Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Toward marker-assisted breeding for bioaccessible provitamin A in sorghum: genetic and environmental basis of carotenoid bioaccessibility

Rae McDowell
Colorado State University
Álvaro Cruz-Carrión
University of Arkansas for Medical Sciences, Arkansas Children's Nutrition Center
Chelsey Fiecke
University of Arkansas for Medical Sciences, Arkansas Children's Nutrition Center
Michael P. Dzakovich
Baylor College of Medicine
Richard E. Boyles
Clemson University
Hugo Cuevas
USDA-ARS, Tropical Agriculture Research Station
Terry Felderhoff
Kansas State University
Chad Hayes
USDA-ARS
Elizabeth A. Cooper
UNC Charlotte
John K. McKay
Colorado State University
Jacqueline M Chaparro
Colorado State University
Jessica E. Prenni
Colorado State University
Geoffrey P. Morris
Colorado State University
Mario Ferruzzi
University of Arkansas for Medical Sciences, Arkansas Children's Nutrition Center
Davina H. Rhodes
davina.rhodes@colostate.edu

Colorado State University

Research Article

Keywords: sorghum, carotenoid, bioaccessibility, QTL, GxE, MAS

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11459018/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Toward marker-assisted breeding for bioaccessible provitamin A in sorghum: genetic and environmental basis of carotenoid bioaccessibility

Rae McDowell[1], Álvaro Cruz-Carrión[2], Chelsey Fiecke[2], Michael P. Dzakovich[3], Richard E. Boyles[4], Hugo Cuevas[5], Terry Felderhoff [6], Chad Hayes[7], Elizabeth A. Cooper[8], John K. McKay[9], Jacqueline M Chaparro[1], Jessica E. Prenni[1], Geoffrey P. Morris[9], Mario Ferruzzi[2,10], and Davina H. Rhodes[1]*

[1]Cell and Molecular Biology Program, Department of Horticulture & Landscape Architecture, Colorado State University, Fort Collins, CO 80523, USA
[2]University of Arkansas for Medical Sciences, Department of Pediatrics; Arkansas Children's Nutrition Center, Little Rock, AR 72202, USA
[3]Department of Pediatrics, Baylor College of Medicine, USDA-ARS Children's Nutrition Research Center, Houston, TX, United States, USA
[4]Department of Plant and Environmental Sciences, Clemson University, Clemson, South Carolina, USA
[5]USDA-ARS, Tropical Agriculture Research Station, 2200 Pedro Albizu Campos Avenue, Mayaguez, Puerto Rico 00680, USA
[6]Department of Agronomy, Kansas State University, Manhattan, KS, 66506, USA
[7]USDA-ARS, Plant Stress & Germplasm Development Research Unit, Cropping Systems Research Laboratory, Lubbock, TX, 79415, USA
[8]Department of Bioinformatics and Genomics, UNC Charlotte, Kannapolis, NC 28081
[9]Department of Soil & Crop Science, Colorado State University, Fort Collins, CO 80523, USA
[10]College of Agriculture and Life Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA

*Corresponding author (davina.rhodes@colostate.edu)

ABBREVIATIONS

MAS: Marker-Assisted Selection, QTL: Quantitative Trait Loci, ICP-MS: Inductively Coupled Plasma Mass Spectrometry, GAE: Gallic Acid Equivalents; GxE: Genotype by Environment interaction, IPCA = Interaction Principal Component Axis, KASP: Kompetitive Allele Specific PCR, PVA: Provitamin A, PVAC: Provitamin A Carotenoids, RB: Relative Bioaccessibility, BC: Bioaccessible Content, LOD: Logarithm of the Odds, LOQ: Limit of Quantification, TKW: Thousand Kernel Weight, SNP: Single Nucleotide Polymorphism, RIL: Recombinant Inbred Line

Key Words: sorghum, carotenoid, bioaccessibility, QTL, GxE, MAS

Key Message

In this first diversity and QTL mapping study of carotenoid bioaccessibility in sorghum, novel QTL predicted carotenoid content and bioaccessibility, supporting marker-assisted selection to help address high genotype-by-environment variation.

ABSTRACT

Vitamin A deficiency is a critical global health issue, particularly in regions reliant on sorghum, a staple cereal with low provitamin A carotenoid (PVAC) content. Enhancing bioaccessible PVAC in sorghum grain is essential for effective biofortification, yet the genetic basis and environmental influences on carotenoid bioaccessibility are not well characterized. We used an *in vitro* digestion assay, genomic mapping, and multi-environment trials on diverse sorghum germplasm and a biparental family to quantify genetic and environmental influences on carotenoid bioaccessibility. We discovered substantial genetic variation for bioaccessible carotenoid traits, with heritability that varied by trait and year; notably, the ratio of α-branch to β-branch carotenoids showed the highest and most stable heritability across environments and generations. Bioaccessibility quantitative trait loci (QTL) colocalized with carotenoid content QTL, and phytates, phenolics, and mineral content showed environment-dependent associations with reduced bioaccessibility. Markers linked to these QTL explained up to 67% of the variation in α- to β-branch carotenoid ratios across diverse germplasm and environments, supporting marker-assisted selection strategies that target metabolic flux toward β-branch carotenoids to improve bioaccessible provitamin A in sorghum, offering a promising approach to reduce vitamin A deficiency in vulnerable populations.

PLAIN LANGUAGE SUMMARY

Vitamin A deficiency leads to serious health problems, including vision loss, decreased immunity, and even death. Sorghum, a grain eaten by millions of people in Africa, naturally has very little provitamin A, the compounds our bodies turn into vitamin A. Plant breeding can increase the provitamin A in sorghum grain, but it has been unclear how much of it the body can absorb and use after digestion. This study tested 128 sorghum varieties from around the world and found that they differ in how much provitamin A remains after digestion, and that these differences were linked to a few specific regions of the sorghum genome in a related family. Availability was also lower, in some environments, when the grain contained higher levels of certain minerals, plant defense compounds, and starch-related traits that can interfere with absorption. Detecting the genetic regions that improve provitamin A availability can help plant breeders develop sorghum varieties that deliver more usable provitamin A to the people who eat sorghum.

1 INTRODUCTION

Vitamin A deficiency is one of the most prevalent global nutrition problems, leading to vision loss, impaired immunity, stunting, and death, particularly in young children and pregnant or lactating women (Balmer and Blomhoff 2002; World Health Organization et al. 2006). Sub-Saharan Africa has the highest prevalence of vitamin A deficiency and relies heavily on the staple cereal sorghum (*Sorghum bicolor* [L.] Moench) (Zhao et al. 2022). Although sorghum is a highly diverse and resilient staple crop (Prasad et al. 2021), it typically accumulates low concentrations of provitamin A carotenoids (PVAC), the dietary precursors that the human body converts to vitamin A. Biofortification aims to reduce vitamin A deficiencies by increasing PVAC levels in sorghum grain.

Biofortification is a strategy used to combat nutrient deficiencies by improving the nutritional quality of staple crops. Although food fortification has helped reduce micronutrient deficiencies in some populations through the addition of nutrients during food processing (Institute of Medicine 2003), crop biofortification offers a cost-effective and sustainable solution, particularly in rural, resource-limited areas (Garg et al. 2018). Cereal crops are key biofortification targets due to their widespread consumption, long shelf life, and typically low levels of micronutrients. Efforts have focused on increasing β-carotene, a provitamin A carotenoid, with target concentrations typically ranging from 9 to 15 μg/g (Cruet-Burgos et al. 2023; HarvestPlus 2015). Biofortified crops such as orange fleshed sweet potato, grown in six million households in Africa and Asia (CGIAR, n.d.; van Jaarsveld et al. 2005), and high β-carotene maize developed by the International Maize and Wheat Improvement Center (CIMMYT) have demonstrated efficacy in improving vitamin A status in target populations (Muzhingi et al. 2011).

The nutrient target value estimates the concentration of a nutrient in a crop needed to meet dietary intake recommendations, taking into account factors such as crop consumption levels, other dietary sources, conversion efficiency of nutrient precursors, and nutrient bioaccessibility. Bioaccessibility is the fraction of a nutrient available for intestinal absorption after digestion, and is influenced by factors including health of the individual and composition of the food matrix. Certain plant compounds, such as phytates and phenolics, can inhibit digestive enzymes such as amylases and lipases (Dzakovich et al. 2023; Samtiya et al. 2020; Sreerama et al. 2010), thereby decreasing the bioaccessibility of essential nutrients (Cominelli et al. 2020). In sorghum, the endosperm makes up 80% of the grain mass, is rich in starches and protein bodies, and interactions within the endosperm may inhibit carotenoid bioaccessibility by making the grain less digestible (Duodu et al. 2003; Rooney and Pflugfelder 1986; Singh and Axtell 1973; Taylor and Duodu 2023; Wong et al. 2009). For example, a study in bananas found that the amount of soluble and resistant starch affected carotenoid bioaccessibility (Munoz et al. 2024). Additionally, minerals such as iron and zinc can decrease carotenoid bioaccessibility by catalyzing lipid peroxidation and forming lipid soap complexes that hinder absorption (Biehler et al. 2011). In sorghum grain, however, the effects of the food matrix on carotenoid bioaccessibility are not well understood.

Genetic tools are increasingly used to improve biofortification breeding (Cruet-Burgos et al. 2023; Liu et al. 2025), enabling low-cost, high-throughput germplasm screening (Hasan et al. 2021; Mohler and Singrün 2005). In sorghum, genetic variants affecting carotenoid biosynthesis, degradation, and precursor pathways have been converted into Kompetitive Allele Specific PCR (KASP) markers for selecting for high-provitamin A carotenoid germplasm (Cruet-Burgos et al. 2026). Notably, markers within and near the *zeaxanthin epoxidase* (*ZEP*) gene have been developed and validated in diverse germplasm in the Senegal national sorghum breeding program (Cruet-Burgos et al. 2026; Banda et al. 2026)**.** However, to maximize the effectiveness of biofortification, it may be necessary to select for both carotenoid bioaccessibility and concentration. Currently, no carotenoid bioaccessibility markers are available for sorghum or other crops, and the extent and genetic basis of variation in carotenoid bioaccessibility remain unknown.

In this study, we assessed the potential for breeding sorghum with enhanced bioaccessible PVAC by using an *in vitro* digestion method (Fig. 1a) (Dzakovich et al. 2023) to quantify both bioaccessible content and relative bioaccessibility of sorghum grain carotenoids (Fig. 1b). We analyzed a diverse panel representing global sorghum genetic diversity (Fig. 1c) and observed substantial variation in carotenoid bioaccessibility traits (Fig. 1d-i). Building on initial findings, we 1) evaluated a subset of accessions in multi-location trials to assess genotype by environment (GxE) interaction effects on carotenoid bioaccessibility; 2) developed a biparental family from medium- and high-carotenoid parental lines to identify putative QTL underlying these traits; and 3) quantified potential inhibitor compounds to gain insights into factors contributing to reduced carotenoid bioaccessibility. Together, these findings provide guidance on breeding strategies aimed at increasing bioaccessible PVAC in sorghum.

2 MATERIALS AND METHODS

2.1 Germplasm

2.1.1 Germplasm for Initial Evaluation of Diversity. To assess variation in carotenoid bioaccessibility in sorghum, 128 accessions from the Sorghum Association Panel (SAP) and carotenoid panel, hereafter referred to as the diversity panel (Casa et al. 2008; Cruet-Burgos et al. 2020, 2023), were grown in 2019 at the Kansas State University Agronomy North Farm in Manhattan, Kansas using a randomized complete block design with two replications. Grain was harvested at maturity, dried, and stored at -80 °C until further use. For analysis, two biological replicates from separate field plots were used to capture intergenotypic variation and the samples were extracted and analyzed in a randomized complete block design.

2.1.2 Germplasm for GxE Analysis. Twelve accessions (PI533877, PI563450, PI563455, PI585347, PI563068, PI563453, PI569812, PI585348, PI563409, PI563454, PI576426, PI585369) from the diversity panel were grown in four locations in the 2022 and 2023 field seasons in three replicate plots per accession per location to represent biological replicates. The four locations were semi-arid (Lubbock, TX, USDA - ARS), humid continental (Manhattan, KS, Kansas State University), humid subtropical (Florence, SC, Clemson University), and tropical

(Mayaguez, PR, USDA - ARS) environments. Panicles at all locations were grown uncovered with full exposure to sunlight. Grain was harvested at maturity, dried, and shipped to Colorado State University in Fort Collins, CO and stored at -80 °C until analysis. Year 1 (2022) samples from PR had a disproportionate amount of green seeds, indicating the presence of chlorophyll, despite having a black layer that signifies grain maturity. Since carotenoid concentrations are higher in grain that is not fully mature (Kean et al. 2007), and chlorophyll, which is usually present in maturing grain, is a fat-soluble pigment that can interfere with carotenoid quantification assays, the data from 2022 PR was excluded from the main analysis. Climatic descriptions of environments can be found in Table S1. Three biological replicates per genotype, per location, and per year were phenotyped.

2.1.3 Germplasm for QTL Analysis. To identify loci underlying variation in carotenoids and carotenoid bioaccessibility, a biparental family segregating for medium and high grain carotenoids was developed. PI655996 (RTx430, medium carotenoid, Female) and PI585348 (high carotenoid, Male) were crossed in 2019 in Manhattan, KS and advanced in Puerto Vallarta, Mexico to the F5:6 generation in 2022 and F6:7 generation in 2023. Seeds were shipped to Colorado State University in Fort Collins, CO and stored at -20 °C for later analysis. For phenotyping, 3 technical replicates of bulked grain per genotype per year were used. The original population consisted of 229 Recombinant Inbred Lines (RILs) and the two parental lines. Six genotypes failed to be genotyped and were removed from analysis, and several genotypes failed to germinate in the field were removed from both years. Genomic quality control identified individuals that were genetically inconsistent with membership in the biparental family. Individuals were evaluated with several criteria including residual heterozygosity, genotype-error recombination patterns, and kinship analysis. Genotypes that displayed characteristics inconsistent with the expected biparental family were excluded according to the filtering criteria as described in Supplemental Methods 1. Because genetically divergent individuals can lead to spurious identification of false-positive QTL during mapping, we filtered conservatively, excluding individuals flagged by multiple genomic QC metrics that indicated inconsistencies in the expected structure of the biparental family (Data S1). Resulting data of 140 RILs and 2 parental lines have the expected distribution on a PCA where RILs are distributed between the two parents as well as the expected kinship matrix (Fig. SM5 , Fig. 3a). For the second year only 127 lines of the 140 were analyzed.

2.2 Grain phenotyping

2.2.1 Sample Porridge Preparation for Bioaccessibility and Nutrient Analysis. For analysis of carotenoid, polyphenol, and phytate content, as well as carotenoid bioaccessibility, grain samples were shipped on dry ice to Arkansas Children's Nutrition Center (Little Rock, AR) and stored at -80 °C until preparation. The grain was first cleaned of excess plant material and 5 g was placed in 15 mL polycarbonate vials (OPS Diagnostics) with 2 large stainless steel beads. The samples were then pre-frozen at -80 °C to prevent overheating from grinding. Samples were ground using a 2010 Geno/Grinder® (SPEX Sample Prep) for 2.5 min at 1400 rpm × 2 cycle with 60 sec rest between cycles to prevent samples from overheating. The flour was then transferred

into 15 ml tubes and blanketed with N_2 gas, sealed with a cap and parafilm, and frozen at -80 °C until further analysis. Samples for raw flour carotenoid extraction (300 mg) were weighed in 2 mL microcentrifuge tubes, and samples for digestion in the bioaccessibility assay (400 mg) were weighed in 15 mL capped tubes. The prepared samples were then blanketed with N_2 and sealed with parafilm and stored at -80 °C until analysis.

