to have limited genetic diversity, and originated predominantly from only one country. Potential novel genetic diversity for carotenoid content was identified through genomic predictions in 2,495 accessions of unexplored germplasm. Oligogenic variation of carotenoids was confirmed, as well as evidence for polygenic variation, suggesting both marker-assisted selection and genomic selection can facilitate breeding efforts.

**Conclusions** Sorghum vitamin A biofortification could be beneficial for millions of people who rely on it as a dietary staple. Carotenoid content in sorghum is low, but high heritability suggests that increasing concentrations through breeding is possible. Low genetic diversity among high carotenoid lines might be the main limitation for breeding efforts, therefore further germplasm characterization is needed to assess the feasibility of biofortification breeding. Based on germplasm here evaluated, most countries' germplasm lacks high carotenoid alleles, thus pre-breeding will be needed. A SNP marker within the *zeaxanthin epoxidase* gene was identified as a good candidate for use in marker-assisted selection. Due to the oligogenic and polygenic variation of sorghum grain carotenoids, both marker-assisted selection and genomic selection can be employed to accelerate breeding efforts.

**Keywords** Sorghum, Carotenoid, Biofortification, Vitamin A, Genomics-assisted breeding, GWAS, Genomic predictions

\*Correspondence:

Geoffrey P. Morris

davina.rhodes@colostate.edu

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Carotenoids are the main source of vitamin A in many developing countries where diets are primarily plant based. Cereals provide the majority of calories in developing countries, and most cereal grains accumulate carotenoids—particularly lutein, zeaxanthin,  $\beta$ -carotene, and  $\beta$ -cryptoxanthin [1]. However, the concentration of provitamin A carotenoids— $\beta$ -carotene,  $\beta$ -cryptoxanthin and  $\alpha$ -carotene—is low in cereals compared to fruits, vegetables, and animal derived products. For example, among cereals, maize contains the highest concentrations of carotenoids [1], however, the majority of yellow maize accessions accumulate less than 2  $\mu\text{g/g}$  of provitamin A carotenoids [2], although accessions with higher concentrations have been identified [3]. This concentration is low compared to carotenoid-containing fruits and vegetables, such as carrots (50  $\mu\text{g/g}$  of  $\beta$ -carotene) [4], melon (1  $\mu\text{g/g}$  of  $\beta$ -carotene) [4, 5] and kale (6  $\mu\text{g/g}$  of  $\beta$ -carotene) [6].

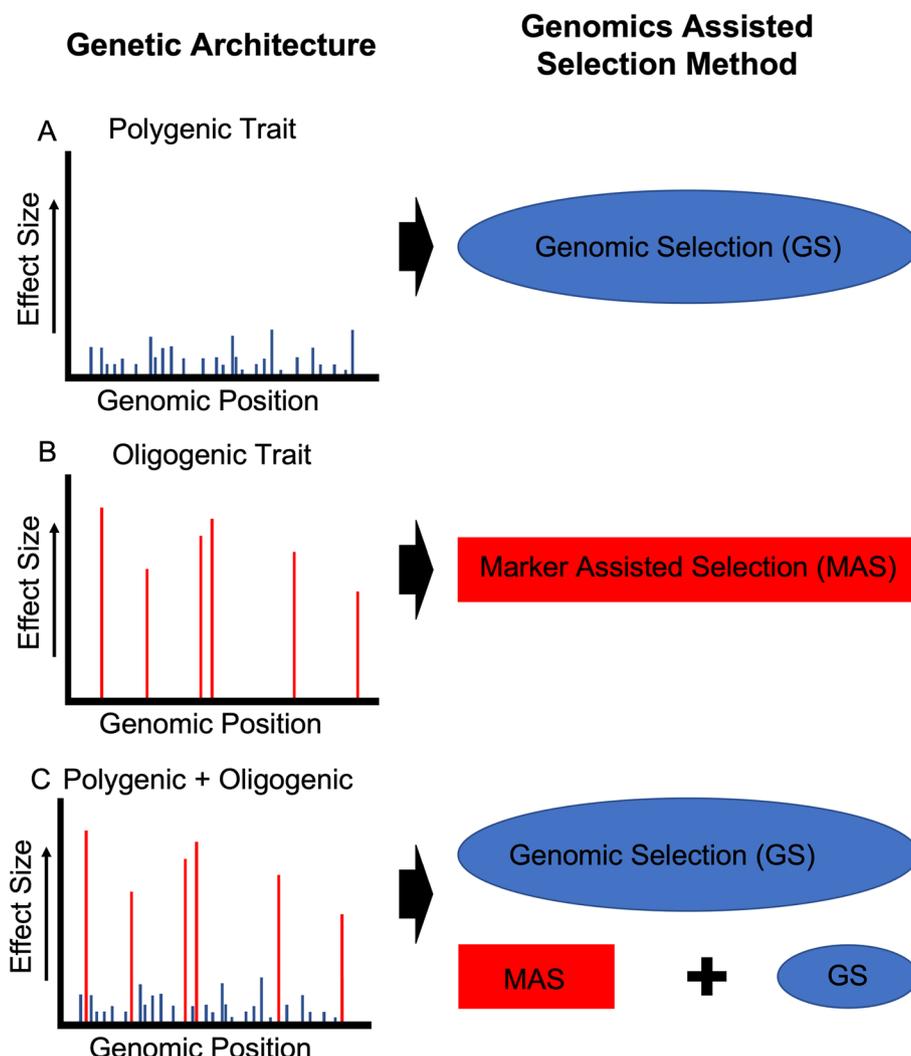
Globally, vitamin A deficiency affects an estimated 190 million preschool age children and 19 million pregnant women, contributing to poor growth, intellectual impairment, vision loss, perinatal complications, and increased mortality [7, 8]. Cereal biofortification is one of the most sustainable strategies to combat vitamin A deficiency in developing countries [9]. HarvestPlus has accomplished successful maize vitamin A biofortification through traditional breeding, with current releases in several countries containing  $\beta$ -carotene ranging from 4–16  $\mu\text{g/g}$  [10]. Given the prevalence of vitamin A deficiency in many developing countries and proven success of current biofortified crops, expanding biofortification efforts to other staple crops could significantly reduce global vitamin A deficiencies.

Sorghum [*Sorghum bicolor* (L.) Moench] is a good candidate for biofortification as it is a staple food in regions with high prevalence of vitamin A deficiency, such as in South East Asia and sub-Saharan Africa [11]. Studies have demonstrated that carotenoids are present in sorghum grains, and  $\beta$ -carotene is the main provitamin A carotenoid, with concentrations up to 3.23  $\mu\text{g/g}$  [12, 13, 13–20]. We estimate a sorghum biofortification target value of approximately 12  $\mu\text{g/g}$   $\beta$ -carotene, although that value will vary depending on the sorghum intake of the target population. Sorghum biofortification using genetic engineering has developed sorghum grain with  $\beta$ -carotene concentrations as high as 12  $\mu\text{g/g}$  [21, 22], but genetically modified sorghum has not been adopted by farmers due to limitations on use of transgenic crops in Africa [23, 24]. However, progress in sorghum carotenoid research suggests that biofortification through breeding is feasible. Genetic studies have demonstrated that there is natural phenotypic variation in sorghum grain

carotenoids and that there are genetic components controlling the trait [15, 16, 20]. Genomics-assisted breeding, via marker-assisted selection (MAS) or genomic selection (GS), has the potential to accelerate biofortification efforts by removing the need to employ complex phenotyping methods. Therefore, due to the potential impact of biofortified sorghum in developing countries, the feasibility of sorghum vitamin A biofortification through breeding needs to be tested.

To use genomics-assisted breeding to develop a carotenoid biofortified sorghum variety, genomic regions associated with variation of provitamin A carotenoids must be identified, as well as efficient selection methods and parental donors. The carotenoid biosynthetic pathway is well understood and conserved in plants [25], facilitating identification of carotenoid genes in sorghum. For  $\beta$ -carotene—the most abundant provitamin A carotenoid in sorghum grain—marker-trait associations have been identified in proximity to *phytoene synthase* (*PSY*), *phytoene desaturase* (*PDS*), and *geranylgeranyl diphosphate synthase* (*GGPS*) genes [15, 16]. For zeaxanthin, marker-trait associations have been identified within *zeaxanthin epoxidase* (*ZEP*), a gene that has also been identified in several other crops as underlying carotenoid variation [26–29]. Previous research in sorghum suggests that carotenoids are oligogenic traits, meaning a moderate number of genes with moderate effect underly the majority of the phenotypic variation detected [15, 16, 20], which is consistent with observations in maize [30] and wheat [31]. The oligogenic variation of carotenoids in cereal grains suggests that MAS may be an effective biofortification method. Alternatively, due to the interconnectedness of carotenoid biosynthesis to other biochemical pathways, a combination of both oligogenic and polygenic models might more accurately explain carotenoid variation. In this instance, GS, or MAS followed by GS, could be employed to simultaneously select for both large and small effect genes (Fig. 1).

Identifying germplasm that harbor alleles for high  $\beta$ -carotene concentrations in sorghum grains is also essential for biofortification breeding. This aspect is perhaps the most challenging, because even though carotenoids are naturally present in sorghum grains, studies have shown that the majority of sorghum varieties have low  $\beta$ -carotene concentrations [14–16, 18, 32]. The limited number of high  $\beta$ -carotene varieties could imply that there is limited genetic diversity for carotenoid concentrations, which would impede gains in breeding efforts. However, the accessions that have been phenotyped for  $\beta$ -carotene concentration is limited, suggesting there could be unexplored germplasm harboring genetic variation. No direct assessment of diversity in high carotenoid sorghum accessions has been conducted, but the



**Fig. 1** Genomics-assisted selection method based on genetic architecture of a trait. Genomics-assisted selection method for a trait under **A)** polygenic; **B)** oligogenic; and **C)** both polygenic and oligogenic control. Effect size is represented by height of vertical lines; number of genes associated with the trait is represented by number of vertical lines; blue vertical lines represent the polygenic component and red vertical lines represent the oligogenic component contributing to the trait

high incidence of yellow endosperm from Nigeria in collections and previous studies [16, 33–35] supports that limited diversity is a possibility. Expanding germplasm evaluations and genetic studies could therefore highlight new or conserved genomic regions associated with  $\beta$ -carotene that can be used for MAS, as well as identify a set of parental lines with sufficient genetic diversity to employ in breeding efforts.

Given the potential impact of a high vitamin A sorghum developed through biofortification breeding, and the current gaps in knowledge for sorghum biofortification, in this study we further explored the potential of genomics-enabled breeding tools for increasing carotenoids in sorghum. We hypothesize that both oligogenic

and polygenic components of variation exist for sorghum carotenoids (Fig. 1), such that both MAS and GS could accelerate breeding efforts. We expanded the number of sorghum accessions phenotyped for carotenoids, identifying additional high carotenoid accessions, and found evidence for both an oligogenic and polygenic component of variation in sorghum grain carotenoids. We also found that the limited number of known high carotenoid accessions have low genetic diversity among them, but genomic predictions identified new potential donor lines that could harbor novel genetic variation for carotenoids. Lastly, we examined allelic diversity in the *ZEP* gene and found evidence of selection for a high carotenoid allele found in only a few countries.