2.2.2 Bioaccessibility Assay. Porridge was prepared and digested according to Dzakovich et al. (2023) in triplicate. Ground sorghum flour was mixed with double distilled water in tubes, vortexed until homogenous, and gelatinized by placing it in 100 °C water for 10 minutes. Samples were removed and 100 mg of canola oil was added and mixed in manually with wooden coffee stirrers. Resulting sorghum porridge samples were then immediately digested. An oral phase solution containing α-amylase was added to porridge samples, incubated, and slowly shaken at body temperature (37 °C) for 10 minutes. A gastric phase solution of saline and pepsin was added to the samples, then adjusted to pH 2.5 with 1 M HCl, followed by blanketing samples in N_2, and finally incubated and gently shaken at 37 °C for an hour. Samples were adjusted to pH 5.0 with 1 M $NaHCO_3$, and then a small intestinal phase solution of pancreatin-lipase and bile were added. Samples were then adjusted to pH 7.0 with 1 M $NaHCO_3$. aline solution was added to the samples, and then samples were blanketed with N_2, incubated, and shaken slowly at 37 °C for two hours. Aliquots of digesta (3 mL each) were removed for carotenoid analysis. Digesta fractions were then blanketed with N_2 and stored at -80 °C for future analysis. Aqueous fractions of digesta were collected by centrifuging the remaining samples for 80 minutes at 3,230 x g at 4 °C, followed by filtering 7 mL of the supernatant using 50 mL centrifuge filter tubes (0.22 µm, cellulose acetate) for 15 minutes at 3,230 x g at 4 °C. Aliquots of aqueous fractions (6 mL each) were then blanketed with N_2, sealed with parafilm, and stored at -80 °C until further analysis.

2.2.3 Carotenoid Extraction and Quantification. Carotenoids in raw flour samples were extracted at the Arkansas Children's Nutrition Center according to the method of Dzakovich et al., (2023) in triplicate using a Tecan Freedom EVO 150 liquid handling system (Tecan; Mannedorf, Switzerland). Briefly, 500 mg flour samples and five to ten 2.38 mm stainless steel balls (Qiagen; 13117) in 15 mL tubes were rehydrated with 250 µL LC-grade water, then spiked with internal standard (30 µL of 150 µM retinyl palmitate in ethyl acetate:methanol [1:1 *v/v*]). Carotenoid-biofortified corn grits (Professor Torbert's Orange Corn Grits) were used as a positive control. 1 mL of acetone: ethyl acetate (3:2, *v/v*) containing 0.01% BHT (*w/v*) was added, samples were vortexed (1400 RPM, 45 seconds) and centrifuged (3000 RPM, 10 minutes, 4 °C), and supernatants were transferred to borosilicate glass tubes. This extraction was repeated twice more using 1 mL methyl *tert*-butyl ether + 0.01% butylated hydroxytoluene (BHT) (*w/v*). Combined extracts were dried under N_2, redissolved in methanol:ethyl acetate (1:1, *v/v*) by reconstituting in 1 mL ethyl acetate, vortexing, adding 1 mL methanol, and vortexing again. Finally, samples were filtered using PTFE syringe filters (0.45 µm) and transferred to amber LC vials.

Carotenoids in digested porridge (aqueous and digesta fractions) were extracted according to the method of Dzakovich et al., (2023) at the Arkansas Children's Nutrition Center. Briefly, aqueous fractions (micellar fraction) were used to quantify bioaccessibility by spiking 6 mL of sample with internal standard (30 µL of 7.5 µM retinyl palmitate in ethyl acetate:methanol [1:1 *v/v*]). Samples were extracted with 6 mL of acetone: petroleum ether (1:3, *v/v*) containing 0.01% BHT (*w/v*), vortexed for 2 minutes at 2500 RPM, and centrifuged for 2 minutes at 2500RPM at 4 °C, and the organic layer was transferred into a glass tube. This extraction was repeated two more times and the organic phases were combined. Combined extracts were dried under N_2 ; tubes were rinsed with 1.0 mL methyl *tert*-butyl ether to recover residual analytes and dried again, then redissolved in 100 µL of ethyl acetate:methanol (1:3, *v/v*) containing 0.01% BHT (*w/v*), sonicated in a water bath for 1 minute, and filtered through 0.22 µm cellulose acetate syringe filters. Filtered samples were transferred to amber LC vials. All carotenoid extractions were performed under low light to minimize photooxidation.A Waters UPLC Acquity H Class system equipped with photodiode array detector was used to quantify the initial carotenoid concentration in the raw flour as well as the aqueous fraction of the digesta (Dzakovich et al., 2023) and then relative bioaccessibility and bioaccessible content were calculated using the following equations:

$$\text{Relative Bioaccessibility } (\%) = \frac{\text{Carotenoid concentration in aqueous fraction of digesta } (\mu g/g)}{\text{Carotenoid concentration in raw flour } (\mu g/g)} \times 100 \quad (1)$$

$$\text{Bioaccessible Content} = \text{Carotenoid concentration in aqueous fraction of digesta } (\mu g/2g) \times 100 \quad (2)$$

2.2.5 Thousand Kernel Weight (TKW). Two hundred whole seeds per sample were counted with a seed counter, dried in an oven for 5 days at 63 ℃ to ensure equal moisture content, and weighed on a scale. TKW was calculated by multiplying the mass of 200 seeds by five. This dataset can be found in Supplemental Data 1 and 2.

2.2.6 Total Phenolics. Total phenolic compounds were quantified in sorghum flour using the Folin-Ciocalteu method (Cruz-Carrión et al. 2023) adapted to a 96-well format. Samples were analyzed in triplicate, and biofortified corn grits (Professor Torbert's Orange Corn Grits) were used as a positive control. Phenolics were extracted according to Li et al. (2020) with minor modifications. Sorghum flour (500 mg) was mixed with 1.5 mL of 2% formic acid in 80% methanol (*v/v*), homogenized in Geno/Grinder® 2010 (SPEX Sample Prep, Metuchen, NJ, USA) at 1750 rpm for 2 min, then centrifuged (4000 rpm for 4 min at 10 ℃) and the supernatant collected (1.25 mL aliquot transferred). The pellet was extracted twice more with 2% formic acid in 98% methanol (*v/v*). Pooled extracts were dried under N_2 gas for 1.5 h and reconstituted in 4 mL 0.1% formic acid in water. Reconstituted samples were vortexed (30 sec, 1500 rpm) and sonicated (2 min) twice. For the microplate assay, 1 mL of sample was transferred to a 96-well collection plate (2 mL, Waters). In a 96-well microplate (350 µL, Waters), 70 µL of water was dispensed into each well, followed by 30 µL of sample extract, water (blank), or gallic acid

standards. After mixing, 50 µL of Folin & Ciocalteu's phenol reagent were added and plates were incubated in the dark for 10 min at room temperature. Then 100 µL of 25% (*w/v*) Na_2CO_3 were added, mixed, and incubated in the dark for 60 min at room temperature. Absorbance was read at 725 nm using a POLARstar Omega Microplate Reader (Cary, NC, USA). Total phenolics were calculated from a gallic acid calibration curve and expressed as mg gallic acid equivalents (GAE) per g of dry sorghum flour. This dataset can be found in Supplemental Data 1 and 2.

2.2.7 Phytic Acid. Phytic acid content was analyzed according to the method of Gao et al. (2007), with minor modifications. Samples were analyzed in triplicate, and biofortified corn grits (Professor Torbert's Orange Corn Grits) were used as a positive control. Sorghum flour or corn grits (250 mg) was vortexed with 5 mL of 2.4% (0.78 N) HCl (2500 rpm, 2 min), and then shaken at 220 rpm for 16 h in an incubator shaker (New Brunswick Scientific I2400) at room temperature. Samples were then centrifuged for 20 min at 1000 x g and 10 °C. The crude acid extract was transferred to 15 mL centrifuge tubes containing 500 mg NaCl, vortexed for 2 minutes at 2500 rpm, shaken at 350 rpm for 20 min and allowed to settle at -20° C for 20 min. Samples were then centrifuged for 20 min at 1000 × g and 10 °C, and the clear supernatants (500 µL) were collected for color development. Supernatants were diluted 25-fold by adding 12 mL of distilled water, then vortexed at 2500 rpm for 1 minute. Diluted samples or standards (1.5 mL) were combined with 0.5 mL modified Wade reagent [0.03% (*w/v*) $FeCl_3 \cdot 6H_2O$ and 0.3% (*w/v*) sulfosalicylic acid], and vortexed (2500 rpm, 1 min). Sodium phytate standards at 0, 5, 10, 20, 30, 40, 50, 60, 80, and 100 mg/L were prepared in the same matrix. Prior to reading, aliquots were transferred to a 96-well plate (300 µL per well) and absorbance of color reaction products was read at 500 nm on a POLARstar Omega Microplate Reader (Cary, NC, USA) against reagent blanks. Phytic acid content was calculated following the method of Latta and Eskin (1980) and expressed as mg phytic acid equivalents per g of dry sorghum flour. Data can be found in Supplemental Data 1 and 2.

2.2.8 ICP-MS Mineral Analysis. Grain minerals were quantified by ICP-MS using three biological replicates per genotype across environments. A single sample was included across batches as a technical control. Grain (2 g) was frozen (-80 °C), lyophilized (-50 °C, 1 week), and stored at -80 °C until analysis. For digestion, 250 mg whole grain was weighed into 75-mL teflon microwave vessels. Digestion was achieved by adding 10 mL of doubly distilled concentrated nitric acid, spiked with internal standard (Indium, 565 ppm). Each vessel was allowed to react for 10 min prior to sealing. The sealed vessels were placed in a programmable MPS 320 microwave digestion system (PerkinElmer). The digestion program was set as follows: Step 1 - temperature increased to 200 °C at a ramp rate of 15 °C min^{-1}, pressure at 100%, and held for 30 min; Step 2 - temperature decreased to 50 °C, pressure reduced to 0%, and held for 15 min. After digestion samples were diluted to a total volume of 30 mL with Milli-Q water. Subsequently, 1.2 mL of each sample was further diluted to a total volume of 15 mL. Samples and batches were run in randomized fashion and each batch contained a blank sample. For quality control across ICP-MS batches, 1 mL of each sample was pooled together and sampled repeatedly across batches along with a pooled blank. Elemental concentrations of aluminum,

arsenic, boron, barium, beryllium, calcium, cadmium, copper, chromium, iron, potassium, lithium, magnesium, manganese, molybdenum, sodium, nickel, phosphorus, lead, sulfur, selenium, strontium, vanadium, tungsten, and zinc were measured using NexION 2000 mass spectrometer (PerkinElmer). Samples were measured in randomized order with analysis of the pooled QC after every 9 samples. A 7-point calibration curve was prepared by serial dilution of commercially available single-element standard stock solutions (Inorganic Ventures). These were matrix matched to the samples, 2.5% HNO_3 and 20 ppb indium. For correction of instrument drift, internal standard solutions consisting of lithium, rhodium, and iridium were added to each sample via the autosampler. Each element was subjected to internal standard corrections and subsequently drift corrected (Broccardo et al. 2013). Corrections were chosen based on minimizing the CV for the QC samples. After drift correction, samples were corrected for the dilution factor. Limits of detection and limits of quantification (LOQ) were calculated as 3 times or 10 times the SD of the blank divided by the slope of the calibration curve, respectively (Broccardo et al. 2013). Final concentrations are given in micrograms per gram of freeze-dried sorghum grain. Values below the LOQ were assigned a value of LOQ/2 (Becker et al. 2002). This dataset can be found in Supplemental Data 2.

2.3 Statistical Analysis

Carotenoid content, carotenoid bioaccessibility, and potential inhibitors were quantified in three biological subsamples from three replicate plots per location to capture inter-environment biological variation. Broad-sense heritability (Eq 3) was calculated using variance components of genotype (G), genotype by environment interaction (GxE), environment (E), and error (ε) derived from a linear regression model implemented in the R lme4 package (Bates et al. 2015) with all factors as random effects, as exemplified below:

```
rand_mod <- lmer(LuteinRelbio ~ (1|Genotype) + (1|Environment) + (1|Genotype:Environment),
data = all)
```

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G\times E}^2 + \sigma_E^2 + \sigma_\varepsilon^2} \quad (3)$$

Breeder's heritability (Eq. 4) (Holland et al. 2002) (heritability with removal of the variance contributions of environmental effects) was calculated using variance components of genotype (G; n = 12), genotype by environment interaction (GxE; n = 36, 46), environment (E; n = 3, 4), Replicates (R; n=3) and error (ε) derived from a linear regression model in RStudio (v4.3) using the lme4 package with all factors as random effects. Genotypes were removed from analysis if there were missing samples that caused singularities, resulting in 100 datapoints for 2022 and 139 for 2023. Biological replicates of accessions grown in separate plots were analyzed. Grain mineral content was also analyzed and visualized in SIMCA to assess potential

outliers before further data analysis. Variation in mean differences between environments and phenotypes was tested by ANOVA followed by Tukey's HSD comparisons ($\alpha = 0.05$). For carotenoids and minerals, values below the instrument's limit of detection (3 times the standard deviation of the blank divided by the slope of the calibration curve) were set to (Limits of Detection) /2, so that these samples could be kept in the analysis rather than exclude or set to zero.

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G\times E}^2}{E} + \frac{\sigma_\varepsilon^2}{G\times R}} \quad (4)$$

Broad sense heritability (Eq. 5) was calculated for the biparental family. σ_G^2 is the variance attributed to genotype/ RIL as a random effect and σ_ε^2 is the variance attributed to error. Variance terms were calculated from a random regression model in lme4.

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_\varepsilon^2} \quad (5)$$

AMMI analysis was conducted in R Studio (v4.3) with 2022 and 2023 GxE data across environments, with each year and location combination treated as a unique environment. Genotype and environment main effects were removed from a calculated means matrix to isolate and obtain the genotype-by-environment interaction matrix. Genotype and environment scores are derived from the singular vectors and singular values produced by SVD (Singular Value Decomposition) matrix, the SVD was applied to the interaction matrix to identify the ICPAs (Interaction principal components), and their scores were plotted in a biplot to assess GxE patterns. Means of carotenoid tertile labels were calculated using both years of data. Year-location combinations were treated as 7 environments to further decompose GxE variance components identified in the heritability calculations (Table 1) to understand non-genotypic driven interactions across environments. Pearson's correlations were calculated with biological replicates across and within environments (Data S2).

2.4 Genotyping and QTL mapping

To map QTL for carotenoid bioaccessibility, a biparental family was genotyped and linkage maps constructed. The $F_{5:6}$ family was grown in a greenhouse at Colorado State University (Fort Collins, CO) in summer 2023. Two leaf punches per plant were collected in 96-well plates and genotyped by AgriPlex Genomics (Cleveland, OH) using the PlexSeq™ next-generation sequencing (NGS) platform (Kumar et al. 2026). AgriPlex analysis yielded 2421 markers. Genotypes were converted to an ABH file in TASSEL version 5 (Bradbury et al. 2007) and monomorphic and missing markers were filtered out. The remaining 639 markers spanned the 10 chromosomes with 92, 72, 96, 49, 88, 69, 59, 42, 52, and 20 markers per chromosome respectively with less than 5% missing genotype calls.

Sequencing revealed contamination from samples that were not a part of the biparental family, indicated by a high proportion of heterozygosity, which is unexpected in recombinant inbred lines. TASSEL version 5 was used to calculate the proportion of heterozygosity for each individual recombinant inbred line. Samples that contained unusually high levels of

heterozygosity or an unusually high number of recombination events were removed from analysis; this process was documented in supplemental methods file 1 and were visualized with ZHANG kinships in GAPIT (Fig. 3) (Lipka et al. 2012; Wang and Zhang 2020). In total, 83 lines were removed, leaving 140 for QTL mapping for year 1 and 127 for year 2. QTL mapping was done in RStudio (v4.3) using *R/QTL* (Broman et al. 2003) with lines defined as inbreds. Interval mapping was used to identify putative QTL (referred to as QTL). The location of a QTL was determined according to its logarithm of the odds (LOD) peak location. A hidden Markov model was used with *calc.genoprob* to calculate the probabilities of true genotypes for the cross object at marker positions (step=0) with a genotyping error probability(error.prob) of 0.001 which represents the Pr(observed genotype | true genotype) to generate a file for QTL mapping. The significance threshold for linkage was determined using Haley-Knott regression (Haley and Knott 1992) with 1000 permutations. Confidence intervals were calculated using *bayesint* in *R/QTL* to get a bayesian credible interval for a QTL with a 95% probability and *lodint* was used to get a support interval for QTL. Details on QTL and their confidence intervals can be found in Supplemental Data 1. Linkage disequilibrium (LD) R^2 was calculated for markers using TASSEL v.5 in global sorghum germplasm (Hu et al. 2019), relevant subsets can be found in Data S1. Over 180 *a priori* candidates were identified for carotenoid and bioaccessibility traits (Supplementary Data 1) from published literature. Candidates were also chosen from hypothesized interactions affecting bioaccessibility, including: starch or carbohydrate metabolism, endosperm development, endosperm texture, kafirin production and sequestration, lipid metabolism, phenolics production, carotenoid metabolism, heat and stress response metabolism and metabolism and regulation of plastids. Positions for candidates were found using Phytozome (Goodstein et al. 2012) with the Tx430 v3.1.1 reference genome.

2.5 QTL Marker Regression Models

SNP markers identified as QTL in mapping analysis from the biparental family were used to predict variation in the diversity panel and GxE experiment phenotypes. To control for false positives, we conducted a permutation test of 100 random markers to calculate an empirical 95% significance threshold (Data S1). The marker genotypes were then converted into factors in a fixed effect regression model, as shown below:

$$\text{model} = \text{lm}(\text{Trait} \sim \text{Marker},\ \text{data} = \text{data}) \quad (6)$$

Data for permutation tests, models, and model outputs can be found in Data S3. SNP data was derived from (Hu et al. 2019) and principal components were derived using Genome Association and Prediction Integrated Tool (GAPIT) version 3 in R (Lipka et al. 2012).

3 RESULTS

3.1 Substantial genetic variation exists for grain bioaccessibility in sorghum

We evaluated 128 diverse accessions and observed a significant range of bioaccessible content and relative bioaccessibility for β-carotene (0.07–11.7 μg/200 g, 3–31%), zeaxanthin

(0.06–173.6 µg/200 g, 0.07–59%), and lutein (1.3–111.3 µg/200 g, 14–55%); supporting the feasibility of breeding for improved bioaccessibility (Fig. 1d-i, Data S1).

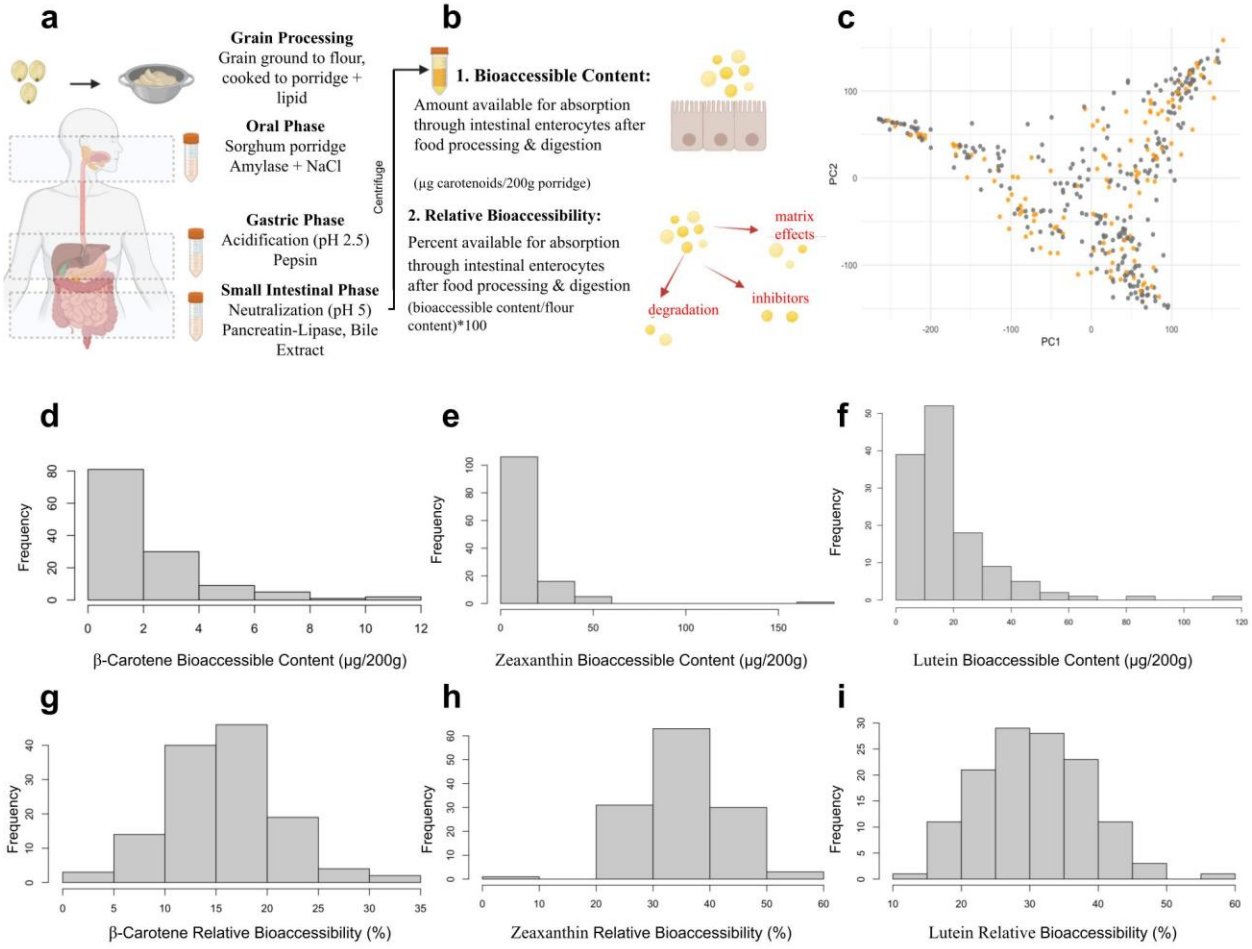


Fig 1 Global sorghum diversity reveals wide variation in carotenoid bioaccessibility (a) Phases of simulated digestion method; (b) two measurements of bioaccessibility derived from the method; (c) principal component analysis (PCA) showing the distribution of the diversity panel accessions phenotyped for bioaccessibility with respect to global genetic diversity. Genotypes in the analysis are depicted in orange and those not included are depicted in grey; (d-f) range of bioaccessible content in the diversity panel accessions for β-carotene, zeaxanthin, and lutein; and (g-i) range of relative bioaccessibility in the diversity panel accessions for β-carotene, zeaxanthin, and lutein (n = 128)

Next, we hypothesized that variation in bioaccessible content primarily reflects 1) variation in starting content in the raw flour or 2) losses during digestion caused by degradation or inhibition. Pearson's correlations analysis in the diversity panel accessions revealed strong positive correlations between raw flour content and bioaccessible content for lutein ($r = 0.95$, $P < 10^{-16}$), zeaxanthin ($r = 0.98$, $P < 10^{-16}$), and β-carotene ($r = 0.94$, $P < 10^{-16}$) (Fig. S1). Similar

correlations were observed across subsequent experiments (Fig. S1, Data S1), indicating that bioaccessible content is largely a function of the starting carotenoid content in the flour.

To identify additional factors contributing to variation in bioaccessibility, we examined Pearson's correlations between carotenoid bioaccessibility and five potential inhibitors: iron, zinc, phenolics, phytates, and thousand kernel weight (TKW) as a proxy for total starch. In the biparental family (PI655996 × PI585348), associations of phytate and phenolic content with relative bioaccessibility were weak and varied by carotenoid and year (Data S1). TKW associations were similarly inconsistent (Data S1).

Since carotenoid content and these candidate inhibitors are highly dependent on environmental context, we performed further correlation analyses within and between environments using 12 accessions and 4 environments. This GxE analysis revealed additional, environment-dependent relationships that were not captured by the pooled correlation analysis above. In 2022, iron was negatively correlated with β-carotene bioaccessible content overall (r = -0.23, $P < 0.05$) and this relationship was consistent within Texas across all three carotenoid bioaccessible contents (β-carotene: r = -0.37, $P < 0.05$; lutein: r = -0.43, $P < 0.05$; and zeaxanthin: r = -0.62, $P < 0.001$). However, in 2023 no significant correlations were observed between iron and bioaccessible carotenoid content except for a positive correlation between iron and zeaxanthin bioaccessible content was observed in Puerto Rico in 2023 (r = 0.47, $P < 0.01$) (Fig. S2). Zinc was also negatively correlated with β-carotene bioaccessible content across environments in 2022 (r = -0.23, $P < 0.05$) (Fig. S2). In 2022 across environments, phytic acid was negatively correlated with β-carotene relative bioaccessibility(r = -0.21, $P < 0.05$), but showed contrasting correlations with zeaxanthin relative bioaccessibility within environments (TX: r = -0.51, $P < 0.001$, SC: r = 0.41, $P < 0.05$) (Fig. S2).

3.2 Some carotenoid bioaccessibility traits are moderately stable across environments.

Breeding for bioaccessible PVAC requires that genetic variation contributes substantially to the phenotypic variation. To test this, we conducted a multi-location GxE study with 12 diverse accessions grown across four environments over two years. To assess trait stability across years for feasibility in nascent breeding programs, data were analyzed separately for 2022 and 2023 (Table S2-S5).

Using the standard heritability calculation, carotenoid bioaccessibility traits exhibited low to moderate heritability, with moderate to high GxE interactions contributing to variance (Table 1, Table S6). For example, β-carotene relative bioaccessibility had low heritability (H^2 = 0.39) in 2022, but practically zero in 2023 (H^2 = 0.02), accompanied by high GxE variance in 2022. This suggests the trait may respond to selection, but complex GxE interactions across environments may complicate use of phenotypic selection in breeding. Interestingly, the ratio of α-branch (lutein) to β-branch (zeaxanthin and β-carotene) bioaccessible content had higher heritability (2022: H^2 = 0.50, 2023: H^2 = 0.27) than other bioaccessible content traits (Table 1), indicating this ratio flux is a more stable and genetically controlled trait than individual bioaccessible carotenoids alone. Bioaccessible content generally had lower GxE contributions compared to

relative bioaccessibility traits, although they were more influenced by environmental variance (Table 1 and Fig. 2 a-d).

However, since the standard heritability calculation underestimates the response to selection across all genotypes and environments, we also considered the heritability by reducing the variance contributions of non-genetic effects (Table 1). β-carotene relative bioaccessibility had moderate to low heritability in both 2022 ($H^2 = 0.52$) and 2023 ($H^2 = 0.15$). As with the standard heritability calculation, the ratio of α-branch to β-branch bioaccessible content showed moderate to high heritability (2022: $H^2 = 0.66$, 2023: $H^2 = 0.85$).

Table 1. Heritability and decomposition of variance for carotenoid bioaccessibility, raw flour carotenoid content, and inhibitor content for GxE

Trait	Breeder's H^2		H^2 (V_G/V_T)		V_E/V_T		V_{GxE}/V_T		V_ε/V_T	
	2022	2023	2022	2023	2022	2023	2022	2023	2022	2023
β-Carotene RB	0.52	0.15	0.39	0.02	0	0.02	0.61	0	0.03	0.96
Lutein RB	0.07	0.17	0.01	0.03	0.12	0	0.79	0.14	0.08	0.83
Zeaxanthin RB	0	0.23	0	0.04	0.02	0	0.89	0.02	0.09	0.94
β-Carotene BC	0.01	0.65	0.11	0.20	0.51	0.06	0.34	0.21	0.04	0.53
Lutein BC	0.55	0.50	0.14	0.11	0.39	0.18	0.41	0.27	0.05	0.44
Zeaxanthin BC	0.71	0.66	0.29	0.19	0.28	0.07	0.39	0.11	0.04	0.63
Ratio α:β BC	0.66	0.85	0.50	0.27	0.06	0.13	0.33	0.02	0.10	0.57
PVAC:total BC	0.14	0.41	0.19	0.10	0.36	0.03	0.35	0.11	0.10	0.77
Total β-carotene*	0.32	0.82	0.11	0.24	0.22	0.19	0.65	0.03	0.02	0.53
Zeaxanthin	0.77	0.78	0.34	0.22	0.34	0.18	0.30	0.07	0.03	0.54
Lutein	0.62	0.77	0.24	0.14	0.28	0.35	0.44	0	0.03	0.51
Ratio α:β	0.89	0.89	0.60	0.42	0.09	0.32	0.28	0.17	0.04	0.10
Ratio PVAC:total	0.82	0.74	0.50	0.55	0.16	0.28	0.31	0.03	0.04	0.14

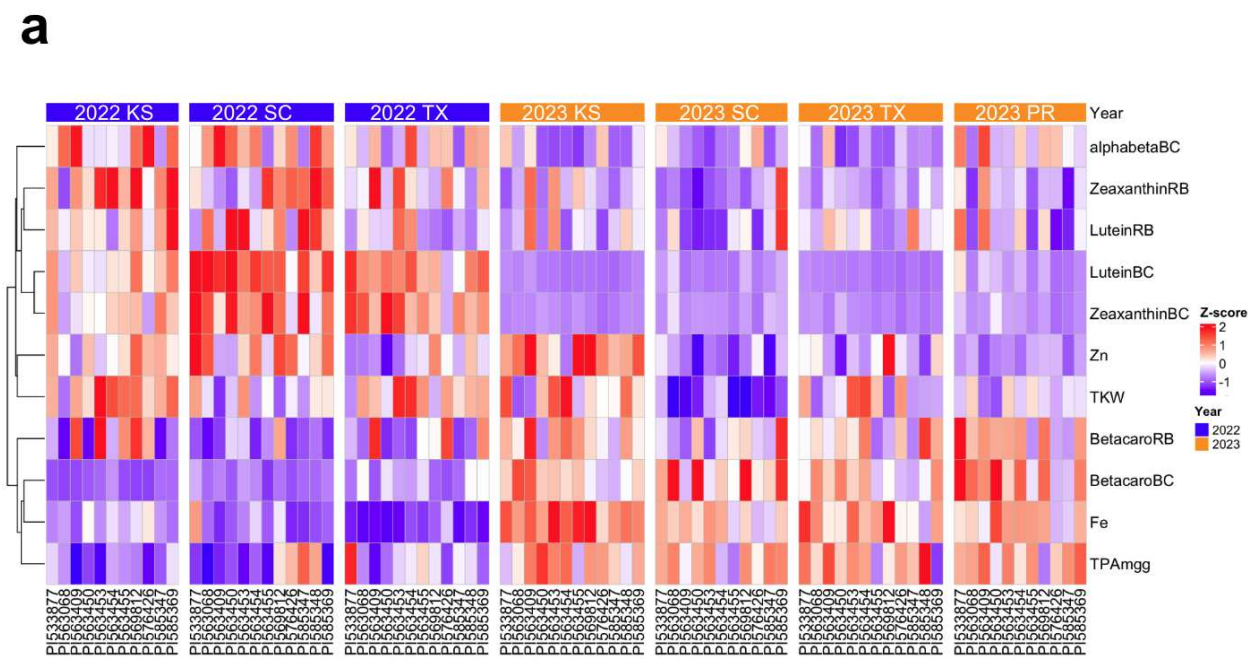
RB: relative bioaccessibility % ; BC: bioaccessible content (ug/200g); PVAC: provitamin A carotenoids

*: signifies a provitamin A carotenoid

To further understand GxE patterns and stability of bioaccessibility traits across environments, we examined z-scores of trait means across environments and environment-years (Fig. 2a). The resulting heatmap showed a clear year effect: 2022 environments generally had higher z-scores for zinc, and lutein/zeaxanthin relative bioaccessibility and bioaccessible content, whereas 2023 environments had higher z-scores for β-carotene relative bioaccessibility and bioaccessible content, iron, and total phytic acid (Fig. 2a). Additionally, we calculated correlations between these variables between environments (Fig. S2), which showed inconsistent sign patterns of hypothesized inhibitor content with bioaccessibility traits across environments.