## Results

### Phenotypic variation of carotenoids in the SAP/CAP collection

To characterize phenotypic variation of carotenoids in sorghum grain, and to confirm previously published phenotype data on one year of samples, we quantified lutein, zeaxanthin,  $\beta$ -carotene, and  $\alpha$ -carotene for 446 accessions in the sorghum association and carotenoid panel (SAP/CAP) global collection using high-performance liquid chromatography HPLC (Table S1). Lutein was the most abundant carotenoid, followed by zeaxanthin,  $\beta$ -carotene, and then  $\alpha$ -carotene. Raw concentrations for lutein ranged from 0.02–4.61  $\mu\text{g/g}$ , for zeaxanthin from 0.01–2.40  $\mu\text{g/g}$  and for  $\beta$ -carotene from 0.03–1.19  $\mu\text{g/g}$ , with means of 0.58  $\mu\text{g/g}$ , 0.18  $\mu\text{g/g}$ , and 0.17  $\mu\text{g/g}$ , respectively.  $\alpha$ -carotene was detected in only 31 accessions, with values ranging from 0.02–0.11  $\mu\text{g/g}$ . Due to the limited number of accessions with detectable concentrations,  $\alpha$ -carotene was omitted from subsequent genetic analysis. High phenotypic correlations were found between  $\beta$ -carotene and zeaxanthin ( $r=0.74$ ;  $p < 10^{-16}$ ),  $\beta$ -carotene and lutein ( $r=0.78$ ;  $p < 10^{-16}$ ), and lutein and zeaxanthin ( $r=0.75$ ;  $p < 10^{-16}$ ). Four accessions, two of which had not previously been phenotyped, had higher concentrations of  $\beta$ -carotene than any accessions previously phenotyped in the SAP/CAP collection.

To account for unbalanced data and accurately predict the genetic merit for carotenoids of the SAP/CAP accessions, best linear unbiased predictors (BLUPs) and heritabilities were calculated for each of the carotenoid traits (Table 1, Table S2). Due to the expected shrinkage effect, lower ranges were obtained for the BLUPs than for the raw concentrations. However, entry-mean basis heritability estimates (Table 1) were high, ranging from 0.78 for  $\beta$ -carotene to 0.92 for zeaxanthin.

### Genome-wide association study of carotenoids in SAP/CAP collection

Next, we sought to characterize the genetic architecture of sorghum carotenoids. A previous study [15] suggested that global sorghum grain carotenoid

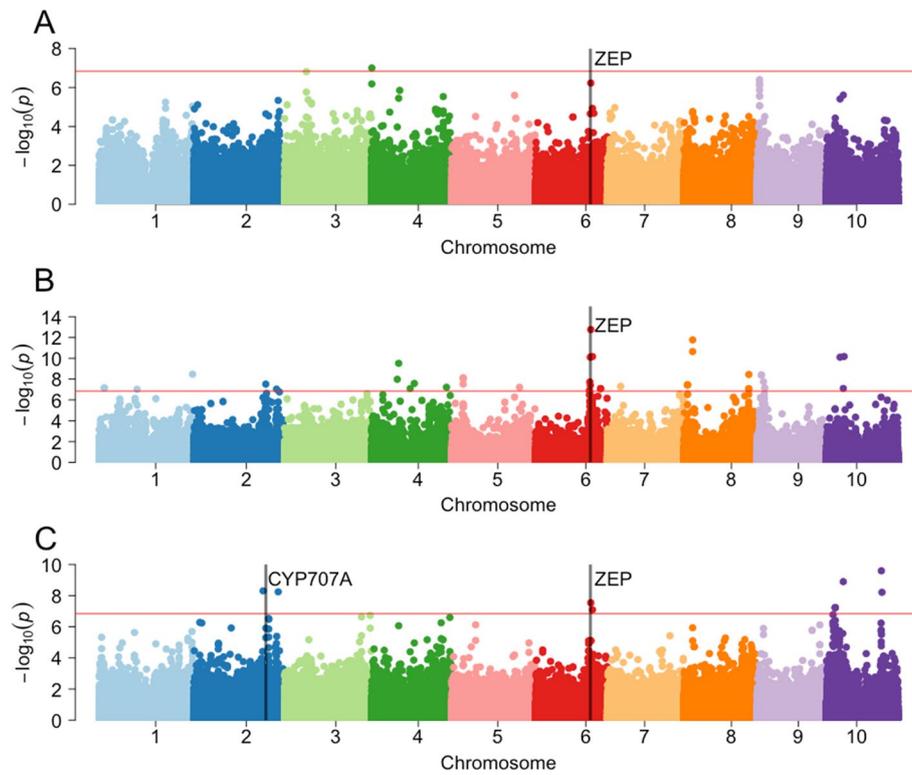
variation is oligogenic, so to further test this hypothesis, we conducted a genome-wide association study (GWAS) using more accessions and replicates. To maximize the number of accessions included, we used BLUPs rather than raw data in order to account for unbalanced data. Marker-trait associations were identified for the BLUPs of the three carotenoid traits evaluated (Fig. 2, Table S3). GWAS was conducted on 345 accessions from the SAP/CAP collection for which we had phenotype and genotype information.

For lutein, only 1 SNP, on chromosome 4 (S04\_275231), was above the Bonferroni threshold of significance (Fig. 2A, Table S3). To identify candidates that may not be found using the stricter Bonferroni multiple comparison corrections, we also considered a more liberal False Discovery Rate (FDR) criteria. Under the  $\text{FDR} < 0.05$  threshold, 7 significant single nucleotide polymorphisms (SNPs) were identified, corresponding to four regions of association on chromosomes 3, 4, 6 and 9. Three of these SNPs were located in a region around 2.17 Mb on chromosome 9, which is not near any a priori candidate genes. The only association in proximity to an a priori candidate gene was at SNP S6\_47123508, near Sobic.006G097500 (401 kb away), an a priori candidate gene that is annotated as a putative ortholog of the maize *ZEP* gene.

Zeaxanthin had the highest number of marker-trait associations above the Bonferroni significance threshold, with 39 significant SNPs in 17 regions across all chromosomes except chromosome 3 (Fig. 2B, Table S3). The most prominent association was on chromosome 6 between 45.9–48.6 Mb, with six significant SNPs, three of which were among the top ten most significant associations. The most significant association for zeaxanthin was the SNP near the *ZEP* gene (S6\_47123508; 401 kb away) that was also associated with lutein. There was also an association on chromosome 2 (S2\_61694864), which is in proximity to Sobic.002G225400 (42 kb away), an a priori candidate gene annotated as an *abscisic acid 8'-hydroxylase 3* (*CYP707A*).  $\beta$ -carotene had ten significant marker-trait associations for a total of six regions of association across chromosomes 2, 6 and 10 (Fig. 2C, Table S3). Chromosome 10 had the highest number of marker-trait associations, particularly within a region around 7.48 Mb. There was also a SNP on chromosome 10 (S10\_14377366) that was significantly associated with both  $\beta$ -carotene and zeaxanthin, which is not in proximity to any a priori candidate genes. Among the ten markers associated with  $\beta$ -carotene, only SNP S06\_47123508, 401 kb from the *ZEP* gene, was in proximity to an a priori candidate gene.

**Table 1** Range, mean, and entry-mean basis heritability ( $H^2$ ) for the BLUPs of lutein, zeaxanthin, and  $\beta$ -carotene for the SAP/CAP collection

|         | Lutein ( $\mu\text{g/g}$ ) | Zeaxanthin ( $\mu\text{g/g}$ ) | $\beta$ -carotene ( $\mu\text{g/g}$ ) |
|---------|----------------------------|--------------------------------|---------------------------------------|
| Minimum | 0.15                       | 0.04                           | 0.07                                  |
| Maximum | 3.09                       | 1.83                           | 0.80                                  |
| Mean    | 0.58                       | 0.18                           | 0.17                                  |
| $H^2$   | 0.80                       | 0.92                           | 0.78                                  |

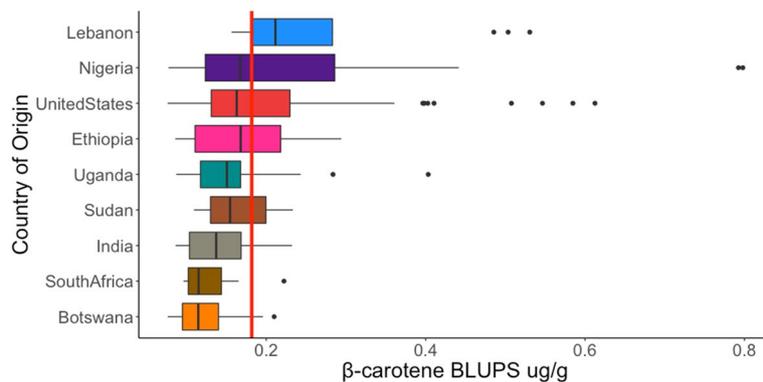


**Fig. 2** Genome-wide association study of carotenoid BLUP estimates using MLM. Manhattan plot of BLUPs for **A**) lutein, **B**) zeaxanthin, and **C**)  $\beta$ -carotene. The red horizontal line represents the genome wide significance threshold for the Bonferroni multiple comparisons correction at  $P=0.05$

**Genetic relationship of sorghum carotenoids in SAP/CAP global collection**

Next, we tested if carotenoid variation is structured by genetic relationship and geographic origin. Since provitamin A carotenoids are our primary target, we focused on  $\beta$ -carotene concentrations for this analysis. Country of origin was obtained from the USDA NPGS GRIN database for the accessions in the SAP/CAP collection.

Countries that had less than eight accessions were discarded from the analysis. In the SAP/CAP collection there were nine countries represented by more than eight accessions: Botswana, Ethiopia, India, Lebanon, Nigeria, South Africa, Sudan, Uganda, and the United States.  $\beta$ -carotene BLUP estimates among this subset of 309 accessions had the same range and average as the full set of SAP/CAP global collection (Table 1, Fig. 3).