Seeing that relative bioaccessibility traits had large GxE and error contributions to variance (Table 1), we used AMMI biplots, which separate genotype and environment main

effects from their interaction, to identify which genotypes and environments contributed most to GxE for these traits (Fig. 2b-d). To test if genotypes with similar starting carotenoid content clustered together, and how GxE affected their bioaccessibility, we colored genotypes by carotenoid content rank. For all three relative bioaccessibility traits, both axis IPCA1 and IPCA2 collectively explained over 60% of the GxE interaction variance (Fig. 2b-d). For β-carotene relative bioaccessibility, top- and middle-rank β-carotene lines clustered on the left side of the first axis (IPCA1), whereas lower-rank lines clustered on the right (Fig. 2b). However, for zeaxanthin relative bioaccessibility, there was no observable clustering pattern for the genotypes. Notably, genotype PI533877 was near the origin in the biplot, indicating that it was less susceptible to GxE driven variation than the other genotypes (Fig. 2c). However, this stability did not extend across the other relative bioaccessibility traits. Lastly, for lutein relative bioaccessibility, there was some concentration dependent clustering, with low lutein lines left of the environmental vectors on the negative diagonal (Fig. 2d). In the biplots, environment vectors pointed in different directions, indicating that environments ranked genotypes differently for relative bioaccessibility (Fig. 2b-d). Together with the variance components, which attributed most relative bioaccessibility variance to GxE in 2022 and to residual error in 2023 (Table 1), this suggests that relative bioaccessibility is environmentally unstable. The large 2023 error variance may partly reflect environmental variation not captured by the model, but this cannot be separated from measurement error in our data.



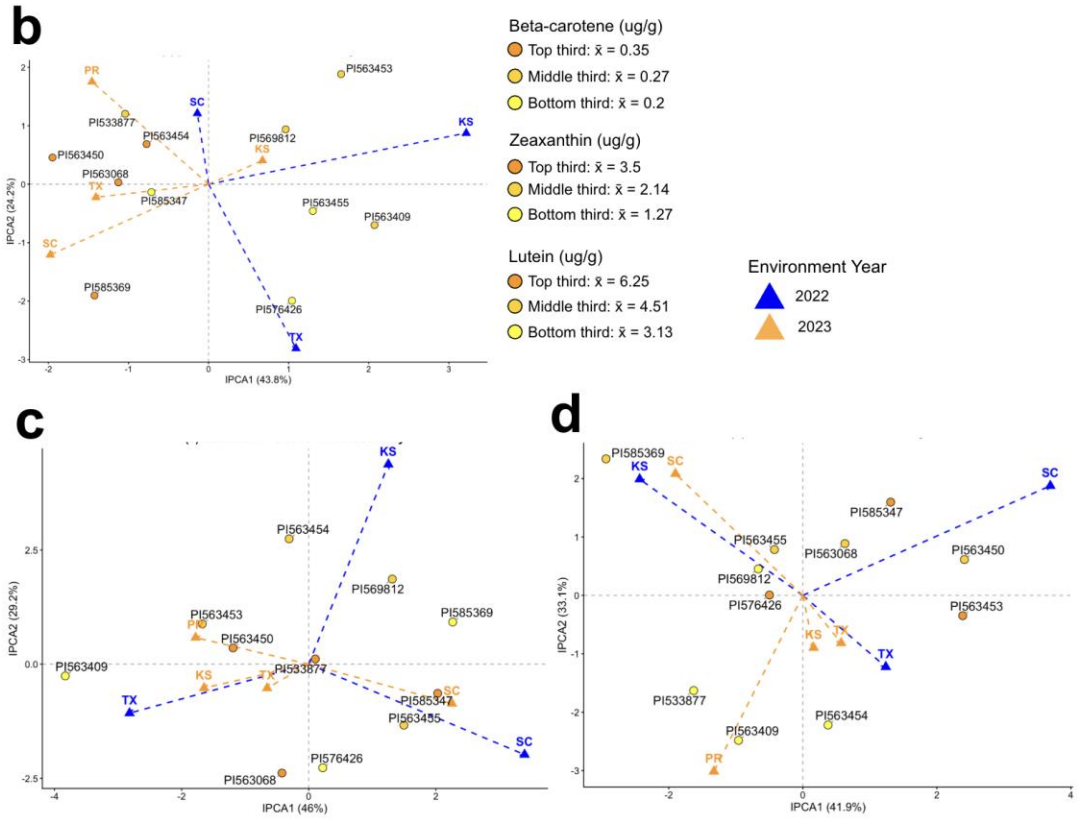


Fig 2 GxE interaction limits relative bioaccessibility trait stability (a) Heat map cladogram of z-scores for bioaccessibility traits (RB: Relative Bioaccessibility %; BC: Bioaccessible content (ug/ 200g); ratios of bioaccessible products (alphabetaBC) hypothetical inhibitors (TKW: Thousand Kernel weight (ug/g), TPAmgg: Total Phenolics (mg/g), Fe and Zn (ppm), across environments (KS: Kansas, SC: South Carolina, TX: Texas, PR: Puerto Rico) and environment years 2022 blue and 2023 in orange; AMMI biplots of (b) β-carotene relative bioaccessibility; genotypes were colored by β-carotene content (ug/g) in raw flour according to 2022 and 2023 averages of top third of lines (dark orange, x̄=0.35 ug/g), middle third (orange, x̄=0.27 ug/g), and bottom third (yellow, x̄=0.2 ug/g). (c) zeaxanthin relative bioaccessibility; genotypes were colored by zeaxanthin content (ug/g) in raw flour according to 2022 and 2023 averages of top third of lines (dark orange, x̄= 3.5 ug/g), middle third (orange, x̄= 2.14 ug/g), and bottom third (yellow, x̄= 1.27 ug/g) and (d) lutein relative bioaccessibility; genotypes were colored by lutein content (ug/g) in raw flour according to 2022 and 2023 averages of top third of lines (dark orange, x̄= 6.25 ug/g), middle third (orange, x̄= 4.51 ug/g), and bottom third (yellow, x̄= 3.13 ug/g). Genotypes with singularities were dropped from AMMI analyses (PI585348)

3.3 Carotenoid bioaccessibility traits share common genetic regulation

To identify the genetic basis of variation in carotenoid bioaccessibility traits, we conducted QTL mapping using a biparental mapping family (PI655996 × PI585348) grown over two years (Fig. 3).

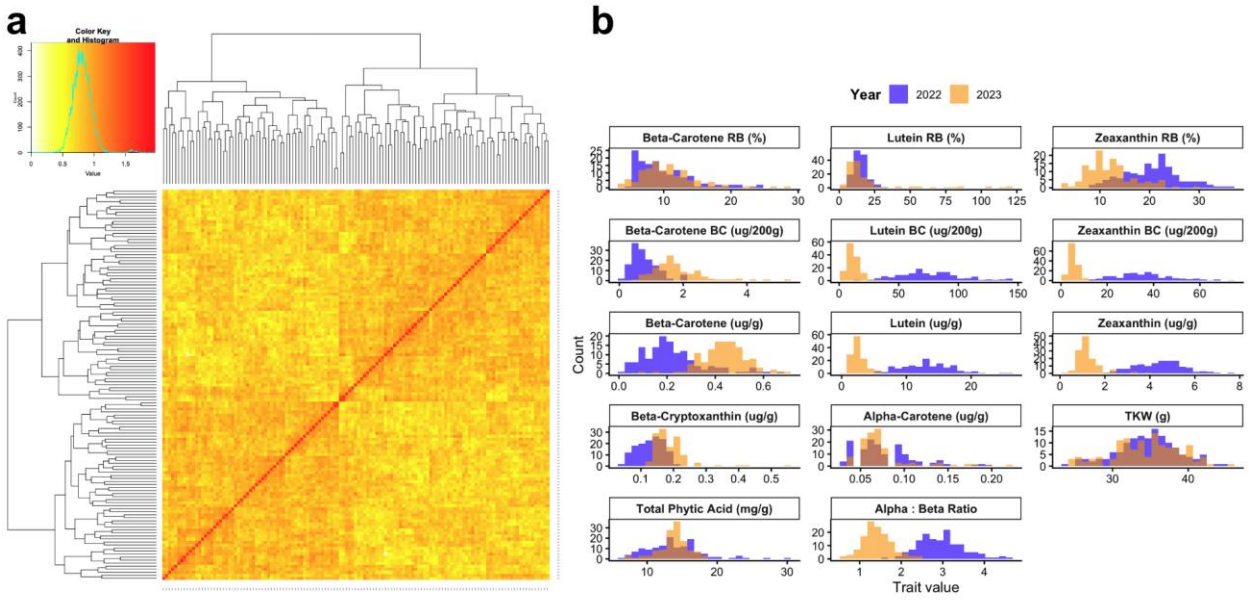


Fig 3 The biparental family segregates for bioaccessibility traits, enabling QTL mapping of the genetic basis (a) Genetic structure evaluated through ZHANG kinship matrix of the biparental family used for QTL mapping and (b) Phenotypic structure displayed through histograms of F5:6 in 2022 (blue) and F6:7 in 2023 (orange) for major carotenoid and carotenoid bioaccessibility phenotypes. RB: Relative bioaccessibility, BC: Bioaccessible content, TKW: Thousand Kernel Weight

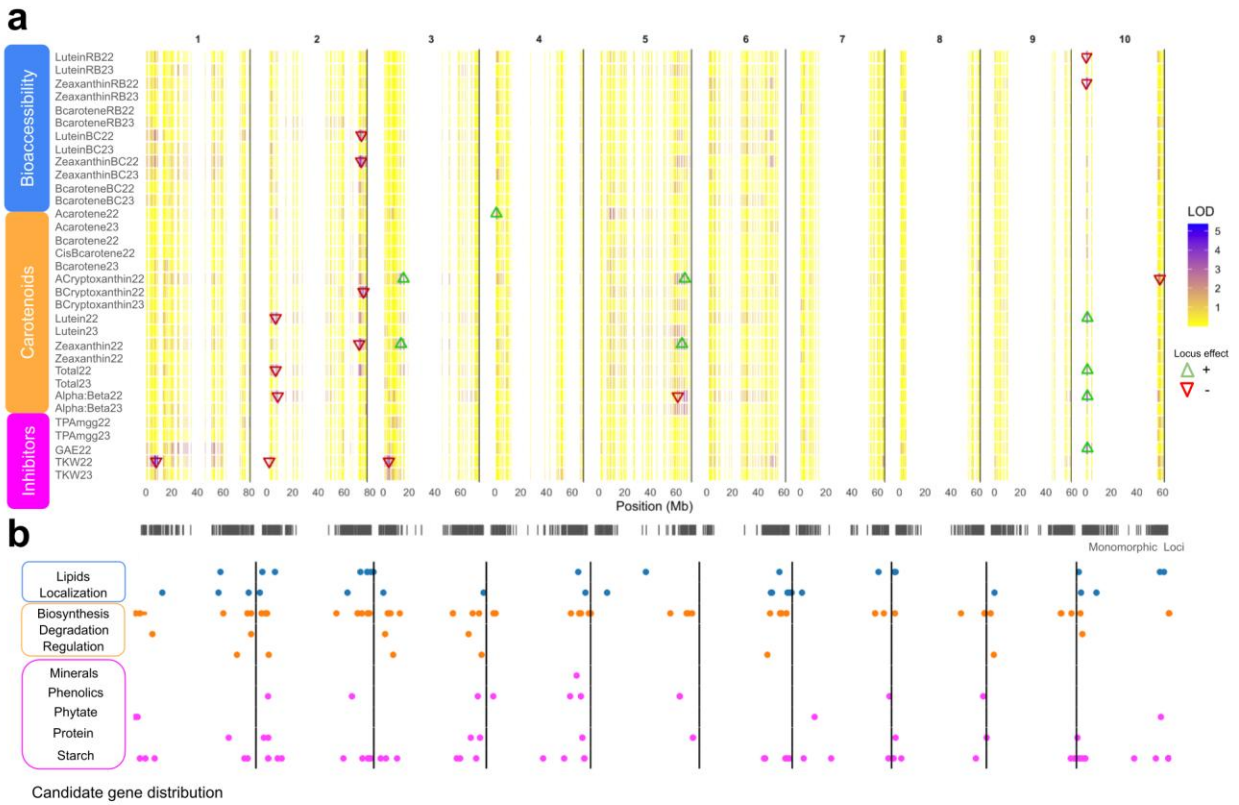
Consistent with our hypothesis that bioaccessibility variation has a genetic component, broad-sense heritability in the biparental family was generally high in 2022 and moderate to high in 2023 consistent with other inbred mapping studies (Fernandez et al. 2008) (Table 2). Notably, β-carotene bioaccessible content had the highest heritability of all the traits in 2022 ($H^2 = 0.95$), but moderate heritability in 2023 ($H^2 = 0.26$). Similar to the GxE study, the ratios of α- to β-branch carotenoids, representing metabolic flux through the biosynthesis pathway, had some of the highest heritabilities, with the bioaccessible content α: β-branch ratio showing the highest heritability in 2023 ($H^2 = 0.51$). This further suggests that selecting for pathway flux ratios could be a successful PVAC breeding strategy.

Table 2. Heritability estimates (V_G/V_T) for bioaccessibility traits in the biparental family

Trait	H^2 (2022)	H^2 (2023)

β-carotene RB	0.93	0.23
Lutein RB	0.90	0.25
Zeaxanthin RB	0.86	0.21
β-carotene BC	0.95	0.26
Lutein BC	0.92	0.29
Zeaxanthin BC	0.91	0.27
% BC PVAC	0.93	0.41
Ratio α:β BC	0.82	0.51

575 BC: bioaccessible content; RB: relative bioaccessibility; PVAC: provitamin A carotenoids



576

Fig 4 Common genetic regulation underlies carotenoid content, bioaccessibility, and inhibitor traits (a) Linkage heat map of LOD scores for bioaccessibility, carotenoid, and inhibitor phenotypes for 2022 and 2023 across 10 chromosomes. Monomorphic loci were excluded from analysis and their positions are depicted by the grey vertical lines. Significant QTL are represented by red triangles (negative effect) and green triangles (positive effect). (b) Candidate gene distribution for bioaccessibility, carotenoid, and inhibitor traits

Table 3. QTL for bioaccessibility, carotenoid, and inhibitor compounds

QTL Peak SNP	Trait	PVE~	Est. effect*	kb^	Candidates and position (bp)	Candidate Gene Annotation (abbrev)
S1_7953720	TKW	28.2	-1.4	0	Sobic.001G103600[a] 7951938-7954166	Tetrapyrrole methylase/sugar transport (ST)
S2_1829560	TKW	28.2	-0.8	17	Sobic.002G020100[a] 1847051-1851888	Chalcone synthase (*CHS*)
S2_6843380	Lutein	24.7	-1.4	>500	Sobic.002G072400[a] 7359871-7365189	Zeta-carotene desaturase (*ZDS*)
	Total	19.9	-1.7			
S2_8423718	Ratio α:β	32.9	-0.2	12	Sobic.002G080100[b] 8435786-8437465	No apical meristem (*NAM*)
S2_71654434	Zeaxanthin	28.1	-0.32			
S2_73160854	Zea BC	13.6	-4.6			
S2_73375802	Lut BC	10.8	-8.6	3	Sobic.002G377600[a] 73437256-73440299	Patatin/Phospholipase A2-related (PLA2)
S2_74972257	β-Crypto	12.1	0.0	12	1: Sobic.002G398600[a] 74984247-74988503	1: CYP734A1 P450 hydroxylase (*CYP*)
				<1	2: Sobic.002G398400[a] 74973029-74976437	2: Glycosyltransferase (*GT*)
S3_4244979	TKW	28.2	-1.0	0	1: Sobic.003G046400[c] 4238059-4248215	1. Transcriptional Corepressor Leunig *WD40*
				297	2: Sobic.003G049600[a] 4541986-4548948	2. Cellulose synthase (*CesA*)
S3_15675914	α-Crypto	26.5	0.1	0	Sobic.003G149000[b] 15675303-15680861	Bidirectional Sugar Transporter (*SWEET*)
S3_13716353	Zeaxanthin	28.1	0.3	50	Sobic.003G141200[c] 13766643-13770312	Protein brassinosteroid insensitive 1 (*BRI1*)
S4_1120014	α-Caro	9.6	0.0	9	Sobic.004G013600[c] 1111432-1115414	Autophagy-related protein 5 (*ATG5*)
S5_61168342	Ratio α:β	32.9	-0.1	117	Sobic.005G145300[c] 61047436-61051332	Flavin-binding kelch repeat F-box protein 1 (*FKF1*)
S5_66622383	α-Crypto	26.5	0.1	0	Sobic.005G183000[c] 66613656-66623637	Ribonuclease Inhibitor, DISEASERSIST, RPP13-RELATED