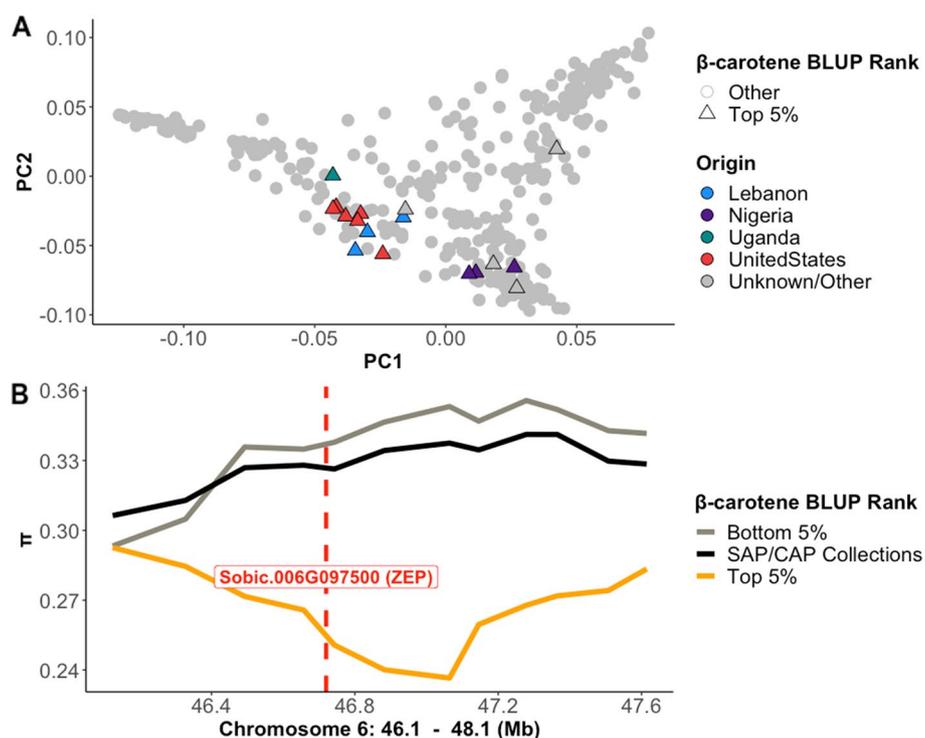
**Fig. 3** Distribution of  $\beta$ -carotene BLUPs among countries in the SAP/CAP collection. The red vertical line represents the SAP/CAP average across all accessions. Countries with less than 8 accessions were excluded

Interestingly, accessions from most of the countries had average values below the global average for  $\beta$ -carotene of  $0.17 \mu\text{g/g}$  (Fig. 3). Furthermore,  $\beta$ -carotene distribution for the accessions of Sudan, South Africa, India, and Botswana were almost completely below the global average. In contrast, accessions from Lebanon had the highest average  $\beta$ -carotene BLUPs estimates with the majority of their accessions above the global average. Notably, Nigeria, had the widest range of variation as well as the highest  $\beta$ -carotene concentrations among the countries.

Based on the limited geographic distribution of high carotenoid sorghums, we hypothesized that the high carotenoid germplasm originates from a narrow genetic pool. To test this hypothesis, we conducted a principal component analysis (PCA) to evaluate the genetic diversity of the high carotenoid accessions identified in the SAP/CAP collection. The high carotenoid accessions were defined as those within the top 5% for  $\beta$ -carotene BLUP estimates, which consisted of 19 accessions ranging from  $0.40$  to  $0.80 \mu\text{g/g}$   $\beta$ -carotene. The majority of the high carotenoid accessions originated in the United States (8 accessions), followed by

Nigeria (3 accessions) and Lebanon (3 accessions) (Fig. 4A). Interestingly, the three high carotenoid accessions from Nigeria grouped together and were clustered separately from most of the other high carotenoid accessions, and another high accession of unknown origin did not group with any other high carotenoid accessions, suggesting three genetically distinct high carotenoid groups (Fig. 4A).

To further test our hypothesis on a narrow genetic pool for high carotenoid lines, we evaluated the genetic diversity surrounding the most prominent SNP identified by GWAS for all three carotenoids (S06\_47123508). We analyzed a window of 1 Mb upstream and downstream of S06\_47123508, which encompassed 1,665 SNPs. Nucleotide diversity was decreased in the high carotenoid accessions, but not in the low carotenoid accessions (defined as the lowest 5% for  $\beta$ -carotene BLUP estimates) or in the complete set of SAP/CAP collection accessions. The most prominent region of low nucleotide diversity was surrounding SNP S06\_47123508, a region which includes the a priori candidate gene encoding *ZEP* (Fig. 4B).



**Fig. 4** Additive genetic relationship of SAP/CAP accessions based on the top 5% and bottom 5% rankings for  $\beta$ -carotene BLUP estimates and allelic diversity surrounding marker trait association. **A** Accessions plotted according to the first two principal components for sorghum kinships coded by country of origin and ranking for  $\beta$ -carotene BLUP estimates. **B** Nucleotide diversity of the region 1 Mb upstream and downstream of marker S06\_47123508. The gray and orange lines represent nucleotide diversity for the bottom and top 5% rankings for  $\beta$ -carotene BLUP estimates, respectively. The black line represents the nucleotide diversity for all of the accessions in the SAP/CAP collections. The red dashed line represents the start position for the *ZEP* gene

### Prediction of carotenoid breeding values in unexplored germplasm collection

Next, we sought to explore if there exists unidentified high carotenoid germplasm in additional germplasm collections. Publicly available genotype data was obtained for germplasm collections from six countries: Ethiopia, Haiti, Niger, Nigeria, Senegal, and Sudan. Together with the SAP/CAP collection, the dataset consisted of 60,129 common SNPs with less than 20% of data missing for 2,488 accessions. There were 361 accessions from Ethiopia, 296 from Haiti, 516 from Niger, 180 from Nigeria, 420 from Senegal, 319 from Sudan, and 396 from the SAP/CAP collection. Most of this germplasm is photoperiod sensitive making it difficult to phenotype in temperate regions such as the United States. For example, among the 396 accessions from the SAP/CAP collection for which we had genotype information, we were only able to phenotype the 345 accessions that were photoperiod sensitive. The inclusion of germplasm from Haiti's breeding program served to generate hypotheses on the best biofortification approach for the program, providing guidance to the breeder by indicating if there is currently promising germplasm in the program or if donor lines must be introduced to introgress high carotenoid alleles.

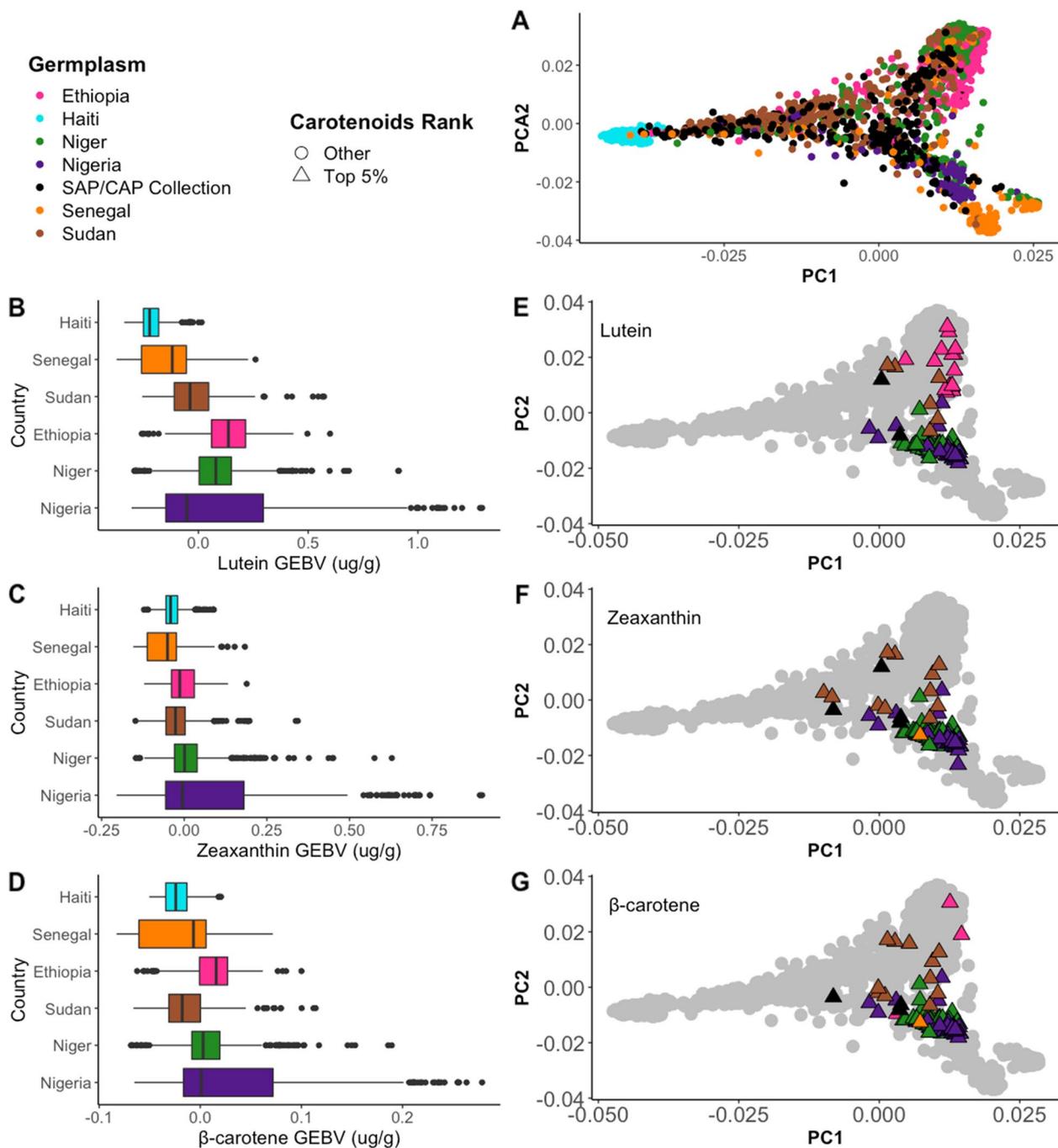
Genomic prediction has the potential to guide resource allocations by identifying the most promising germplasm to test in future work. We first explored the feasibility of the SAP/CAP collection as a training population for the unexplored germplasm collections. For this, the genetic relationship among the unexplored germplasm collections and the SAP/CAP collection was tested with a PCA, highlighting the country of origin for each accession (Fig. 5A, Fig. S2). Based on the scattered distribution observed and the presence of accessions across PCA clusters, the SAP/CAP germplasm collection is an appropriate training population (Fig. S2). Germplasm from Haiti, Ethiopia, Niger, Nigeria, and Senegal formed independent clusters, indicating genetic similarities within but not between countries. Haiti segregated the most distantly, followed by more sparsely grouped germplasm from Senegal and Nigeria. The distant genetic relationship of Haitian germplasm with the other countries was expected as these materials are from a breeding program that went through a recent bottleneck after a sugarcane aphid infestation [36]. Germplasm from Niger and Ethiopia clustered very close together, but separate from the other countries. As expected based on previous studies [37, 38], accessions from the Sudan collection and SAP/CAP collection were scattered across all clusters, rather than clustering together, indicating high genetic diversity. The scattered distribution of the SAP/CAP collection confirms that it

is an appropriate training population for genomic predictions in the unexplored germplasm.