S5_64411841	Zeaxanthin	28.1	0.3	205	Sobic.005G165566[b] 64206131-64206488	Non-Specific Lipid Transfer Protein (*LTP*)
S10_635046	Lut RB	11.9	-1.7	38	Sobic.010G008200[b] 673084- 673797	Bifunctional inhibitor/lipid- transfer protein/seed storage 2S albumin (SS)
	Zea RB	16.2	-2.4			
S10_1493293	Lutein	24.7	1.3	49	Sobic.010G018600[b] 1542247-1544495	Alcohol/retinol dehydrogenase *(RDH*)
	Phenols	12.9	0.0			
	Ratio α:β	32.9	0.2			
	Total	19.9	1.5			
S10_57750536	α-Crypto	26.5	-0.1	24	Sobic.010G235300[c] 57774185-57775808	Stearyl acyl carrier protein desaturase (*SAD*)

PVE: percent phenotypic variation explained

~PVE for lutein, ratio α:β, zeaxanthin, TKW, α-cryptoxanthin, and total carotenoid content includes multiple markers in the model

*Effect sizes are relative to the PI585348 donor allele

^Distance between peak SNP and start/end of candidate, distance of 0 indicates within coding region

[a]*A priori* candidate

[b]*A posteriori* candidate

[c]*Post hoc* candidate

3.3.1 Carotenoid bioaccessibility mapping

We hypothesized that carotenoid bioaccessibility is under oligogenic control (as opposed to an infinitesimal model). Mapping identified several significant QTL associated with bioaccessibility traits across chromosomes 2 and 10 (Fig. 4, Table 3, Data S1). We further hypothesized that variation in lutein, zeaxanthin, and β-carotene bioaccessibility share common regulatory controls. Bioaccessibility traits for lutein, zeaxanthin, and β-carotene were positively correlated (Fig. S1, Data S1), and some QTL overlapped between these traits. On chromosome 10, lutein relative bioaccessibility and zeaxanthin relative bioaccessibility QTL co-localized (S10_635046), although this region is not near any *a priori* candidate genes. On chromosome 2, a QTL for lutein bioaccessible content (S2_73375802) was identified close to *a priori* candidate gene Sobic.002G377600 (3 kb away), a putative patatin/phospholipase A2-related gene (Table 3, Fig. 4). A putative phytoene desaturase gene (Sobic.002G383400) also lies within the confidence interval, but far from the LOD peak. Within the same confidence interval, a QTL for zeaxanthin bioaccessible content was found (S2_73160854), however was not in close proximity to any candidates.

3.3.2 Carotenoid content mapping

To further test our hypothesis that carotenoid bioaccessible content is a function of the raw carotenoid content in the flour, we predicted overlap between QTL for bioaccessibility and

carotenoid traits. In 2022, 11 unique significant QTL were identified across chromosomes 2,3,4,5, and 10 for α-carotene, α-cryptoxanthin, β-cryptoxanthin, lutein, zeaxanthin, total carotenoids and the α- to β-branch ratio (Fig. 4a, Table 3). No significant QTL were detected for any traits in 2023. Notably, confidence intervals for lutein content and the α- to β-branch ratios on chromosome 10 (S10_1493293) overlapped with the lutein relative bioaccessibility and zeaxanthin relative bioaccessibility QTL (S10_635046). Similarly, on chromosome 2, the confidence intervals for β-cryptoxanthin (S2_74972257), zeaxanthin bioaccessible content (S2_73160854), and lutein bioaccessible content (S2_73375802) overlapped (Fig. 3, Table 3), in which S2_74972257 and S2_73375802 are in moderate linkage ($R^2 = 0.65$), and S2_73160854 and S2_73375802 are in strong linkage ($R^2 = 0.97$)(Data S1).

In addition to the overlapping QTL regions between carotenoid traits and carotenoid bioaccessibility traits, the QTL that are unique to carotenoid traits are also potential targets for marker development. On chromosome 2, the confidence interval for a lutein QTL (S2_6843380) contained an *a priori* biosynthesis candidate, Sobic.002G072400, which encodes a zeta-carotene desaturase, although the gene is fairly far from the peak. However, it appears to be in partial linkage with markers proximal to the candidate ($R^2 = 0.78$: S2_6843380 - S2_7347374, $R^2 = 0.50$: S2_6843380 - S2_8423718) (Fig. 3, Table 3, Data S1).

A β-cryptoxanthin QTL on chromosome 2 (S2_74972257) was detected near an *a priori* candidate gene encoding a glycosyltransferase (Sobic.002G398400, 772 bp away) and an *a priori* candidate gene encoding a CYP734A1 P450 hydroxylase (Sobic.002G398600, 12 kb away). On chromosome 3, two QTL were found for α-cryptoxanthin (S3_15675914) and zeaxanthin (S3_13716353), however these QTL are not in linkage. The QTL S3_15675914 was within a Bidirectional Sugar Transporter (*SWEET*) Sobic.003G149000, where SWEET transporters have been shown to be differentially expressed during ripening in *Prunus salicina* when carotenoid concentrations are increasing (Lin et al. 2023). The QTL for zeaxanthin S3_13716353 was about 50 kb away from the candidate Sobic.003G141200 which encodes Protein brassinosteroid insensitive 1 (*BRI1*), which has been shown in *Brassica oleracea var. Alboglabra* to interact with the carotenoid loci *CRTISO* and *PSY2* (Zhang et al. 2024).

A QTL for the α- to β-branch ratio on chromosome 2 was near an *a posteriori* candidate gene coding a No Apical Meristem (*NAM*) protein (Sobic.002G080100), 12 kb from S2_8423718 (Fig. 3, Table 3). NAM proteins have been implicated in senescence and nutrient remobilization in plants, and have been shown to be positive regulators of carotenoid accumulation (Zhu et al. 2014). On chromosome 4, a QTL for α-carotene content was found at S4_1120014, and it was about 9 kb away from Sobic.004G013600, annotated as *ATG5* where the *ATG* gene family is involved in senescence, autophagy and nutrient recycling (Contreras et al. 2026; Magen et al. 2022; Minina et al. 2018).

Three additional QTL were identified on chromosome 5 for the α:β ratio (S5_61168342), α-cryptoxanthin (S5_66622383), and zeaxanthin (S5_64411841), and these QTL were not in significant linkage with one another, indicating they are independent loci. The QTL S5_61168342 was approximately 117 kb away from Sobic.005G145300 which encodes a

Flavin-binding kelch repeat F-box protein 1 (FKF1), where in *Cucumis sativus L.* it was implicated to interact with β-carotene hydrolase 1 (CHYβ1) and is involved in photoperiod and stress response (Musumba et al. 2025). There were no candidates in close proximity to S5_66622383, and S5_64411841, however there was candidate (205 kb) away from S5_64411841, Sobic.005G165566, which encodes a Non-Specific Lipid Transfer Protein (*LTP*). A QTL was identified for lutein and the α- to β-branch ratio that peaked on chromosome 10 (S10_1493293) in proximity (49 kb away) to the *a posteriori* candidate Sobic.010G018600, encoding an alcohol dehydrogenase (Fig. 3, Table 3). Overexpression of alcohol dehydrogenase has been associated with an increase in proanthocyanidins and carotenoid-derived compounds (Tesniere et al. 2006). Finally, a QTL for α-cryptoxanthin content on chromosome 10 (S10_57750536) was 24 kb away from a *post hoc* candidate gene encoding a stearyl acyl carrier protein desaturase with putative chloroplastic function (Sobic.010G235300) (Fig. 4, Table 3, Data S1).

3.3.3 Inhibitor content mapping

To test our hypotheses that inhibitor compounds in the grain significantly contribute to variation in carotenoid bioaccessibility (Fig. S2, Table S3), we predicted that inhibitor content and carotenoid bioaccessibility would have overlapping QTL. There was colocalization of confidence intervals of QTL on chromosome 10 for total phenolics, lutein content, and α- to β-branch ratios (S10_1493293) as well as zeaxanthin relative bioaccessibility (S10_635046) and lutein relative bioaccessibility (S10_635046) (Fig. 4, Table 3). Both QTL are in moderate linkage (R^2 = 0.66: S10_1493293 - S10_635046) (Data S1).

3.4 Biparental family markers predict bioaccessibility variation in diverse accessions

We next tested if the QTL identified in the inbred biparental family were predictive of variation in carotenoid bioaccessibility and carotenoid content in diverse accessions and environments. Trait means for each allele were calculated using data from the GxE study and the diverse panel. To control for false-positive associations between LOD peak SNPs and phenotypes, an empirical significance threshold was determined using a permutation test of 100 random markers (Data S1). Using this threshold, most SNPs individually explained a significant proportion of variation for each trait (Fig. S3-S8). Three SNPs, S2_1829560, S3_4244979, and S10_1493293 showed insufficient allelic variation or lacked marker data in the diverse accessions and were therefore excluded from allele-mean comparisons.

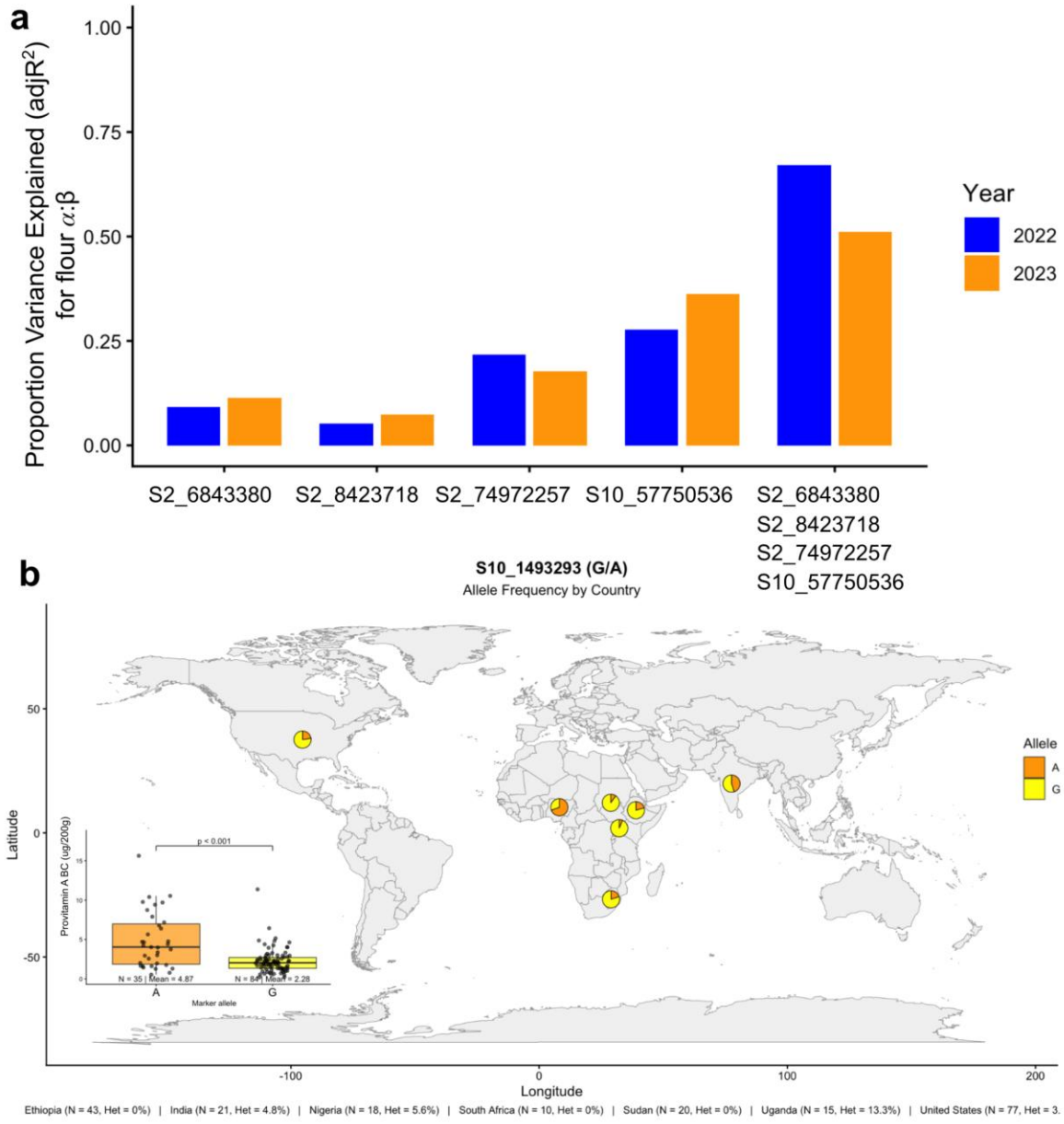


Fig 5 Biparental QTL markers predict carotenoid bioaccessibility across diverse sorghum germplasm and environments (a) α:β branch flux across experimental data from environments in diverse accessions across years**.** Shown are markers for lutein (S2_6843380), the α:β branch ratio (S2_8423718), β-cryptoxanthin (S2_74972257), and α-cryptoxanthin (S10_57750536) identified in the 2022 dataset. (b) Allelic distribution of QTL colocalized for Lutein, GAE, and α:β branch flux (S10_1493293) across georeferenced lines and allele comparisons for mean bioaccessible provitamin A content with Welch's t-test (P <0.001)

Together, four markers (S2_8423718, S2_6843380, S2_74972257, S10_57750536) explained a significant proportion of the phenotypic variation in the α- to β-branch ratios in the

raw flour in both years: 67% in 2022 ($P < 10^{-15}$), and 51% in 2023 ($P < 10^{-13}$) (Fig. 5a). They also explained 43% (2022; $P < 10^{-9}$) and 31% (2023; $P < 10^{-8}$) of the variation in the α- to β-branch ratios of bioaccessible content. The same markers explained about 30% of the variation in β-carotene relative bioaccessibility in 2022 ($P < 10^{-5}$), but none in 2023 (Data S1). Additionally, marker S2_73375802, which was associated with bioaccessible content in 2022, explained only 5-9% variation in bioaccessible content in the diversity panel dataset and did not explain variation in relative bioaccessibility (Fig. S3). Marker S2_71654434 associated with zeaxanthin content, explained almost 16% variation in 2022 for zeaxanthin bioaccessible content, however only explained about 2% of this variation in 2023 (Data S1). Although marker S10_1493293—a QTL for lutein, GAE, total carotenoids, and α:β ratio—was monomorphic for the A allele in the GxE genotypes, it segregated globally across 204 diversity lines from Ethiopia (n = 43), India (n = 21), Nigeria (n =18), South Africa (n = 10), Sudan (n = 20), Uganda (n = 15), and the US (n = 77) (Fig. 5b). The G allele, associated with α-branch carotenoids, was generally more prevalent across these lines; however, in Nigeria, the β-branch-associated A allele was more common among the georeferenced landrace lines. In the diversity panel, this allele is segregating with moderate frequency(A=0.28, G=0.68, R= 0.04), suggesting it can be selected for in global germplasm. Additionally this marker was also predictive in the diversity panel of differences in bioaccessible provitamin A content with mean values of 4.9 ug/200g porridge for the A allele and 2.3 ug/200g porridge for the G allele (P<0.001)(Fig 5b).