Next, we estimated genomic estimated breeding values (GEBV) from the BLUPs of  $\beta$ -carotene, lutein, and zeaxanthin in the unexplored germplasm collections and the SAP/CAP collection (Table S4). Lutein GEBV ranged from -0.37 to 2.16  $\mu\text{g/g}$ , with an average prediction accuracy of 0.62 and a genomic heritability of 0.96 (Table 2). For zeaxanthin, GEBV ranged from -0.20 to 1.44  $\mu\text{g/g}$ , with a prediction accuracy of 0.69 and a genomic heritability of 1.00 (Table 2). Lastly, for  $\beta$ -carotene, GEBV values ranged from -0.08 to 0.46  $\mu\text{g/g}$ , with a prediction accuracy of 0.67 and a genomic heritability of 0.75 (Table 2). Interestingly, there were no accessions in the unexplored germplasm that had predicted GEBV for  $\beta$ -carotene higher than the highest values in the SAP/CAP collection, however there were some accessions that had values among the highest in all the collections (Fig. S1 and Table S4). Finally, as seen in the SAP/CAP accessions, high correlations were identified for GEBV between  $\beta$ -carotene and zeaxanthin ( $r=0.89$ ;  $p < 10^{-16}$ ),  $\beta$ -carotene and lutein ( $r=0.87$ ;  $p < 10^{-16}$ ), and lutein and zeaxanthin ( $r=0.85$ ;  $p < 10^{-16}$ ).

To further explore geographic patterns of sorghum carotenoid distribution beyond the SAP/CAP collection, we aggregated GEBV by country using the unexplored germplasm collections (Fig. 5B-D). Nigeria had the highest GEBV and range of values for all three carotenoids, followed by Niger. In contrast, Haiti had some of the smallest carotenoid GEBV values, as well as the smallest range of values. Interestingly, Ethiopia had several accessions with high GEBV for lutein, but only three high accessions for  $\beta$ -carotene, and no high accessions for zeaxanthin. Similarly, Senegal had one accession with a high GEBV for  $\beta$ -carotene and zeaxanthin, but not for lutein. These differences suggest that although the three carotenoids are highly correlated—consistent with common genetic controls—there are independent genetic controls, as well.

Next, we looked at the genetic relationships among the predicted top 5% for each carotenoid using a PCA for the GEBV (Fig. 5E-G, Table S5). The pattern of distribution differed by carotenoids, but the majority of the accessions were clustered by country. Lutein (Fig. 5E) had two major clusters corresponding to Ethiopian accessions and a combination of accessions mostly from Nigeria and Niger. For zeaxanthin (Fig. 5F) and  $\beta$ -carotene (Fig. 5G), the clustering was similar, with the accessions of Nigeria and Niger forming the tightest cluster. Accessions from Sudan and the SAP/CAP germplasm were scattered for the three carotenoids, suggesting they are genetically distinct. Taken together, a proportion of accessions predominantly from Nigeria



**Fig. 5** Carotenoids in unexplored germplasm collections. **A** PCA of additive genetic relationship for unexplored germplasm collections and SAP/CAP collection; boxplot of distribution of GEBVs aggregated by country and ordered by lowest to highest carotenoid for **B**) lutein, **C**) zeaxanthin, and **D**)  $\beta$ -carotene; PCA of the genetic relationships of the top 5% of **E**) lutein, **F**) zeaxanthin, and **G**)  $\beta$ -carotene GEBVs in the unexplored germplasm collections

and Niger formed the most distinct cluster in the PCA for the three carotenoids, indicating they are genetically similar. The accessions with the highest GEBV for  $\beta$ -carotene were also part of this cluster.

**Allelic diversity and geographic distribution of ZEP allele**

To further test the hypothesis that high carotenoid lines originate from a narrow genetic pool, we analyzed the SNPs inside the *ZEP* gene in the SAP/CAP collection

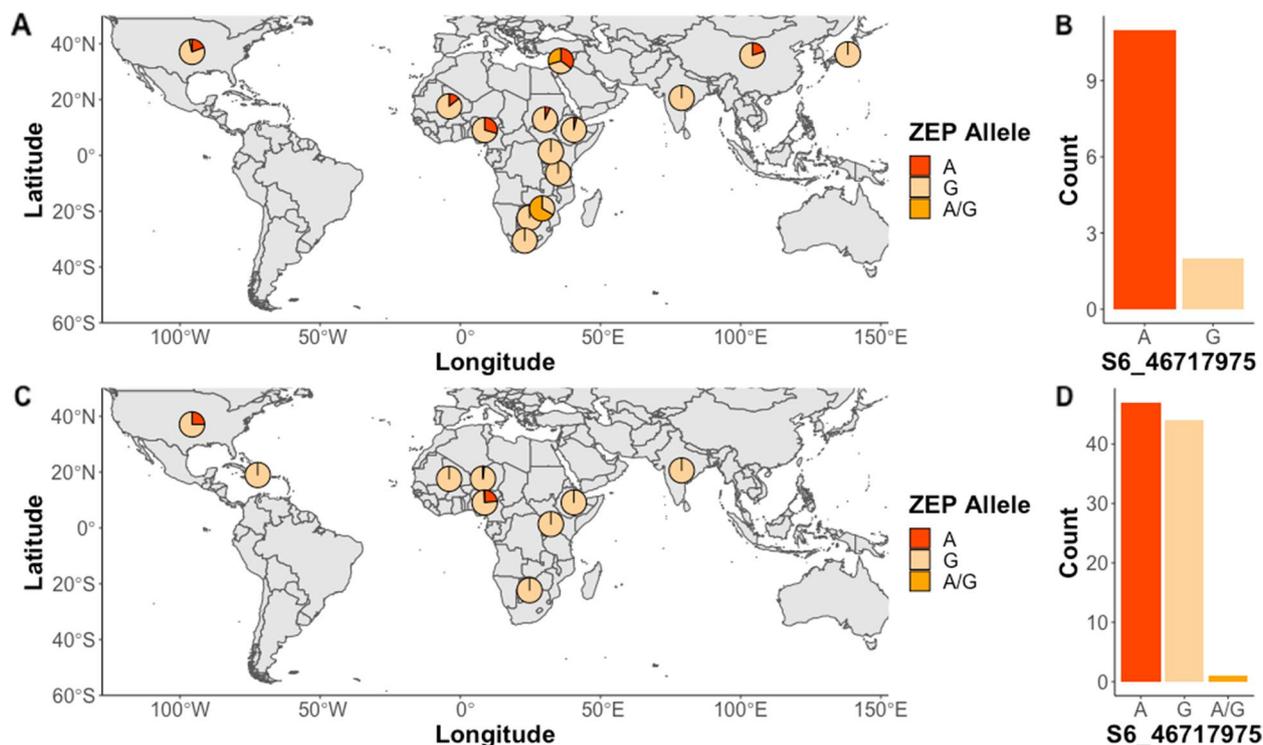
**Table 2** Range of GEBV, average prediction accuracy and genomic heritability ( $H^2$ ) for the lutein, zeaxanthin, and  $\beta$ -carotene for the unexplored germplasm collections and SAP/CAP collection

|                             | GEBV Lutein ( $\mu\text{g/g}$ ) | GEBV Zeaxanthin ( $\mu\text{g/g}$ ) | GEBV $\beta$ -carotene ( $\mu\text{g/g}$ ) |
|-----------------------------|---------------------------------|-------------------------------------|--|
| Minimum                     | -0.37                           | -0.20                               | -0.08                                      |
| Maximum                     | 2.16                            | 1.44                                | 0.46                                       |
| Average Prediction Accuracy | 0.62                            | 0.69                                | 0.67                                       |
| $H^2$                       | 0.96                            | 1.00                                | 0.75                                       |

and unexplored germplasm collections. In the SAP/CAP collection, we identified 14 SNPs in the *ZEP* gene with minor allele frequency (MAF) > 0.05. Due to low marker density, the majority of these SNPs were absent in the unexplored germplasm collections. However, SNP S06\_46717975 was present in the SAP/CAP collection and the SNP data set for Haiti, Niger, and Nigeria germplasm (Table S6). This SNP is found within the *ZEP* gene and was previously identified by our group as associated

with zeaxanthin variation [15]. S06\_46717975 was found to be bi-allelic with A/G variants present among the germplasm. The minor allele 'A' is moderately common globally, with 10% presence in SAP/CAP collection. However, among countries there are striking differences in the allele frequency; for instance, 24% in Nigeria versus 2% in Niger and 0% in Haiti germplasm.

We next explored if there were any patterns between allelic variant, geographic distribution, and carotenoid content (Fig. 6). In the SAP/CAP collection, there was a correlation between allelic type and country of origin with the United States, Lebanon, and Nigeria, the countries with the highest prevalence of the 'A' allele (Fig. 6A). Among the high carotenoid accessions in the SAP/CAP collection (defined as the top 5% for  $\beta$ -carotene concentration), the 'A' allele was present in 85% of them (Fig. 6B). We then analyzed the alleles in the unexplored germplasm collections and the SAP/CAP accessions that were not phenotyped. Similar patterns were observed for the geographic distribution of the 'A' allele with the highest prevalence in the United States and Nigeria (Fig. 6C). Surprisingly, the difference in the distribution of the 'A' and 'G' alleles was not nearly as pronounced in



**Fig. 6** Geographic distribution of the SNP S06\_46717975 inside the *ZEP* gene and the distribution of allelic classes in high carotenoid germplasm. **A** Geographic distribution for accessions in the SAP/CAP collection. **B** Distribution of allelic classes for the top 5% rankings for  $\beta$ -carotene BLUP estimates. **C** Geographic distribution for accessions in unexplored germplasm collections and the SAP/CAP accessions without phenotype; **D** Distribution of allelic classes for the top 5% rankings for  $\beta$ -carotene GEBV where A is the minor allele

the predicted high carotenoid lines based on  $\beta$ -carotene GEBV (Fig. 6D).

## Discussion

### Genetic diversity among high carotenoid lines

A vitamin biofortified sorghum variety has the potential to positively impact the livelihood of millions that rely on it as a dietary staple. We estimate a  $\beta$ -carotene target value of 12  $\mu\text{g/g}$  for biofortified sorghum grain. Although the highest  $\beta$ -carotene content measured in our study was 1.19  $\mu\text{g/g}$ , a previous study reported a sorghum variety with  $\beta$ -carotene concentrations as high as 3.23  $\mu\text{g/g}$  [20]. In that study, crosses with high carotenoid parents resulted in F2 progeny with  $\beta$ -carotene as high as 3.57  $\mu\text{g/g}$ , suggesting that classical breeding can increase concentrations further. Overall, our findings, along with these previous studies, suggest that sorghum provitamin A carotenoid biofortification is feasible using breeding coupled with modern genomic breeding tools.

In order to ensure continuous genetic gains and trait improvement, genetic diversity is necessary. For carotenoid content in sorghum, however, this might be a limitation, because high carotenoid lines appear to be highly related (Figs. 4A and 5E-G). The tight clustering of countries (Fig. 5E-G) and few countries with high carotenoid lines suggest that there has not been much exchange of germplasm among the countries and most germplasm might not have high carotenoid alleles. Interestingly, here we identified Nigeria, Lebanon, and the United States as the major origins with high carotenoid germplasm both in the SAP/CAP collection and the unexplored germplasm collections. However, it seems that the high carotenoid lines from Lebanon and the United States are from Nigerian origin. In the 1950s, a breeder named O.J. Webster collected yellow endosperm kaura sorghums from Nigeria, which were subsequently used for breeding material in the United States [39]. Some of these kaura lines were then sent by another breeder, R.E. Karper, from the United States to the Arid Land Agricultural Development (ALAD) Program in Lebanon, which eventually became the International Center for Agricultural Research in the Dry Areas (ICARDA) [40, 41]. This relationship between the kaura accessions from the United States and Lebanon explains the close genetic similarity identified in our study between accessions from the United States and Lebanon (Fig. 4A).

Selection for kaura types could also be the underlying driver of the limited genetic diversity and selection signals observed for the high carotenoid lines (Fig. 4B). In Nigeria, kaura types are one of the most common sorghum landraces grown due to their high yield, drought resistance, and grain quality [42]. They also have

widely-sought agronomic traits as they are generally of short stature, have large yellow seeds, and are photoperiod insensitive. In the United States, selection for kaura types could also have contributed further to the limited genetic diversity. The first yellow hybrids developed in the United States, using Nigerian germplasm, had stronger root development, improved stay-green, and resistance to charcoal rot [39], which could have led to the incorporation of high carotenoid alleles into multiple pedigrees. Also, among the high carotenoid accessions identified in our study, several of them are listed in GRIN as kaura (*durra-caudatums*), which could support the indirect selection for carotenoids among the kaura types. All together, these results suggest that kaura types from Nigeria are the main source of high carotenoid alleles and that efforts to increase diversity can focus on them.

Here, we identified 107 accessions from Nigeria, Niger, Ethiopia, Senegal, Sudan and the SAP/CAP collection (Table S5) that have a high GEBV for  $\beta$ -carotene. These accessions need to be phenotyped with HPLC to test the hypothesis that they have high  $\beta$ -carotene concentrations and have potential as donor parents for breeding efforts. It should be noted that if the genetic diversity of the SAP/CAP collection does not fully encompass the range of genetic diversity of the unexplored germplasm, then we may have underestimated some of the predicted values and failed to identify some high carotenoid lines. Importantly, some of the lines hypothesized to have high carotenoids based on GEBV are highly genetically divergent from the high-carotenoid Nigerian germplasm (e.g. Ethiopia and Sudanese lines in Fig. 5, Table S5), suggesting that previously untapped high-carotenoid germplasm exists. It should be noted that since there are so many high carotenoid accessions of the kaura type, it is possible that the signal in the GEBV is related more to the botanical type and geographic origin than high carotenoid concentrations, per se. In the future this can be tested through crosses that break up population structure using germplasm that vary for carotenoid levels. For example, biparental populations could be developed with kaura parental lines crossed with germplasm from other regions. Given that sorghum germplasm does not meet current target values, breeding will be necessary to increase  $\beta$ -carotene concentrations. The high heritabilities for  $\beta$ -carotene [15, 16] suggest that increasing concentrations through breeding is possible. Developing crosses among high  $\beta$ -carotene lines in the SAP/CAP collection as well as those identified by genomic prediction, can provide insights into if there is enough genetic diversity to reach target values. Genomics-assisted breeding via MAS or GS has the potential to accelerate efforts by simplifying selection methods.

### Marker-assisted selection for sorghum carotenoids

MAS could be a viable alternative to select carotenoids in sorghum given that GWAS suggests an oligogenic architecture (Fig. 2 and [15]). MAS for carotenoids has been tested in cassava with a marker linked to the *PSY* gene, which initiates the first committed step in the carotenoid biosynthesis pathway, demonstrating prediction accuracies above 0.8 [43]. MAS has also been implemented in maize biofortification efforts with markers in linkage with the biosynthesis genes *lcyE* [44], *crtRB1* [44–46], *ZEP* [45], and *opaque 2* [47]. For a successful implementation of MAS for sorghum carotenoids, breeder-friendly markers (i.e. convenient and low-cost) with tight linkage and high LD with target alleles must be developed [48].

Chromosome 6 might be a good place to start for sorghum carotenoid marker development due to the high number of associations detected for  $\beta$ -branch carotenoids, where most of the provitamin A carotenoids are synthesized. Four regions of association on chromosome 6 have been identified: 46.7 Mb [15], 50.3–53.5 Mb [16], 57.4 Mb [15] and 47.1 Mb in this study. For  $\beta$ -carotene, associations have been detected in proximity to *phytoene desaturase* (*PDS*, Sobic.006G232600) [15], the second enzyme in the carotenoid biosynthesis pathway. For zeaxanthin, associations have been detected near *zeta-carotene desaturase* (*ZDS*, Sobic.006G177400) [16] and *ZEP* (Sobic.006G097500) [15]. Interestingly, in this study we also identified significant associations near *ZEP* for lutein, zeaxanthin, and  $\beta$ -carotene (Fig. 2, Table S3). These genes have also been associated with natural variation of carotenoids in maize [30]. If no linkage drag is present, the prevalence of associations on chromosome six could mean that positive alleles for multiple genes could be introduced simultaneously, reducing the generations needed. Understanding the allelic diversity of these genes in sorghum germplasm and the expression profiles among high carotenoid germplasm could further demonstrate their potential for utilization in vitamin A biofortification.

The *ZEP* gene in sorghum could be a candidate to initiate such efforts. Based on this and our previous study [15] *ZEP* is a core gene controlling variation in the sorghum  $\beta$ -branch carotenoids, i.e. zeaxanthin and  $\beta$ -carotene. *ZEP* is also a good candidate for breeding efforts through MAS as it seems to have allelic variants with strong correlation with carotenoid concentrations. Marker S06\_46717975 is a biallelic allele (A/G) in the *ZEP* coding sequence and is in proximity to S06\_47123508, here associated with  $\beta$ -carotene and zeaxanthin. In the SAP/CAP collection, the allele was minor with only 10% of the accessions having the 'A' allele. The geographic distribution of the 'A' allele also correlates with Nigeria, United States, and Lebanon, the countries that had the highest

observed or predicted  $\beta$ -carotene concentrations (Fig. 6). Interestingly, among the unexplored germplasm collections, the top 5% had a more balanced distribution of the A/G variants. One hypothesis is that the top 5% of  $\beta$ -carotene GEBV might capture a wider carotenoid content and diversity than what is present in the SAP/CAP collection and therefore the allele has not been fixed (Table S6). The higher prevalence of countries in the top 5% of GEBV for  $\beta$ -carotene, the lower GEBV when compared to the SAP/CAP collection (Fig. S1), and the more sparsely grouped cluster (Fig. 5G) here observed supports this hypothesis. However, we hypothesize that due to the high correlation of allelic variant and carotenoid content observed and predicted, marker S06\_46717975 could be used for MAS and to identify potential donor lines. Further germplasm evaluations are needed to assess this marker's predictive ability.

Finally, it should be noted that as with previous studies, GWAS here identified a total of 27 regions of association, but only two of those regions were near known carotenoid pathway genes. This demonstrates that there are still many unknown genes involved in carotenoid variation, which are perhaps regulatory pathway controls or unidentified homologues of carotenoid biosynthesis or degradation genes. Further studies, such as transcriptomics, are needed to help find the causal genes in linkage with markers identified through GWAS.

### Genomic selection for sorghum carotenoids

Genomic selection is an alternative to MAS that is increasingly used for complex traits as genotypic costs decrease. For quality traits GS could potentially reduce the cost compared to phenotyping, and reduce the need for specialized equipment and training. In wheat, GS has been proven to be superior for quality traits over MAS as it allows for the selection for both small and large effect loci [49–51]. GS for carotenoids has yet to be implemented in breeding programs, but it has been tested in cassava [52] and maize [53]. Here we report the first study on genomic prediction for sorghum carotenoids. Genomic predictions are designed to capture polygenic variance and allow for selection on complex traits. GS accuracy and efficiency is dependent on several factors, particularly heritability of the trait, because it often directly translates into the prediction accuracy. The high heritability estimates (0.78, Table 1) and prediction accuracy (0.67, Table 2) here obtained for  $\beta$ -carotene would suggest that there is a polygenic component to sorghum carotenoids and GS can be an efficient method for biofortification. One hypothesis that explains why we see evidence for both oligogenic and polygenic variation is that sorghum carotenoids are omnigenic traits, in which a small number of core genes directly regulate

carotenoids and a large number of peripheral genes that are expressed in the grain indirectly regulate carotenoids [54]. This hypothesis could be tested with a genome-wide expression study in high carotenoid germplasm.

Despite its potential, there are several factors to consider in GS for sorghum carotenoids. First, the results of this study suggest that most countries' germplasm currently lack enough phenotypic and genotypic diversity for sorghum carotenoids. This suggests that the next step for provitamin A biofortification would be to introduce high carotenoid alleles into these breeding programs via pre-breeding. Given that oligogenic variation for grain carotenoids exists, this initial introduction of alleles could be accomplished with MAS. Second, even though GS has the potential to reduce cost of phenotyping, simulation studies suggest that depending on population size, genotyping costs must be under \$15 (U.S. Dollars) to be more cost-effective than simple phenotypic selection [55]. This genotyping cost can make GS unrealistic for breeding programs in developing countries, which would be the ones to benefit the most from a biofortified sorghum, as it is estimated that genotyping several hundred SNP markers remains at \$14 (U.S. Dollars) [56].

Lastly, incorporation of a GS scheme for a young breeding program can be very challenging. GS has the biggest potential for genetic gain per unit of time when breeding cycles are closed rapidly and effectively [55–57]. However, many breeding programs in developing countries have slow breeding cycles with recycling improved lines as a parent often taking well over 10 years [58]. Therefore, under these scenarios we suggest the direction for biofortification breeding will be to first introduce major genes through MAS in breeding programs. After the introduction of these alleles, and as genotyping cost continues to decrease, MAS in tandem with GS can then be used for continuous improvement. If carotenoid variation in sorghum is in fact both oligogenic and polygenic, then the incorporation of MAS, GS, and rapid breeding cycles could substantially increase  $\beta$ -carotene to target values and ensure continuous genetic gains.

## Conclusions

In this study we evaluated carotenoid concentrations in SAP/CAP collection, identifying the accessions with the highest  $\beta$ -carotene concentrations. Also, it was established that current concentrations of  $\beta$ -carotene are low and current known high  $\beta$ -carotene germplasm has a narrow genetic diversity. We used the SAP/CAP collection as a training population to predict the genetic merit or GEBV via genomic prediction for unexplored germplasm. Based on GEBV, we present 107 accessions with

the highest predicted concentrations for  $\beta$ -carotene that potentially represent novel genetic variation for the trait. Finally, we proposed that MAS should be initially used to introduce high carotenoid alleles like S06\_46717975 inside the *ZEP* gene into breeding programs followed by GS for continuous improvement.