4 DISCUSSION

4.1 Bioaccessibility phenotyping scales moderately well for early-stage trait discovery

In this study we evaluated the scalability of an *in vitro* carotenoid bioaccessibility digestion method for early-stage trait discovery in sorghum grain (Hayes et al. 2020; Dzakovich et al. 2023). Despite the time requirements, the throughput was sufficient for initial trait discovery in sorghum grain and allowed for the identification of QTL underlying variation in carotenoid bioaccessibility.

An inherent limitation of bioaccessibility phenotyping is the difficulty of accurately quantifying bioaccessibility in samples with very low nutrient concentrations. In the crops most in need of PVAC biofortification, many genotypes have PVAC concentrations below the limit of detection after cooking and digestion and the multiple steps of the bioaccessibility assay increases the potential for measurement error. This error can disproportionately reduce repeatability, as reflected in the high residual variance for relative bioaccessibility traits in 2023 (Table 1). It is also possible this error reflects uncaptured variance within the biological replicates. We also observed a large year effect that is unlikely to be explained by environmental differences alone and may reflect differences in sample processing or analysis between years, e.g., separate digestion batches or different storage times between steps before analysis. However, AMMI biplot analysis revealed that although heritability calculations showed high contributions of error to variance, complex GxE interactions are contributing to a large proportion of variance across the genotypes and environments. Given these limitations of the

method, we propose that bioaccessibility phenotyping is most useful in the early trait discovery phase to identify potential donor lines and QTL for marker development, and again in the late phase of breeding, when applied to a small number of selected lines.

4.2 Bioaccessibility traits show sufficient genetic variation and a modest number of QTL

Genetic diversity of a trait is a prerequisite for conventional breeding, and our results support our hypothesis that significant genetic variation underlies PVAC bioaccessibility traits. Variation in bioaccessible content and relative bioaccessibility among diverse accessions, together with QTL identified in the biparental family, supports the feasibility for improvement of PVAC bioaccessibility through breeding (Fig. 1d-i; Fig. 3b). QTL for carotenoid content and bioaccessibility overlapped on chromosome 2 and 10, consistent with the strong relationship between starting carotenoid concentration and bioaccessible content (Fig. 4, Table 3).

Two QTL regions are particularly interesting for further investigation. On chromosome 10, overlapping QTL (S10_635046) for lutein and zeaxanthin relative bioaccessibility were 38 kb from *a posteriori* candidate Sobic.010G008200, putatively involved in lipid and protein storage in the grain (Table 3) and could influence carotenoid release or micellization during digestion (da Silva et al. 2020; Ke et al. 2023; Mun and McClements 2017; Pagnussat et al. 2015; Finkina E.I. et al. 2016; Yan et al. 2024). On chromosome 2, overlapping QTL (S2_73375802 and S2_73160854) for lutein and zeaxanthin bioaccessible content were near *a priori* candidate gene Sobic.002G377600, encoding a putative A2 phospholipase protein that could influence grain lipid composition and carotenoid bioaccessibility (Maoyin Li et al. 2013; Failla et al. 2019; Iddir et al. 2021). However, these candidate genes and their potential mechanisms require further validation. From a breeding perspective, linkage relationships between QTL need to be considered. On chromosome 10, linked QTL had opposing effects (Table 3), suggesting that recombination may be necessary to combine favorable alleles. In contrast, some linked QTL on chromosome 2 had effects in the same direction (S2_6843380 and S2_8423718; S2_74972257 and S2_73375802), which may facilitate simultaneous introgression of these loci.

The QTL identified in this study showed moderate associations with carotenoid bioaccessibility traits and carotenoid content (Fig. S3-S8) and were in proximity to previously reported QTL (Table S7). Additionally, these markers showed additivity: a combination of four markers explained significant variation in the carotenoid pathway flux (Fig. 5a), suggesting that combining moderate effect loci may support initial genetic gains through MAS. However, the predictive utility of these markers in different genetic backgrounds still remains to be tested. For example, S10_1493293, associated with variation in lutein, total phenolics, and the ratio of the α- and β-branch carotenoid products, explained 20% of the variation in α-cryptoxanthin bioaccessible content in the diversity panel ($P = 2 \times 10^{-5}$). This marker could not be evaluated in the 12 accessions used in the GxE study, because they all carried the A allele, likely because they are all yellow lines, but it segregated in global germplasm and was associated with differences in mean provitamin A bioaccessible content (Fig. 5b). Similarly, S2_6843380 underlying lutein variation in 2022 explained approximately 21% of the variation in lutein bioaccessible content (*P*

$< 10^{-5}$; Fig. S4, Data S1), with alleles segregating in the diversity panel (Fig. S7). Together, these findings suggest potential for MAS, although inconsistent marker associations across years and genotypes highlight the need for further validation in different genetic backgrounds and environments.

4.3 Breeding for bioaccessible PVAC requires addressing inhibitors, heritability, and pathway flux

4.3.1 Inhibitory factors reduce carotenoid bioaccessibility in an environment-dependent manner

We hypothesized that carotenoid bioaccessibility varies with grain components such as starch, protein, phenolics, phytates, and minerals. We found some support for this hypothesis, although associations were inconsistent across environments and years (Fig. S2). Negative correlations between iron and zinc concentrations and carotenoid bioaccessible content were observed in some environments, consistent with previous *in vitro* studies demonstrating that divalent minerals can decrease carotenoid micellization and uptake (Biehler et al. 2011). More studies are needed to determine if mineral-associated reductions in carotenoid bioaccessibility are biologically meaningful, or if stacking vitamins and minerals in a single variety provides a net nutritional benefit. Similarly, phytic acid and phenolics were negatively associated with bioaccessibility in some environments, but these relationships were not consistent across traits or years (Fig. S2). Further research examining the relationships between bioaccessibility and specific classes of sorghum grain phenolic compounds, such as anthocyanins or condensed tannins (Marques et al. 2021; D. H. Rhodes et al. 2014; D. Rhodes et al. 2017), may help identify compounds that most strongly reduce carotenoid bioaccessibility.

We observed generally negative correlations between TKW and bioaccessible content of some carotenoids (Data S1). A TKW-associated QTL (S1_7953720) also explained some variation in bioaccessibility traits among the diverse accessions across environments (Fig. S8), whereas another TKW QTL (S2_1829560) was near a previously reported TKW and grain width QTL (Cao et al. 2025). These associations could reflect a dilution effect, in which larger endosperms have lower carotenoid concentrations, or physical interactions between carotenoids and starches or proteins in the endosperm. However, TKW is only an indirect indicator of grain composition, and further quantification of starches and kafirin proteins is needed to determine if these factors contribute to variation in carotenoid bioaccessibility.

4.3.2 Heritability enables genetic gains despite strong GxE effects

Heritability provides insight into the relative contributions of genotype and environment to phenotypic variation and can be used to predict the potential rate of genetic gain over time. Overall, carotenoid bioaccessibility traits had lower heritability than carotenoid content traits (Table 1). Interestingly, the proportion of α- to β-branch carotenoids and their corresponding bioaccessible fractions showed higher heritability than the other traits (Table 1), indicating there is heritable variation in flux at the branching point of the carotenoid biosynthesis pathway. Variation in pathway flux has previously been exploited as a biofortification strategy in maize, through selection of a weak lycopene ε-cyclase allele which shunted metabolic flux toward the β-branch, resulting in increased accumulation of β-branch carotenoids (Harjes et al. 2008). The

relatively high heritability of pathway flux further suggests that selecting for favorable flux patterns, rather than individual carotenoids, may enable genetic gains *across* environments. Variation in heritability between years likely reflects the sensitivity of carotenoid traits to environmental conditions such as light and water availability. Drought, which has become more prevalent across the United States (Z. Li et al. 2024), increases the conversion of carotenoids into apocarotenoids, including precursors of abscisic acid (Manzi et al. 2015). Because grain carotenoid concentrations in these lines were already near the limit of detection, such environmentally driven losses would further limit the ability to detect genetic variation.

The large GxE effects observed for bioaccessibility traits (Table 1, Fig. 2 a-d) suggest that a selection strategy *within* environments as opposed to *across* environments may yield greater genetic gains per breeding cycle. Relative bioaccessibility was the trait most affected by GxE interactions (Table 1, Fig. 2b), particularly in 2022 (Table 1, Fig. 2b-d). Further research is needed to better understand the contribution of environmental factors, including light intensity, water availability (Saenz et al. 2023; Xiang et al. 2022), and soil mineral content, on variation in bioaccessibility traits. Additionally, further studies need to be conducted on higher carotenoid content lines to better assess variation with more power, as this study was limited by low carotenoid levels near the limit of detection.

4.3.3 Genetic and metabolic insights integrate into a phased breeding strategy for bioaccessible PVAC

The identification of substantial genetic variation, including three QTL associated with bioaccessibility traits and additional colocalized QTL for carotenoid content and inhibitor compounds (Fig. 6a), suggests that MAS could be an effective breeding strategy for improving bioaccessible PVAC. Based on the low observed levels of β-carotene bioaccessible content and relative bioaccessibility (0.07-11.7 µg/200g porridge, 3-31%), and average sorghum consumption in regions of Africa with high sorghum intake and high rates of vitamin A deficiency (FAOSTAT 2021), we maintain our previous recommendation for an initial raw flour breeding target concentration of 12 µg/g β-carotene to provide 25% of the estimated average requirement (EAR) for vitamin A based on estimates for storage and cooking degradation, consumption, and bioconversion efficiency (Cruet-Burgos et al. 2023). The dynamic range of relative bioaccessibility across environments, genotypes, and years suggest that the amount of bioaccessible carotenoids could vary greatly in genotypes with similar starting content, suggesting that bioaccessibility will need to be assessed in the breeding pipeline. Achieving this target across diverse breeding contexts requires a phased breeding strategy that applies our genetic and metabolic findings at each stage of variety development, from initial trait discovery through final product release.

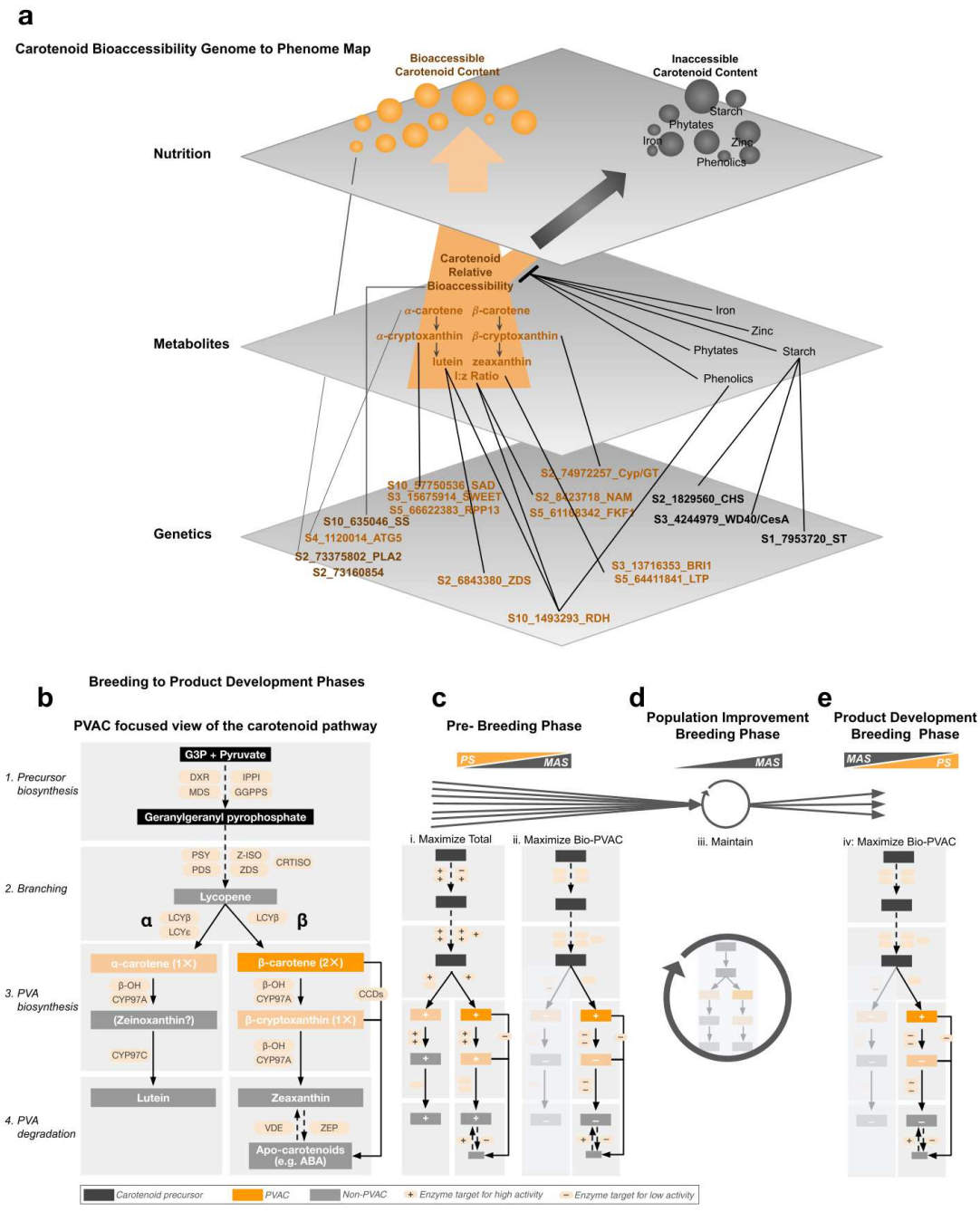
We integrated our genetic and breeding insights into a strategy for developing sorghum with enhanced bioaccessible PVAC (Fig. 6). A genome-to-phenome framework linking identified QTL and metabolites to final bioaccessibility phenotypes provides a basis for selection (Fig. 6a). This framework is illustrated in a PVAC-focused representation of the carotenoid biosynthesis pathway in sorghum grain (Fig. 6b), which highlights the differential contribution of

individual PVAC to retinol yield, underscoring the importance of β-carotene in combating vitamin A deficiency.

During the *Pre-Breeding Phase*, broad genetic diversity should be evaluated for carotenoid content and bioaccessibility using phenotypic selection in early stages to characterize the range of diversity, identify potential donor lines, and discover QTL for marker development, as demonstrated in this study. This should be followed by increasing reliance on MAS once markers are developed. Initial selection should prioritize maximizing total carotenoids, as the strong correlations we observed between carotenoid concentration and bioaccessible content suggest that gains in carotenoid concentrations may also improve bioaccessibility (Fig. 6c_i). Subsequent selection should focus on increasing β-carotene and its bioaccessibility by directing metabolic flux into the β-branch of the pathway while minimizing flux toward non-PVA end products (Fig. 6c_ii). In the *Population Improvement Breeding Phase*, MAS should be used in breeding cycles to introgress (where absent) or maintain favorable β-carotene and bioaccessible β-carotene alleles in breeding populations (Fig. 6d). Finally, in the *Product Development Breeding Phase*, MAS should be used early and phenotypic selection later to advance a small number of elite lines optimized for high β-carotene and bioaccessible β-carotene content, with reduced inhibitors if possible (Fig. 6e).