## Methods

### Plant material

Grain carotenoid concentration was evaluated for a total of 446 sorghum accessions, which included 316 from the sorghum association panel (SAP) [59] and 130 accessions from the carotenoid panel (CAP), a set of accessions chosen for presence of yellow endosperm and/or yellow grain [15, 60]. The two panels were grown, selfed, and harvested by the authors at Kansas State University Agronomy North Farm in Manhattan, Kansas with a randomized complete block design with two replications during the summer of 2019. At maturity, grain was harvested, dried and stored at  $-80^{\circ}\text{C}$  until carotenoid quantification.

### Carotenoid quantification

Carotenoid extractions were performed following a modified solid phase method [61]. All steps of the extraction were carried out under yellow light to avoid photodegradation. Briefly, approximately 5 seeds were ground to flour using a Bead Ruptor Elite (Omni International, Kennesaw, GA) and 20 mg of the sorghum flour were transferred to a 1.5 mL eppendorf tube. Next, 20 mg of ascorbic acid and 400  $\mu\text{L}$  of absolute ethanol with a 1 mg/mL concentration of butylated hydroxytoluene (BHT) were added. The tubes were vortexed for 1 min and placed in an  $80^{\circ}\text{C}$  water bath for 5 min. Following the incubation, 20  $\mu\text{L}$  of a solution of potassium hydroxide (80% w/v, in water) was added and tubes vortexed for 1 min. Next, samples were returned to the water bath for 15 min and mixed every 5 min. Samples were then brought to room temperature and centrifuged for 5 min at 1800 rcf. Supernatant was transferred to a new 1.5 mL eppendorf tube. An additional 400  $\mu\text{L}$  of absolute ethanol with 1 mg/mL of BHT was added to the residue, vortexed for 1 min and centrifuged for 5 min at 1800 rcf. The supernatant was combined with the above extract, vortexed for 30 s and centrifuged for 5 min at 5000 rcf. The supernatant was then transferred to a new 1.5 mL eppendorf tube and evaporated to dryness with a gentle  $\text{N}_2$  stream at room temperature. Finally, the residue was reconstituted in 100  $\mu\text{L}$  of Methanol:Ethyl Acetate (1:1) and centrifuge for 5 min at 5000 rcf. An aliquot of 40  $\mu\text{L}$  of the clear supernatant was utilized for the HPLC analysis. Resolution of lutein, zeaxanthin,  $\beta$ -carotene,  $\alpha$ -carotene and  $\beta$ -cryptoxanthin

was conducted using an Agilent 1290 Infinity UHPLC with Eclipse Plus C18 column (Agilent Technologies, Santa Clara, California, U.S).

### Statistical analysis of carotenoids and heritability estimation

Concentrations for lutein, zeaxanthin and  $\beta$ -carotene were analyzed with ASReml-R package [62], which accounts for missing data. For the three carotenoids, we implemented a randomized complete block design model with genotype and block as random effects. The gamma parameterization with a maximum iteration number of 100 was used for the analysis. Best linear unbiased predictors (BLUPs) were obtained as predictors of genetic merit for lutein, zeaxanthin and  $\beta$ -carotene and were used for subsequent analysis. Broad sense heritability on an entry-mean was also calculated for lutein, zeaxanthin and  $\beta$ -carotene as followed:

$$H^2 = \frac{\sigma_{Genotype}^2}{\sigma_{Phenotype}^2} = \frac{\sigma_{Genotype}^2}{\sigma_{Genotype}^2 + \frac{\sigma_{error}^2}{r}}$$

where  $H^2$  represents the broad sense heritability on a entry-mean basis,  $\sigma_{Genotype}^2$  represents the genotypic variance,  $\sigma_{Phenotype}^2$  the phenotypic variance and  $\sigma_{error}^2$  the residual or error variance, and  $r$  represents the blocks. Pearson correlations were calculated between carotenoid pairs for lutein, zeaxanthin, and  $\beta$ -carotene.

### Genome-wide association study

The genetic architecture and the genomic regions underlying carotenoid variation in sorghum grain were investigated through a genome-wide association study (GWAS) implemented in GAPIT [63], version “2022.4.16, GAPIT 3.1”. The genotype information for the sorghum association panel and the carotenoid panel was obtained from previous studies [15, 64]. The SNP datasets are available for download from the Dryad Data Repository (doi:10.5061/dryad.63h8fd4). After filtering the single nucleotide polymorphism (SNP) data set (Sorghum bicolor v3.1 genome version) by a minimum allele frequency of 0.05, 348,181 biallelic SNPs remained. A total of 345 accessions for the SAP/CAP collection had both genotype information and BLUP estimates for lutein,  $\beta$ -carotene, and zeaxanthin. A mixed linear model (MLM) (model=“MLM”) was used with a marker derived kinship and ten principal components (PCA.total=10) to control for relationship and population structure, respectively. To account for multiple comparisons, the Bonferroni correction with  $P=0.05$  was used to identify significant SNPs. Significant associations were compared with candidate genes that are annotated as enzymes involved in the carotenoid pathway in

Phytozome or that have been identified in other carotenoid association studies (Table S7).

### Diversity of high carotenoids lines in the SAP/CAP global collection

Genetic diversity among the high carotenoid lines identified was examined. We prioritized assessing the genetic diversity among the accessions with the highest  $\beta$ -carotene BLUP estimates, because  $\beta$ -carotene is the most abundant provitamin A carotenoid in sorghum. Accessions were ranked as top 5% and bottom 5% based on their BLUP estimate for  $\beta$ -carotene. A marker-derived additive relationship matrix or kinship, was estimated with the ‘A.mat’ function in the rrBLUP R package [65]. The eigenvalues for the first two principal components were estimated with R function ‘eigen’ for the additive relationship matrix and the grouping of the top 5% was examined. Genetic diversity of regions associated with  $\beta$ -carotene variation was determined using a window of 1 megabase (Mb) upstream and downstream of significant SNPs identified by GWAS in proximity to a priori candidate genes. The linkage disequilibrium (LD) for the region was calculated with rTASSEL [66] for all the SNPs within the 2 Mb window and setting heterozygous as “missing” (Fig. S3). Nucleotide diversity ( $\pi$ ) per base pair was calculated with rTASSEL [66] using a step size of 100 and a window size of 500.

### Genotype from unexplored germplasm collections

Publicly available genotype data from unexplored germplasm collections were gathered. Accessions and their corresponding genotype information from Ethiopia [67], Haiti [36], Niger [68], Nigeria [64, 69, 70], Senegal [71] and Sudan [37] were obtained from published data or by contacting the authors. Common SNPs between the unexplored germplasm collections and the SAP/CAP global collection that had at least 80% of the data present were identified. Missing SNP data were then imputed using Beagle [72] with the default parameters. To assess the genetic relationships among the accessions in the unexplored germplasm and the SAP/CAP collection a realized additive relationship matrix was calculated first using the ‘A.mat’ function in rrBLUP R package [65]. The additive relationship matrix was then used to perform a principal component analysis (PCA) in R.

### Genomic prediction of GEBV for carotenoids in unexplored germplasm

Predictions of the genomic estimated breeding values (GEBV) and genomic heritability were conducted using the genomic data from unexplored germplasm, representing country collections of Ethiopia, Haiti, Niger, Nigeria, Senegal, Sudan, and the SAP/CAP collection.

The accessions in the SAP/CAP collection for which we had genotype information and BLUPs estimates for lutein, zeaxanthin, and  $\beta$ -carotene were used as the training population ( $n=345$ ). GEBV were estimated for lutein, zeaxanthin and  $\beta$ -carotene using the G-BLUP model with the additive relationship matrix or kinship as implemented in the rrBLUP package in R [65]. A fivefold cross validation approach was used for each carotenoid to determine prediction accuracy. The cross validation was repeated for 100 cycles. Genomic heritability for lutein, zeaxanthin and  $\beta$ -carotene was estimated during each k-fold and cycle. Prediction accuracy was also estimated by calculating the correlation between the genomic prediction and the validation values divided by square root of heritability [73]. The unexplored germplasm was ranked in the top 5% for lutein, zeaxanthin and  $\beta$ -carotene based on the GEBV estimates. A marker-derived kinship was estimated with the 'A.mat' function in the rrBLUP R package [65] for the unexplored germplasm. The eigenvalues for the first two principal components were estimated with R function 'eigen' for the additive relationship matrix and the grouping of the top 5% for each of the carotenoids was examined. Additionally, the distribution of the GEBVs for  $\beta$ -carotene was evaluated by country. Lastly, we compared the GEBV for the three carotenoids in the SAP/CAP collection and unexplored germplasm collections.

#### Allelic diversity and geographic distribution of ZEP allele

Distribution of the ZEP allele for the SAP/CAP collection and the unexplored germplasm collections was examined using country of origins and the allelic classes for the SNP S06\_46717975. Countries that had less than 3 accessions were discarded from the analysis. We also aggregated the allelic variants present in the high  $\beta$ -carotene accessions from the SAP/CAP collection and the unexplored germplasm collections, as defined by the top 5% of BLUP or GEBV for  $\beta$ -carotene.

#### Abbreviations

|      |  |
|------|--|
| BLUP | Best linear unbiased prediction        |
| CAP  | Carotenoid panel                       |
| FDR  | False discovery rate                   |
| GWAS | Genome-wide association study          |
| GEBV | Genomic estimated breeding value       |
| GGPS | Geranylgeranyl diphosphate synthase    |
| GS   | Genomic selection                      |
| HPLC | High-performance liquid chromatography |
| MAS  | Marker-assisted selection              |
| Mb   | Megabase                               |
| MLM  | Mixed linear model                     |
| PDS  | <i>Phytoene desaturase</i>             |
| PSY  | <i>Phytoene synthase</i>               |
| PCA  | Principal component analysis           |
| SAP  | Sorghum association panel              |
| SNP  | Single nucleotide polymorphism         |
| ZEP  | <i>Zeaxanthin epoxidase</i>            |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04176-0>.

**Additional file 1: Table S1.** Raw carotenoid concentrations in SAP/CAP. **Table S2.** BLUPs for carotenoids in the SAP/CAP. **Table S3.** Carotenoids marker-trait associations in the SAP/CAP. **Table S4.** Carotenoid GEBV in SAP/CAP and unexplored germplasm. **Table S5.** Top 5%  $\beta$ -carotene GEBV in unexplored germplasm. **Table S6.** Allelic distribution of ZEP SNP S06\_46717975. **Table S7.** A priori carotenoid candidate genes. **Table S8.** Common name, endosperm color, kernel color, race, and working group of evaluated accessions.