Conclusions

Overall, our study demonstrates substantial variation in sorghum carotenoid bioaccessibility across genotypes and environments. These data suggest that phenolics and minerals may decrease carotenoid bioaccessibility, indicating that these traits may be incompatible targets for simultaneous biofortification but will be targets to select against to maximize provitamin A delivery. Further research is needed to clarify the extent and mechanisms of these inhibitory effects. Furthermore, this study demonstrates the variability in stability of carotenoid bioaccessibility traits as well as limitations to using this novel phenotype at scale in a breeding program. Selecting for the ratio of α- to β-branch carotenoids to direct metabolic flux into the β-branch may be a more robust breeding strategy than selecting for total carotenoid content, since β-branch provitamin A carotenoids likely have higher vitamin A conversion rates and their flux is more heritable and stable across environments. MAS may be used to select pathway variants that maximize bioaccessible carotenoid content, although additional validation is needed to confirm the efficacy of the markers identified here. The overlap between QTL for bioaccessibility traits and carotenoid traits suggests that MAS may simultaneously improve both bioaccessible carotenoids and carotenoid compounds, potentially due to linkage or pleiotropy of causal genes. Collectively, these findings offer the most comprehensive understanding to date and provide initial recommendations for breeding for enhanced bioaccessible provitamin A sorghum.


a
Carotenoid Bioaccessibility Genome to Phenome Map
Bioaccessible Carotenoid Content
Inaccessible Carotenoid Content
Starch
Phytates
Iron
Zinc
Phenolics
Nutrition
Carotenoid Relative Bioaccessibility
α-carotene β-carotene
α-cryptoxanthin β-cryptoxanthin
lutein zeaxanthin
l:z Ratio
Metabolites
Genetics
S10_97750536_SAD
S3_15675914_SWEET
S5_66622383_RPP13
S10_635046_SS
S4_1120014_ATG5
S2_73375802_PLA2
S2_73160854
S2_74972257_Cyp/GT
S2_8423718_NAM
S5_61108342_FKF1
S2_6843380_ZDS
S3_13716353_BRI1
S5_64411841_LTP
S10_1493293_RDH
S2_1829560_CHS
S3_4244979_WD40/CesA
S1_7953720_ST
b
Breeding to Product Development Phases
PVAC focused view of the carotenoid pathway
1. Precursor biosynthesis
2. Branching
3. PVA biosynthesis
4. PVA degradation
G3P + Pyruvate
DXR
MDS
IPPI
GGPPS
Geranylgeranyl pyrophosphate
PSY
PDS
Z-ISO
ZDS
CRTISO
Lycopene
α
LCYβ
LCYε
LCYβ
β
α-carotene (1×)
β-carotene (2×)
β-OH
CYP97A
CCDs
(Zeinoxanthin?)
β-cryptoxanthin (1×)
CYP97C
Lutein
Zeaxanthin
VDE
ZEP
Apo-carotenoids (e.g. ABA)
c
Pre- Breeding Phase
PS
MAS
i. Maximize Total
ii. Maximize Bio-PVAC
d
Population Improvement Breeding Phase
MAS
iii. Maintain
e
Product Development Breeding Phase
MAS
PS
iv: Maximize Bio-PVAC
Carotenoid precursor
PVAC
Non-PVAC
Enzyme target for high activity
Enzyme target for low activity

894

Fig 6 Genetic strategy for breeding sorghum with enhanced bioaccessible PVAC (a) Genome-to-phenome framework for bioaccessibility traits, illustrating identified QTL, associated metabolomic endophenotypes, and their relationships to final bioaccessibility phenotypes. (b) PVAC-focused representation of the carotenoid biosynthetic pathway in sorghum grain. Metabolites are shown as rectangles, with carotenoid precursors in black, PVAC in orange, and non-PVAC in gray. Enzymes are shown as ovals. The number of retinol molecules released upon cleavage of each PVAC (1× or 2×) is indicated. (c-e) Schematic of proposed breeding phases when carotenoids are considered a priority ("must have") trait; the structure of each schematic corresponds to the pathway shown in panel B. (c) *Pre-breeding phase*, in which phenotypic selection (PS) is emphasized in early stages and marker assisted selection (MAS) is emphasized in later stages. Multiple horizontal arrows represent the broad genetic diversity initially evaluated and then a smaller subset to be subsequently introgressed into the breeding population. Selection objectives include (i) maximizing favorable alleles for total carotenoids , followed by (ii) maximizing favorable alleles for β-carotene and bioaccessible β-carotene by directing metabolic flux into the β-branch. (d) *Population improvement breeding phase*, during which MAS is emphasized. A circular arrow indicates the breeding cycles aimed at maintaining favorable alleles for β-carotene and bioaccessible β-carotene. (e) *Product development breeding phase*, in which MAS is emphasized initially and PS is emphasized in later stages. The few horizontal arrows represent the limited number of elite lines advanced to specifically maximize β-carotene and bioaccessible β-carotene (Maximize Bio-PVAC). Positive symbols indicate targets for selection of high-activity alleles using PS (rectangle) or MAS (ovals), whereas negative symbols indicate selection for low-activity alleles.

AUTHOR CONTRIBUTIONS

RM conducted the investigation, methodology, data curation, formal analysis, visualization, project administration, validation, and writing - original draft. **ACC** contributed to the methodology and writing - review and editing. **CF** contributed to the methodology and writing - review and editing. **MD** contributed to the methodology and writing - review and editing. **RB** contributed to the methodology and writing - review and editing. **HC** contributed to the methodology and writing - review and editing. **TF** contributed to the methodology and writing - review and editing. **CH** contributed to the methodology and writing - review and editing. **JM** contributed to the methodology and writing - review and editing. **JC** contributed to the methodology and writing - review and editing. **JP** contributed to the methodology and writing - review and editing. **GPM** contributed to the conceptualization, supervision, visualization, methodology, funding acquisition, and writing - review and editing. **EC** contributed to the conceptualization, methodology, funding acquisition, and writing - review and editing. **MF**

contributed to the conceptualization, methodology, funding acquisition, and writing - review and editing. **DHR** contributed to the conceptualization, supervision, methodology, project administration, resources visualization, funding acquisition, and writing - review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

All data files are found in the supplemental information, and additional data and analysis files can be found on github. https://github.com/rmcdower/Bioaccessibility-genomics-sorghum
Original publication with diversity panel accession genotype data (Hu et al. 2019)
Genotype data for Biparental family can be found as an ABH file and SNP calls in supplemental data file Data S1

ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Agriculture's National Institute of Food and Agriculture Foundational Knowledge of Plant Products, project award no. 2022-67013-36441, and by the Gates Foundation through the grant "Green Evolution - Accelerating Dryland Cereals Improvement for Africa" (INV-053669). Additionally, the United States Department of Agriculture Agricultural Research Service (USDA-ARS) funded the Arkansas Children's Nutrition Center through project plan #6026-10700-001-018-S.

ORCID

Rae McDowell https://orcid.org/0009-0000-5863-598X
Alvaro Cruz-Carrion https://orcid.org/0000-0002-4460-2340
Chelsey Fiecke https://orcid.org/0000-0002-8573-477X
Michael Dzakovich https://orcid.org/0000-0002-1375-427X
Richard Boyles *https://orcid.org/0000-0003-1366-7659*
Hugo Cuevas *https://orcid.org/0000-0001-7272-6933*
Terry Felderhoff *https://orcid.org/0000-0002-6610-2398*
Chad Hayes https://orcid.org/0000-0002-6535-4163
Elizabeth A. Cooper *https://orcid.org/0000-0002-2688-7418*
John McKay *https://orcid.org/0000-0003-4311-5513*
Jacqueline M Chaparro *https://orcid.org/0000-0003-1021-8719*
Jessica Prenni https://orcid.org/0000-0002-0337-8450
Geoffrey P. Morris https://orcid.org/0000-0002-3067-3359
Mario Ferruzzi https://orcid.org/0000-0003-1384-2672

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

SUPPLEMENTAL MATERIAL

Data S1- ABH file with phenotypes, Gene Candidate List, QTL information, Pearson's correlations, GxE_Genotypes, QTL models, Linkage, Permutation tests, Genotype information, Removed Genotypes (.xlsx)

Data S2- All GxE data for 2022 and 2023: Carotenoid content, Carotenoid Bioaccessibility, GAE, TKW, Phytate, ICP-MS, Correlations (.xlsx)

Supplemental methods 1 (.pdf)

Proportion of heterozygosity in family pre QC (Fig. SM1)

Kinship matrix before and after QC (Fig. SM2)

Example visualization of high heterozygosity lines. (Fig. SM3)

Example of unusual recombination patterns in suspect progeny. (Fig. SM4)

Pairwise recombination fractions and PCA scores for individuals (Fig. SM5)

Supplemental Information (.pdf)

Trait correlations within the diversity panel, GxE, Biparental family (Fig. S1-S2)

Boxplots of QTL and trait variance (Fig. S3-S8)

Allele Effects of markers in QTL data for Parental A and Parent B alleles (Fig. S9)

Average climatic conditions per environment (Table S1)

Mean carotenoid bioaccessibility traits across environments. (Table S2-S3)

Concentration of bioaccessibility traits pooled together across environments (Table S4-S5)

H2 of inhibitor metabolites in 2022 and 2023 GXE (Table S6)

QTL identified in published literature (Table S7)

REFERENCES

Balmer, James E., and Rune Blomhoff. 2002. “Gene Expression Regulation by Retinoic Acid.” *Journal of Lipid Research* 43 (11): 1773–808. https://doi.org/10.1194/jlr.R100015-JLR200.

Banda, Linly, Chloee M. McLaughlin, Avril M. Harder, et al. 2026. “Pangenomics and Machine Learning Reveal Genetic Variation to Optimize Carotenoids in Sorghum Grain.” *bioRxiv*, January 1, 2026.09.22.753247. https://doi.org/10.64898/2026.09.22.753247.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” Articles. *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, Kerstin, Susanne Kaus, Christian Krause, et al. 2002. “German Environmental Survey 1998 (GerES III): Environmental Pollutants in Blood of the German Population.” *International Journal of Hygiene and Environmental Health* 205 (4): 297–308. https://doi.org/10.1078/1438-4639-00155.

Biehler, Eric, Lucien Hoffmann, Elmar Krause, and Torsten Bohn. 2011. “Divalent Minerals Decrease Micellarization and Uptake of Carotenoids and Digestion Products into Caco-2 Cells12.” *The Journal of Nutrition* 141 (10): 1769–76. https://doi.org/10.3945/jn.111.143388.

Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. Buckler. 2007. “TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.” *Bioinformatics* 23 (19): 2633–35. https://doi.org/10.1093/bioinformatics/btm308.

Broccardo, Carolyn J., Kevin L. Schauer, Wendy M. Kohrt, Robert S. Schwartz, James P. Murphy, and Jessica E. Prenni. 2013. “Multiplexed Analysis of Steroid Hormones in Human Serum Using Novel Microflow Tile Technology and LC–MS/MS.” *Journal of Chromatography B* 934 (September): 16–21. https://doi.org/10.1016/j.jchromb.2013.06.031.

Broman, Karl W., Hao Wu, Śaunak Sen, and Gary A. Churchill. 2003. “R/Qtl: QTL Mapping in Experimental Crosses.” *Bioinformatics* 19 (7): 889–90. https://doi.org/10.1093/bioinformatics/btg112.

Cao, Ning, Yanqing Ding, Jianxia Xu, et al. 2025. “QTL Analysis of Sorghum Grain Traits Based on High-Density Genetic Map.” *Journal of Applied Genetics* 66 (3): 557–67. https://doi.org/10.1007/s13353-024-00904-w.

Casa, Alexandra M., Gael Pressoir, Patrick J. Brown, et al. 2008. “Community Resources and Strategies for Association Mapping in Sorghum.” *Crop Science* 48 (1): 30–40. https://doi.org/10.2135/cropsci2007.02.0080.

CGIAR. n.d. “Biofortified Orange-Fleshed Sweetpotato.” *Innovation Explorer*. https://www.cgiar.org/innovations/biofortified-orange-fleshed-sweetpotato/.

Cominelli, Eleonora, Roberto Pilu, and Francesca Sparvoli. 2020. “Phytic Acid and Mineral Biofortification Strategies: From Plant Science to Breeding and Biotechnological Approaches.” *Plants* 9 (5). https://doi.org/10.3390/plants9050553.

Contreras, Estefanía, Inmaculada Sánchez-Vicente, Elena Pastor-Mora, et al. 2026. “Autophagy-Related Genes AtATG5 and AtATG7 Influence Reserve Mobilization and ABA Responses during Arabidopsis Seed Germination.” *Journal of Experimental Botany*, September 17, erag468. https://doi.org/10.1093/jxb/erag468.

Cruet-Burgos, Clara, Linly Banda, Jacques M. Faye, et al. 2026. “Advancing Provitamin A Biofortification in Sorghum: From Allele Mining and Marker Development to Prebreeding and Breeding.” *Theoretical and Applied Genetics* 139 (5): 136. https://doi.org/10.1007/s00122-026-05245-2.

Cruet-Burgos, Clara, Sarah Cox, Brian P. Ioerger, et al. 2020. “Advancing Provitamin A Biofortification in Sorghum: Genome-Wide Association Studies of Grain Carotenoids in Global Germplasm.” *The Plant Genome* 13 (1): e20013. https://doi.org/10.1002/tpg2.20013.

Cruet-Burgos, Clara, Geoffrey P. Morris, and Davina H. Rhodes. 2023. “Characterization of Grain Carotenoids in Global Sorghum Germplasm to Guide Genomics-Assisted Breeding Strategies.” *BMC Plant Biology* 23 (1): 165. https://doi.org/10.1186/s12870-023-04176-0.

Cruz-Carrión, Álvaro, Ma. Josefina Ruiz De Azua, Begoña Muguerza, et al. 2023. “Organic vs. Non-Organic Plant-Based Foods—A Comparative Study on Phenolic Content and Antioxidant Capacity.” *Plants* 12 (1): 183. https://doi.org/10.3390/plants12010183.

Duodu, K. G., J. R. N. Taylor, P. S. Belton, and B. R. Hamaker. 2003. “Factors Affecting Sorghum Protein Digestibility.” *Journal of Cereal Science* 38 (2): 117–31. https://doi.org/10.1016/S0733-5210(03)00016-X.

Dzakovich, Michael P., Hawi Debelo, Marc C. Albertsen, et al. 2023. “Trait Stacking Simultaneously Enhances Provitamin A Carotenoid and Mineral Bioaccessibility in Biofortified Sorghum Bicolor.” *Food & Function* 14 (15): 7053–65. https://doi.org/10.1039/D2FO03606A.

Failla, Mark L., Daniele B. Rodrigues, and Chureeporn Chitchumroonchokchai. 2019. “Bioavailability and Metabolism of Carotenoid Esters.” In *Carotenoid Esters in Foods:*

Physical, Chemical and Biological Properties, edited by Adriana Z. Mercadante. The Royal Society of Chemistry. https://doi.org/10.1039/9781788015851-00390.

FAOSTAT. 2021. "2021 Crops and Livestock Products, Global Sorghum Production (Yield)."

Finkina E.I., Melnikova D.N., Bogdanov I.V., and Ovchinnikova T.V. 2016. "Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications." *Acta Naturae* 8 (2): 47–61. https://doi.org/10.32607/20758251-2016-8-2-47-61.

Garg, Monika, Natasha Sharma, Saloni Sharma, et al. 2018. "Biofortified Crops Generated by Breeding, Agronomy, and Transgenic Approaches Are Improving Lives of Millions of People around the World." *Frontiers in Nutrition* Volume 5-2018. https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2018.00012.

Goodstein, David M., Shengqiang Shu, Russell Howson, et al. 2012. "Phytozome: A Comparative Platform for Green Plant Genomics." *Nucleic Acids Research* 40 (D1): D1178–86. https://doi.org/10.1093/nar/gkr944.

Haley, C. S., and S. A. Knott. 1992. "A Simple Regression Method for Mapping Quantitative Trait Loci in Line Crosses Using Flanking Markers." *Heredity* 69 (4): 315–24. https://doi.org/10.1038/hdy.1992.131.