**Additional file 2: Fig. S1.** Comparison of GEBV for lutein, zeaxanthin, and  $\beta$ -carotene for the SAP/CAP global collection and the unexplored germplasm collections. Boxplot of GEBV for A) lutein; B) zeaxanthin; and C)  $\beta$ -carotene.

**Additional file 3: Fig. S2.** PCA of genetic relationship between SAP/CAP and unexplored germplasm.

**Additional file 4: Fig. S3.** LD for 1 Mb region upstream and downstream of marker S06\_47123508.

#### Acknowledgements

Not applicable.

#### Authors' contributions

CCB conceptualized the project, collected, analyzed, and interpreted the data, and wrote the manuscript. DR conceptualized the project and assisted in interpretation of the results, writing, and editing; GM participated in interpretation of results and editing. All authors read and approved the final manuscript.

#### Funding

This work received financial support provided by Foundation for Food & Agriculture Research (FF-NIA20-000000036).

#### Availability of data and materials

The variant data for this study have been deposited in the European Variation Archive (EVA) at EMBL-EBI under accession number PRJEB60406. <https://www.ebi.ac.uk/eva/?evastudy=PRJEB60406> Original publication with the genotype data can be found in the following citations:

- Ethiopia germplasm [67]
- Haiti germplasm [36]
- Niger germplasm [68]
- Nigeria germplasm [64, 69, 70]
- Senegal germplasm [71]
- Sudan germplasm [37]
- SAP/CAP germplasm [64]

#### Declarations

##### Ethics approval and consent to participate

Seeds were obtained and are freely available from the U.S. National Plant Germplasm System. Permission to plant at the Agronomy North Farm was granted by Kansas State University. All methods were carried out in accordance with relevant guidelines and regulation.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Horticulture & Landscape Architecture, Colorado State University, Fort Collins, CO 80523, USA. <sup>2</sup>Department of Soil & Crop Science, Colorado State University, Fort Collins, CO 80523, USA.

Received: 19 June 2022 Accepted: 17 March 2023  
Published online: 28 March 2023

## References

- Trono D. Carotenoids in Cereal Food Crops: Composition and Retention throughout Grain Storage and Food Processing. *Plants*. 2019;8(12):551. Available from: <https://www.mdpi.com/2223-7747/8/12/551>. [cited 13 Jan 2022].
- Tanumihardjo SA, editor. Carotenoids and Human Health. Humana Press. Totowa, NJ. 2013. [cited 15 May 2022]. <https://doi.org/10.1007/978-1-62703-203-2>.
- Menkir A, Rocheford T, Maziya-Dixon B, Tanumihardjo S. Exploiting natural variation in exotic germplasm for increasing provitamin-A carotenoids in tropical maize. *Euphytica*. 2015;205(1):203–17. <https://doi.org/10.1007/s10681-015-1426-z>. [cited 23 Jan 2022].
- Jeffery JL, Turner ND, King SR. Carotenoid bioaccessibility from nine raw carotenoid-storing fruits and vegetables using an in vitro model. *J Sci Food Agric*. 2012 [cited 31 May 2022];92(13):2603–10. <https://doi.org/10.1002/jsfa.5768>.
- Bouis H. Reducing Mineral and Vitamin Deficiencies through Biofortification: Progress Under HarvestPlus. In: Biesalski HK, Birner R, editors. *World Review of Nutrition and Dietetics*. S. Karger AG; 2018 [cited 9 May 2022]. p. 112–22. Available from: <https://www.karger.com/Article/FullText/484342>
- Sikora E, Bodziarczyk I. Composition And Antioxidant Activity Of Kale (Brassica Oleracea L. Var. Acephala) Raw And Cooked. *Acta Sci Pol Technol Aliment*. 2012;10.
- von Grebmer K, Saltzman A, Birol E, Wiesmann D, Prasai N, Yin S, Yohannes Y, Menon P, Thompson J, Sonntag A. 2014 Global Hunger Index: The Challenge of Hidden Hunger. Bonn, Washington, D.C., and Dublin: Welt-hungerhilfe, International Food Policy Research Institute, and Concern Worldwide; 2014. <https://doi.org/10.2499/9780896299580>.
- WHO. Global prevalence of vitamin A deficiency in populations at risk: 1995–2005. WHO Global database on vitamin A deficiency. 2009 [cited 13 Jun 2019]. Available from: [https://apps.who.int/iris/bitstream/handle/10665/44110/9789241598019\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/44110/9789241598019_eng.pdf)
- Bouis H, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH. Biofortification: a new tool to reduce micronutrient malnutrition. *Food Nutr Bull*. 2011;32(1):S31–40.
- Harvest Plus. Biofortified Crops Released. 2022 [cited 12 June 2022]. Available from: [https://bcr.harvestplus.org/varieties\\_released\\_list?crop=Vitamin\\_A\\_Maize](https://bcr.harvestplus.org/varieties_released_list?crop=Vitamin_A_Maize)
- FAOSTAT. Food and Agriculture Organization of the United Nations. 2021 [cited 2021 May 26]. Available from: <http://www.fao.org/faostat/en/#data/QC>
- Affy AEMM, El-Beltagi HS, El-Salam SMA, Omran AA. Biochemical changes in phenols, flavonoids, tannins, vitamin E,  $\beta$ -carotene and antioxidant activity during soaking of three white sorghum varieties. *Asian Pac J Trop Biomed*. 2012 [cited 13 Jan 2022];2(3):203–9. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S2221169112600422>
- Blessin CW, Dimler RJ, Webster OJ. Carotenoids of corn and sorghum II. Carotenoid loss in yellow-endosperm sorghum grain during weathering. *Cereal Chem*. 1962;39:389–92.
- Cardoso L de M, Pinheiro SS, da Silva LL, de Menezes CB, de Carvalho CWP, Tardin FD, et al. Tocochromanols and carotenoids in sorghum (*Sorghum bicolor* L.): Diversity and stability to the heat treatment. *Food Chem*. 2015 [cited 23 Jan 2022];172:900–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0308814614015027>
- Cruet-Burgos C, Cox S, Ioerger BP, Perumal R, Hu Z, Herald TJ, et al. Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *Plant Genome*. 2020 [cited 13 Jan 2022];13(1). <https://doi.org/10.1002/tpg2.20013>
- Fernandez MGS, Hamblin MT, Li L, Rooney WL, Tuinstra MR, Kresovich S. Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Sci*. 2008;48(5):1732–43. <https://doi.org/10.2135/cropsci2007.12.0684>. [cited 13 Jan 2022].
- Fu WN. Agronomic Characteristics, Protein, And Carotenoid Composition Of Some Grain Sorghum Varieties, Strains, And Hybrids—With Emphasis On Yellow Endosperm Types [Master of Science]. Oklahoma State University; 1960.
- Kean EG, Bordenave N, Ejeta G, Hamaker BR, Ferruzzi MG. Carotenoid bioaccessibility from whole grain and decorticated yellow endosperm sorghum porridge. *J Cereal Sci*. 2011 [cited 13 Jan 2022];54(3):450–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0733521011001500>
- Shen Y, Su X, Rhodes DH, Herald TJ, Xu J, Chen X, et al. The pigments of sorghum pericarp are associated with the contents of carotenoids and pro-vitamin A. *Int J Food Nutr Sci*. 2017;6(3):48–56.
- Worzella WW, Khalidy R, Badawi Y, Daghir S. Inheritance of beta-carotene in grain sorghum hybrids. *Crop Sci*. 1965;5(6):591–2.
- Che P, Zhao ZY, Glassman K, Dolde D, Hu TX, Jones TJ, et al. Elevated vitamin E content improves all- trans  $\beta$ -carotene accumulation and stability in biofortified sorghum. *Proc Natl Acad Sci*. 2016;113(39):11040–5. <https://doi.org/10.1073/pnas.1605689113>. [cited 13 Jan 2022].
- Zhao ZY. The Africa Biofortified Sorghum Project—Applying Biotechnology to Develop Nutritionally Improved Sorghum for Africa. In: Xu Z, Li J, Xue Y, Yang W, editors. *Biotechnology and Sustainable Agriculture 2006 and Beyond*. Springer Netherlands: Dordrecht; 2007. p. 273–7.
- Tembo L. Production and Adoption of Transgenic Crops in Sub-Saharan Africa. *Asian Res J Agric*. 2021 [cited 10 June 2022];32–41. Available from: <https://www.journalarja.com/index.php/ARJA/article/view/30122>
- Wambugu F, Kamanga D, editors. *Biotechnology in Africa*. Springer International Publishing. Cham. 2014 [cited 13 Jan 2022]. (Science Policy Reports; vol. 7). <https://doi.org/10.1007/978-3-319-04001-1>
- Hirschberg J. Carotenoid biosynthesis in flowering plants. *Curr Opin Plant Biol*. 2001 [cited 10 June 2022];4(3):210–8. Available from: <https://www.sciencedirect.com/science/article/pii/S1369526600001631>
- Azmach G, Menkir A, Spillane C, Gedil M. Genetic Loci Controlling Carotenoid Biosynthesis in Diverse Tropical Maize Lines. *G3 GenesGenomes-Genetics*. 2018 [cited 21 Jan 2022];8(3):1049–65. Available from: <https://academic.oup.com/g3journal/article/8/3/1049/6027158>
- Ikoma Y, Matsumoto H, Kato M. Diversity in the carotenoid profiles and the expression of genes related to carotenoid accumulation among citrus genotypes. *Breed Sci*. 2016 [cited 31 May 2022];66(1):139–47. Available from: [https://www.jstage.jst.go.jp/article/jsbs/66/1/66\\_139/\\_article](https://www.jstage.jst.go.jp/article/jsbs/66/1/66_139/_article)
- Jourdan M, Gagné S, Dubois-Laurent C, Maghraoui M, Huet S, Suel A, et al. Carotenoid content and root color of cultivated carrot: a candidate-gene association study using an original broad unstructured population. *PLoS ONE*. 2015;10(1):e0116674. <https://doi.org/10.1371/journal.pone.0116674>. [cited 31 May 2022].
- Zhou X, McQuinn R, Fei Z, Wolters AMA, Van Eck J, Brown C, et al. Regulatory control of high levels of carotenoid accumulation in potato tubers. *Plant Cell Environ*. 2011;34(6):1020–30. <https://doi.org/10.1111/j.1365-3040.2011.02301.x>. [cited 31 May 2022].
- Diepenbrock CH, Ilut DC, Magallanes-Lundback M, Kandianis CJ, Lipka AE, Bradbury PJ, et al. Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *Plant Cell*. 2021 [cited 20 Jan 2022];33(4):882–900. Available from: <https://academic.oup.com/plcell/article/33/4/882/6126473>
- Kumar J, Saripalli G, Gahlaut V, Goel N, Meher PK, Mishra KK, et al. Genetics of Fe, Zn,  $\beta$ -carotene, GPC and yield traits in bread wheat (*Triticum aestivum* L.) using multi-locus and multi-traits GWAS. *Euphytica*. 2018;214(11):219. <https://doi.org/10.1007/s10681-018-2284-2>. [cited 15 May 2022].
- Abdel-Aal ES, Akhtar H, Zaheer K, Ali R. Dietary Sources of Lutein and Zeaxanthin Carotenoids and Their Role in Eye Health. *Nutrients*. 2013 [cited 13 Jan 2022];5(4):1169–85. Available from: <http://www.mdpi.com/2072-6643/5/4/1169>
- Blessin CW, VanEtten CH, Wiebe R. Carotenoid content of the grain from yellow endosperm-type sorghums. *Cereal Chem*. 1958;35:359–65.
- Sun W, Hu Y. eQTL Mapping Using RNA-seq Data. *Stat Biosci*. 2013;5(1):198–219. [cited 13 Jan 2022]. <https://doi.org/10.1007/s12561-012-9068-3>
- Suryanarayana Rao R, Rukmini C, Mohan VS.  $\beta$ -Carotene context of some yellow-endosperm varieties of sorghum. *Indian J Agric Sci*. 1967;38(4):368–72.
- Muleta KT, Felderhoff T, Winans N, Walstead R, Charles JR, Armstrong JS, et al. The recent evolutionary rescue of a staple crop depended on over

- half a century of global germplasm exchange. *Evolutionary Biology*; 2022 [cited 15 May 2022]. <https://doi.org/10.1101/2021.05.11.443651>
37. Cuevas HE, Prom LK. Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan Sorghum Core collection. *BMC Genomics*. 2020;21(1):88. <https://doi.org/10.1186/s12864-020-6489-0>. [cited 3 May 2022].
  38. 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*. 2013;110(2):453–8. <https://doi.org/10.1073/pnas.1215985110>. [cited 13 Jan 2022].
  39. Smith C, Frederiksen R. History of Cultivar Development in the United States: From "Memoirs of A.B. Maunder--Sorghum Breeder". Sorghum: Origin, History, Technology, and Production and edited by Wayne Smith C, Frederiksen RA. ISBN: 978-0-471-24237-6. <https://www.wiley.com/enus/Sorghum%3A+Origin%2C+History%2C+Technology%2C+and+Production-p-9780471242376>.
  40. Lebanon and ICARDA. Ties that Bind. ISBN: : 92-9127-225-6. <https://mel.cgiar.org/reporting/download/hash/bDF7mlNh>.
  41. Mengesha MH, Rao KEP. World Sorghum Germplasm Collection And Conservation. In: Plant Genetics and Breeding. Cali, Colombia; 1990 [cited 24 Jul 2019]. p. 90–104. Available from: <http://oar.icrisat.org/5157/>
  42. Prasada Rao KE, Obilana AT, Mengesha MH. Collection of Kaura, Fara-Fara and Guineense sorghums in Northern Nigeria. *J Agric Tradit Bot Appliquée*. 1985 [cited 13 Jan 2022];32(1):73–81. Available from: [https://www.persee.fr/doc/jatba\\_0183-5173\\_1985\\_num\\_32\\_1\\_3930](https://www.persee.fr/doc/jatba_0183-5173_1985_num_32_1_3930)
  43. Gelli M, Konda AR, Liu K, Zhang C, Clemente TE, Holding DR, et al. Validation of QTL mapping and transcriptome profiling for identification of candidate genes associated with nitrogen stress tolerance in sorghum. *BMC Plant Biol*. 2017;17(1):123. <https://doi.org/10.1186/s12870-017-1064-9>. [cited 13 Jan 2022].
  44. Babu R, Rojas NP, Gao S, Yan J, Pixley K. Validation of the effects of molecular marker polymorphisms in LcyE and CrTRB1 on provitamin A concentrations for 26 tropical maize populations. *Theor Appl Genet*. 2013 [cited 27 May 2022];126(2):389–99. <https://doi.org/10.1007/s00122-012-1987-3>
  45. Gebremeskel S, Garcia-Oliveira AL, Menkir A, Adetimirin V, Gedil M. Effectiveness of predictive markers for marker assisted selection of pro-vitamin A carotenoids in medium-late maturing maize (*Zea mays* L.) inbred lines. *J Cereal Sci*. 2018 [cited 27 May 2022];79:27–34. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0733521017304010>
  46. Muthusamy V, Hossain F, Thirunavukkarasu N, Choudhary M, Saha S, Bhat JS, et al. Development of  $\beta$ -Carotene Rich Maize Hybrids through Marker-Assisted Introgression of  $\beta$ -carotene hydroxylase Allele. Parida SK, editor. *PLoS ONE*. 2014;9(12):e113583. <https://doi.org/10.1371/journal.pone.0113583>. [cited 27 May 2022].
  47. Gupta HS, Raman B, Agrawal PK, Mahajan V, Hossain F, Thirunavukkarasu N. Accelerated development of quality protein maize hybrid through marker-assisted introgression of opaque-2 allele. Gupta P, editor. *Plant Breed*. 2013;132(1):77–82. <https://doi.org/10.1111/pbr.12009>. [cited 27 May 2022].
  48. Cobb JN, Biswas PS, Platten JD. Back to the future: revisiting MAS as a tool for modern plant breeding. *Theor Appl Genet*. 2019;132(3):647–67. <https://doi.org/10.1007/s00122-018-3266-4>. [cited 14 June 2022].
  49. Guzman C, Peña RJ, Singh R, Autrique E, Dreisigacker S, Crossa J, et al. Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl Transl Genomics*. 2016 [cited 5 June 2022];1:3–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S221206611630031X>
  50. Plavšin I, Gunjača J, Šatović Z, Šarčević H, Ivić M, Dvojković K, et al. An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. *Plants*. 2021 [cited 5 June 2022];10(4):745. Available from: <https://www.mdpi.com/2223-7747/10/4/745>
  51. Yao J, Zhao D, Chen X, Zhang Y, Wang J. Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). *Crop J*. 2018 [cited 13 Jan 2022];6(4):353–65. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2211451411830062X>
  52. Esuma W, Ozimati A, Kulakow P, Gore MA, Wolfe MD, Nuwamanya E, et al. Effectiveness of genomic selection for improving provitamin A carotenoid content and associated traits in cassava. Holland JB, editor. *G3 GenesGenomesGenetics*. 2021;11(9):jkab160. <https://doi.org/10.1093/g3journal/jkab160/6272516> [cited 26 May 2022].
  53. 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 [cited 13 Jan 2022];198(4):1699–716. Available from: <https://academic.oup.com/genetics/article/198/4/1699/5935899>
  54. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017 [cited 10 June 2022];169(7):1177–86. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867417306293>
  55. Muleta KT, Pressoir G, Morris GP. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. *G3 GenesGenomesGenetics*. 2019 [cited 13 Jan 2022];9(2):391–401. Available from: <https://academic.oup.com/g3journal/article/9/2/391/6026675>
  56. Bernardo R. Upgrading a maize breeding program via two-cycle genomewide selection: Same cost, same or less time, and larger gains. *Crop Sci*. 2021;61(4):2444–55. <https://doi.org/10.1002/csc2.20516>. [cited 5 June 2022].
  57. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Sci*. 2010;50(5):1681–90. <https://doi.org/10.2135/cropsci2009.11.0662>. [cited 5 June 2022].
  58. Atlin GN, Cairns JE, Das B. Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Glob Food Secur*. 2017 [cited 13 Jan 2022];12:31–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2211912416300931>
  59. Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, et al. Community Resources and Strategies for Association Mapping in Sorghum. *Crop Sci*. 2008;48(1):30–40. <https://doi.org/10.2135/cropsci2007.02.0080>. [cited 13 Jan 2022].
  60. Salas Fernandez MG, Kapran I, Souley S, Abdou M, Maiga IH, Acharya CB, et al. Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. *Genet Resour Crop Evol*. 2009;56(7):991–1000. <https://doi.org/10.1007/s10722-009-9417-3>. [cited 13 Jan 2022].
  61. Irakli MN, Samanidou VF, Katsantonis DN, Biliaderis CG, Papadoyannis IN. Phytochemical profiles and antioxidant capacity of pigmented and non-pigmented genotypes of rice (*Oryza sativa* L.). *Cereal Res Commun*. 2016;44(1):98–110. <https://doi.org/10.1556/0806.43.2015.033>. [cited 13 Jan 2022].
  62. Butler DG, Cullis BR, Gilmour AR, Gogel BG, Thompson R. ASReml-R Reference Manual Version 4. Hemel Hempstead, HP1 1ES, UK: VSN International Ltd; 2017.
  63. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;28(18):2397–9. <https://doi.org/10.1093/bioinformatics/bts444>. [cited 17 May 2022].
  64. Hu Z, Olatoye MO, Marla S, Morris GP. An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. *Plant Genome*. 2019;12(1):180044. <https://doi.org/10.3835/plantgenome2018.06.0044>. [cited 22 Apr 2022].
  65. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*. 2011;4(3):250–5. <https://doi.org/10.3835/plantgenome2011.08.0024>. [cited 13 Jan 2022].
  66. Monier B, Casstevens TM, Bradbury PJ, Buckler ES. rTASSEL: an R interface to TASSEL for association mapping of complex traits. *Bioinformatics*. 2020. <https://doi.org/10.1101/2020.07.21.209114>. [cited 17 May 2022].
  67. Cuevas HE, Rosa-Valentin G, Hayes CM, Rooney WL, Hoffmann L. Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics*. 2017;18(1):108. <https://doi.org/10.1186/s12864-016-3475-7>. [cited 3 May 2022].
  68. Maina F, Bouchet S, Marla SR, Hu Z, Wang J, Mamadou A, et al. Population genomics of sorghum (*Sorghum bicolor*) across diverse agroclimatic zones of Niger. Ungerer M, editor. *Genome*. 2018;61(4):223–32. <https://doi.org/10.1139/gen-2017-0131>. [cited 3 May 2022].
  69. 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(6):e1400218. <https://doi.org/10.1126/sciadv.1400218>. [cited 3 May 2022].

70. Olatoye MO, Hu Z, Maina F, Morris GP. Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. *G3 GenesGenomesGenetics*. 2018 [cited 3 May 2022];8(10):3269–81. Available from: <https://academic.oup.com/g3journal/article/8/10/3269/6026990>
71. Faye JM, Maina F, Hu Z, Fonckea D, Cisse N, Morris GP. Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. *Ecol Evol*. 2019;9(10):6038–51. <https://doi.org/10.1002/ece3.5187>. [cited 3 May 2022].
72. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet*. 2018 [cited 17 May 2022];103(3):338–48. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002929718302428>
73. Dekkers JCM. Prediction of response to marker-assisted and genomic selection using selection index theory: selection index theory for genomic selection. *J Anim Breed Genet*. 2007;124(6):331–41. <https://doi.org/10.1111/j.1439-0388.2007.00701.x>. [cited 4 May 2022].

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