Harjes, Carlos E., Torbert R. Rocheford, Ling Bai, et al. 2008. "Natural Genetic Variation in Lycopene Epsilon Cyclase Tapped for Maize Biofortification." *Science* 319 (5861): 330–33. https://doi.org/10.1126/science.1150255.

HarvestPlus. 2015. *2014 Annual Report HarvestPlus*. HarvestPlus.

Hasan, Nazarul, Sana Choudhary, Neha Naaz, Nidhi Sharma, and Rafiul Amin Laskar. 2021. "Recent Advancements in Molecular Marker-Assisted Selection and Applications in Plant Breeding Programmes." *Journal of Genetic Engineering and Biotechnology* 19 (1): 128. https://doi.org/10.1186/s43141-021-00231-1.

Hayes, Micaela, Marti Pottorff, Colin Kay, Mary Ann Lila, Massimo Iorizzo, and Mario Ferruzzi. 2020. "Characterization of Processing Influences on In Vitro Bioaccessibility of Carotenoids and Chlorophylls from Six Spinach Genotypes." *Current Developments in Nutrition* 4 (June): nzaa052_024. https://doi.org/10.1093/cdn/nzaa052_024.

Holland, James B., Wyman E. Nyquist, and Cuauhtemoc T. Cervantes-Martínez. 2002. "Estimating and Interpreting Heritability for Plant Breeding: An Update." In *Plant Breeding Reviews*. https://doi.org/10.1002/9780470650202.ch2.

Hu, Zhenbin, Marcus O. Olatoye, Sandeep Marla, and Geoffrey P. Morris. 2019. "An Integrated Genotyping-by-Sequencing Polymorphism Map for Over 10,000 Sorghum Genotypes." *The Plant Genome* 12 (1): 180044. https://doi.org/10.3835/plantgenome2018.06.0044.

Iddir, Mohammed, Juan Felipe Porras Yaruro, Yvan Larondelle, and Torsten Bohn. 2021. "Gastric Lipase Can Significantly Increase Lipolysis and Carotenoid Bioaccessibility from Plant Food Matrices in the Harmonized INFOGEST Static in Vitro Digestion Model." *Food & Function* 12 (19): 9043–53. https://doi.org/10.1039/D1FO00786F.

Institute of Medicine. 2003. *Dietary Reference Intakes: Guiding Principles for Nutrition Labeling and Fortification*. The National Academies Press. https://doi.org/10.17226/10872.

Jaarsveld, Paul J. van, Mieke Faber, Sherry A. Tanumihardjo, Penelope Nestel, Carl J. Lombard, and Ambrose J. Spinnler Benadé. 2005. "β-Carotene–Rich Orange-Fleshed Sweet Potato Improves the Vitamin A Status of Primary School Children Assessed with the Modified-Relative-Dose-Response Test1–33." *The American Journal of Clinical Nutrition* 81 (5): 1080–87. https://doi.org/10.1093/ajcn/81.5.1080.

Ke, Yingying, Lizhen Deng, Taotao Dai, et al. 2023. "Effects of Cell Wall Polysaccharides on the Bioaccessibility of Carotenoids, Polyphenols, and Minerals: An Overview." *Critical Reviews in Food Science and Nutrition* 63 (32): 11385–98. https://doi.org/10.1080/10408398.2022.2089626.

Kean, Ellie G., Gebisa Ejeta, Bruce R. Hamaker, and Mario G. Ferruzzi. 2007. "Characterization of Carotenoid Pigments in Mature and Developing Kernels of Selected Yellow-Endosperm Sorghum Varieties." *Journal of Agricultural and Food Chemistry* 55 (7): 2619–26. https://doi.org/10.1021/jf062939v.

Kumar, Vivek, Robert R. Klein, Benjamin Kaufman, et al. 2026. "Development and Evaluation of a Cost-Effective, Mid-Density SNP Array as a Sorghum Community Genotyping Resource." *bioRxiv*, January 1, 2026.02.20.706663. https://doi.org/10.64898/2026.02.20.706663.

Li, Maoyin, Sung Chul Bahn, Chuchuan Fan, et al. 2013. "Patatin-Related Phospholipase pPLAIIIδ Increases Seed Oil Content with Long-Chain Fatty Acids in Arabidopsis." *Plant Physiology* 162 (1): 39–51. https://doi.org/10.1104/pp.113.216994.

Li, Min, Cheikh Ndiaye, Sydney Corbin, E. Allen Foegeding, and Mario G. Ferruzzi. 2020. "Starch-Phenolic Complexes Are Built on Physical CH-π Interactions and Can Persist after Hydrothermal Treatments Altering Hydrodynamic Radius and Digestibility of Model Starch-Based Foods." *Food Chemistry* 308 (March): 125577. https://doi.org/10.1016/j.foodchem.2019.125577.

Li, Zhiying, Jason E. Smerdon, Richard Seager, Noel Siegert, and Justin S. Mankin. 2024. "Emergent Trends Complicate the Interpretation of the United States Drought Monitor (USDM)." *AGU Advances* 5 (2): e2023AV001070. https://doi.org/10.1029/2023AV001070.

Lin, Zhimin, Xiaoyan Yi, Muhammad M. Ali, Lijuan Zhang, Shaojuan Wang, and Faxing Chen. 2023. "Transcriptome Insights into Candidate Genes of the SWEET Family and

Carotenoid Biosynthesis during Fruit Growth and Development in Prunus Salicina 'Huangguan.'" *Plants* 12 (19): 3513. https://doi.org/10.3390/plants12193513.

Lipka, Alexander E., Feng Tian, Qishan Wang, et al. 2012. "GAPIT: Genome Association and Prediction Integrated Tool." *Bioinformatics* 28 (18): 2397–99. https://doi.org/10.1093/bioinformatics/bts444.

Liu, Fangyuan, Baye Wodajo, Kangxu Zhao, Sanyuan Tang, Qi Xie, and Peng Xie. 2025. "Unravelling Sorghum Functional Genomics and Molecular Breeding: Past Achievements and Future Prospects." *Plant Genome Biology* 52 (6): 719–32. https://doi.org/10.1016/j.jgg.2024.07.016.

Magen, Sahar, Heike Seybold, Daniel Laloum, and Tamar Avin-Wittenberg. 2022. "Metabolism and Autophagy in Plants—a Perfect Match." *FEBS Letters* 596 (17): 2133–51. https://doi.org/10.1002/1873-3468.14359.

Manzi, Matías, Joanna Lado, María Jesús Rodrigo, Lorenzo Zacarías, Vicent Arbona, and Aurelio Gómez-Cadenas. 2015. "Root ABA Accumulation in Long-Term Water-Stressed Plants Is Sustained by Hormone Transport from Aerial Organs." *Plant and Cell Physiology* 56 (12): 2457–66. https://doi.org/10.1093/pcp/pcv161.

Marques, Marcella Camargo, Adriele Hacke, Celso Andrade Camargo Neto, and Lilian RB Mariutti. 2021. "Impact of Phenolic Compounds in the Digestion and Absorption of Carotenoids." *Current Opinion in Food Science* 39 (June): 190–96. https://doi.org/10.1016/j.cofs.2021.03.006.

Minina, Elena A., Panagiotis N. Moschou, Ramesh R. Vetukuri, et al. 2018. "Transcriptional Stimulation of Rate-Limiting Components of the Autophagic Pathway Improves Plant Fitness." *Journal of Experimental Botany* 69 (6): 1415–32. https://doi.org/10.1093/jxb/ery010.

Mohler, V., and C. Singrün. 2005. "General Considerations: Marker-Assisted Selection." In *Molecular Marker Systems in Plant Breeding and Crop Improvement*, edited by Horst Lörz and Gerhard Wenzel. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-26538-4_18.

Mun, Saehun, and David Julian McClements. 2017. "Influence of Simulated In-Mouth Processing (Size Reduction and Alpha-Amylase Addition) on Lipid Digestion and β-Carotene Bioaccessibility in Starch-Based Filled Hydrogels." *LWT* 80 (July): 113–20. https://doi.org/10.1016/j.lwt.2017.02.011.

Munoz, Bryan, Micaela Hayes, Penelope Perkins-Veazie, et al. 2024. "Genotype and Ripening Method Affect Carotenoid Content and Bio-Accessibility in Banana." *Food & Function* 15 (7): 3433–45. https://doi.org/10.1039/d3fo04632j.

Musumba, Faith Adisa, Peter Ankutse, Chunyan Cheng, et al. 2025. "Photoperiod-Regulated Gene Interactions and Genotypic Responses in Chlorophyll Dynamics, Oxidative Stress, and Ultrastructure of Xishuangbanna Cucumber (Cucumis Sativus L.)."

Scientia Horticulturae 345 (April): 114150. https://doi.org/10.1016/j.scienta.2025.114150.

Muzhingi, Tawanda, Tendekayi H. Gadaga, Andrew H. Siwela, Michael A. Grusak, Robert M. Russell, and Guangwen Tang. 2011. "Yellow Maize with High β-Carotene Is an Effective Source of Vitamin A in Healthy Zimbabwean Men1234." *The American Journal of Clinical Nutrition* 94 (2): 510–19. https://doi.org/10.3945/ajcn.110.006486.

Pagnussat, Luciana A., Natalia Oyarburo, Carlos Cimmino, Marcela L. Pinedo, and Laura de la Canal. 2015. "On the Role of a Lipid-Transfer Protein. Arabidopsis Ltp3 Mutant Is Compromised in Germination and Seedling Growth." *Plant Signaling & Behavior* 10 (12): e1105417. https://doi.org/10.1080/15592324.2015.1105417.

Prasad, V. B. Rajendra, Mahalingam Govindaraj, Maduraimuthu Djanaguiraman, et al. 2021. "Drought and High Temperature Stress in Sorghum: Physiological, Genetic, and Molecular Insights and Breeding Approaches." *International Journal of Molecular Sciences* 22 (18). https://doi.org/10.3390/ijms22189826.

Rhodes, Davina, Priyadarshini Gadgil, Ramasamy Perumal, Tesfaye Tesso, and Thomas J. Herald. 2017. "Natural Variation and Genome-Wide Association Study of Antioxidants in a Diverse Sorghum Collection." *Cereal Chemistry* 94 (2): 190–98. https://doi.org/10.1094/CCHEM-03-16-0075-R.

Rhodes, Davina H., Leo Jr. Hoffmann, William L. Rooney, Punna Ramu, Geoffrey P. Morris, and Stephen Kresovich. 2014. "Genome-Wide Association Study of Grain Polyphenol Concentrations in Global Sorghum [Sorghum Bicolor (L.) Moench] Germplasm." *Journal of Agricultural and Food Chemistry* 62 (45): 10916–27. https://doi.org/10.1021/jf503651t.

Rooney, L. W., and R. L. Pflugfelder. 1986. "Factors Affecting Starch Digestibility with Special Emphasis on Sorghum and Corn1." *Journal of Animal Science* 63 (5): 1607–23. https://doi.org/10.2527/jas1986.6351607x.

Saenz, Ezequiel, Lucas Borrás, Gonzalo Parra, Brenda L. Gambin, and José A. Gerde. 2023. "Crop Management Options to Help Increase Maize Grain Carotenoid Levels." *Journal of Cereal Science* 114 (November): 103789. https://doi.org/10.1016/j.jcs.2023.103789.

Samtiya, Mrinal, Rotimi E. Aluko, and Tejpal Dhewa. 2020. "Plant Food Anti-Nutritional Factors and Their Reduction Strategies: An Overview." *Food Production, Processing and Nutrition* 2 (1): 6. https://doi.org/10.1186/s43014-020-0020-5.

Silva, Flávia Camila Vieira da, Evenilton Pessoa Costa, Valdirene Moreira Gomes, and André de Oliveira Carvalho. 2020. "Inhibition Mechanism of Human Salivary α-Amylase by Lipid Transfer Protein from Vigna Unguiculata." *Computational Biology and Chemistry* 85 (April): 107193. https://doi.org/10.1016/j.compbiolchem.2019.107193.

Singh, Rameshwar, and John D. Axtell. 1973. “High Lysine Mutant Gene (Hl That Improves Protein Quality and Biological Value of Grain Sorghum1.” *Crop Science* 13 (5): cropsci1973.0011183X001300050012x. https://doi.org/10.2135/cropsci1973.0011183X001300050012x.

Sreerama, Yadahally N., Dennis A. Neelam, Vadakkoot B. Sashikala, and Vishwas M. Pratape. 2010. “Distribution of Nutrients and Antinutrients in Milled Fractions of Chickpea and Horse Gram: Seed Coat Phenolics and Their Distinct Modes of Enzyme Inhibition.” *Journal of Agricultural and Food Chemistry* 58 (7): 4322–30. https://doi.org/10.1021/jf903101k.

Taylor, John R. N., and Kwaku G. Duodu. 2023. “Resistant-Type Starch in Sorghum Foods—Factors Involved and Health Implications.” *Starch - Stärke* 75 (9–10): 2100296. https://doi.org/10.1002/star.202100296.

Tesniere, Catherine, Laurent Torregrosa, Martine Pradal, et al. 2006. “Effects of Genetic Manipulation of Alcohol Dehydrogenase Levels on the Response to Stress and the Synthesis of Secondary Metabolites in Grapevine Leaves.” *Journal of Experimental Botany* 57 (1): 91–99. https://doi.org/10.1093/jxb/erj007.

Wang, Jiabo, and Zhiwu Zhang. 2020. “GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction.” *bioRxiv*, January 1, 2020.11.29.403170. https://doi.org/10.1101/2020.11.29.403170.

Wong, Joshua H., Tsang Lau, Nick Cai, et al. 2009. “Digestibility of Protein and Starch from Sorghum (Sorghum Bicolor) Is Linked to Biochemical and Structural Features of Grain Endosperm.” *Journal of Cereal Science* 49 (1): 73–82. https://doi.org/10.1016/j.jcs.2008.07.013.

World Health Organization, World Food Programme, and United Nations Children’s Fund. 2006. “Preventing and Controlling Micronutrient Deficiencies in Populations Affected by an Emergency: Multiple Vitamin and Mineral Supplements for Pregnant and Lactating Women, and for Children Aged 6 to 59 Months.” World Health Organization, April 7. https://www.who.int/publications/m/item/WHO-WFP-UNICEF-statement-micronutrients-deficiencies-emergency.

Xiang, Nan, Yihan Zhao, Siyun Wang, and Xinbo Guo. 2022. “The Modulation of Light Quality on Carotenoids in Maize (Zea Mays L.) Sprouts.” *Food Chemistry: Molecular Sciences* 5 (December): 100128. https://doi.org/10.1016/j.fochms.2022.100128.

Yan, Xueyu, Jinman Huang, Leiheng Huang, et al. 2024. “Effects of Dietary Lipids on Bioaccessibility and Bioavailability of Natural Carotenoids.” *LWT* 200 (May): 116171. https://doi.org/10.1016/j.lwt.2024.116171.

Zhang, Chenlu, Qiannan Liang, Yilin Wang, et al. 2024. “BoaBZR1.1 Mediates Brassinosteroid-Induced Carotenoid Biosynthesis in Chinese Kale.” *Horticulture Research* 11 (6): uhae104. https://doi.org/10.1093/hr/uhae104.

Zhao, Tian, Shiwei Liu, Ruijie Zhang, et al. 2022. “Global Burden of Vitamin A Deficiency in 204 Countries and Territories from 1990–2019.” *Nutrients* 14 (5). https://doi.org/10.3390/nu14050950.

Zhu, Mingku, Guoping Chen, Shuang Zhou, et al. 2014. “A New Tomato NAC (NAM/ATAF1/2/CUC2) Transcription Factor, SlNAC4, Functions as a Positive Regulator of Fruit Ripening and Carotenoid Accumulation.” *Plant and Cell Physiology* 55 (1): 119–35. https://doi.org/10.1093/pcp/pct162.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SD2GxEalldata1.xlsx
- SupplementBioaccessibilityManuscriptSupplementalInfo1.docx
- SupplementalMethods1.docx
- SD1BiparentalCaroBio1.xlsx