DISSERTATION

DEVELOPMENT OF MOLECULAR BREEDING RESOURCES FOR INCREASED PRO-VITAMIN A CAROTENOIDS IN SORGHUM GRAIN

Submitted by

Clara María Cruet-Burgos

Horticulture and Landscape Architecture

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2022

Doctoral Committee:

Advisor: Davina Rhodes

Esten Mason
Jessica Prenni
Gael Pressoir

Copyright by Clara M. Cruet Burgos 2022

All Rights Reserved

ABSTRACT

DEVELOPMENT OF MOLECULAR BREEDING RESOURCES FOR INCREASED PRO-VITAMIN A CAROTENOIDS IN SORGHUM GRAIN

Vitamin A deficiency (VAD) affects millions of people in countries in Africa and South-East Asia, contributing to decreased immune response and increased morbidity and mortality from common infections. Sorghum [*Sorghum bicolor* L. (Moench)] is a staple cereal crop in these regions, thus, sorghum carotenoid biofortification is a potential method to improve the vitamin A status of these communities. The overall aim of this research was to determine the feasibility of biofortification breeding for sorghum grain carotenoids, and to develop genomic tools to assist in molecular breeding. Global sorghum germplasm collections were evaluated for pro-vitamin A carotenoids, and concentrations were found to be below target values. Due to the low number of accessions with above average pro-vitamin A content, the genetic diversity of the high carotenoid lines in the global germplasm was assessed. High carotenoid accessions were found to be highly related, hence, to increase genetic diversity for breeding, a genomic prediction model was used to identify additional germplasm with potentially high concentrations of pro-vitamin A carotenoids. Through a genome-wide association study, it was confirmed that carotenoid variation in sorghum grain is oligogenic, but there was also evidence of a polygenic component. Therefore both marker-assisted selection (MAS) and genomic selection (GS) may be effective in accelerating breeding

efforts. KASP markers in linkage with genomic regions associated with carotenoid concentrations were developed and validated in six $F_{2:3}$ populations. Two markers in the intronic region of the carotenoid pathway *β-OH* gene were identified as good candidates to use for MAS due to their predictive ability. A marker inside the coding sequence of the carotenoid pathway *ZEP* gene was also identified as a good marker for MAS. An RNA-seq experiment identified additional genes in the MEP, carotenoid biosynthesis and carotenoid degradation pathways that could be used for MAS. The results of these studies provide a foundation for vitamin A biofortification through genomics-assisted breeding.

TABLE OF CONTENTS

ABSTRACT II
TABLE OF CONTENTS IV
LIST OF TABLES VI
LIST OF FIGURES VII
CHAPTER 1. BIOFORTIFICATION POTENTIAL OF SORGHUM GRAIN CAROTENOIDS USING MOLECULAR BREEDING 1
INTRODUCTION *1*
DISCOVERY PHASE FOR SORGHUM PRO-VITAMIN A BIOFORTIFICATION *10*
THE DEVELOPMENT PHASE FOR A VITAMIN A BIOFORTIFIED SORGHUM *22*
BREEDING FOR A VITAMIN A BIOFORTIFIED SORGHUM *29*
FUTURE CONSIDERATIONS *35*
CONCLUSION *35*
LITERATURE CITED *37*
CHAPTER 2 CHARACTERIZATION OF GRAIN CAROTENOIDS IN GLOBAL SORGHUM GERMPLASM TO GUIDE GENOMICS-ASSISTED BREEDING STRATEGIES 53
INTRODUCTION *53*
METHODS *59*
RESULTS *64*
DISCUSSION *78*
CONCLUSIONS *85*
LITERATURE CITED *86*
CHAPTER 3. DEVELOPMENT AND VALIDATION OF BREEDER-FRIENDLY KASP MARKERS FOR VITAMIN A BIOFORTIFICATION IN SORGHUM 98
INTRODUCTION *98*
METHODS *101*
RESULTS *105*
DISCUSSION *114*
CONCLUSIONS *116*
LITERATURE CITED *118*
CHAPTER 4. UNRAVELING TRANSCRIPTOMICS OF SORGHUM GRAIN CAROTENOIDS: A STEP FORWARD FOR BIOFORTIFICATION 123
INTRODUCTION *123*
METHODS *126*
RESULTS *131*
DISCUSSION *148*
CONCLUSION *157*

LITERATURE CITED *158*
CHAPTER 5. CONCLUSION 167
APPENDIX A 170
APPENDIX B 174
LIST OF ABBREVIATIONS 204

LIST OF TABLES

TABLE 1.1 SORGHUM CONSUMPTION PATTERNS, PREVALENCE OF VITAMIN A DEFICIENCY (VAD) AND CURRENT PUBLIC HEALTH MEASURES IN TARGET COUNTRIES FOR A VITAMIN A BIOFORTIFIED SORGHUM. 12

TABLE 1.2 SORGHUM A PRIORI CANDIDATE GENES ASSOCIATED WITH LUTEIN, ZEAXANTHIN OR B-CAROTENE IN GENOMIC STUDIES AND THEIR MAIZE HOMOLOGS. 34

TABLE 2.1. RANGE, MEAN, AND ENTRY-MEAN BASIS HERITABILITY (H^2) FOR THE BLUPS OF LUTEIN, ZEAXANTHIN, AND B-CAROTENE FOR THE SAP/CAP COLLECTION. 65

TABLE 2.2 RANGE OF GEBV, AVERAGE PREDICTION ACCURACY AND GENOMIC HERITABILITY (H^2) FOR THE LUTEIN, ZEAXANTHIN, AND B-CAROTENE FOR THE UNEXPLORED GERMPLASM COLLECTIONS AND SAP/CAP COLLECTION. 73

TABLE 3.1 SNP USED TO DEVELOP KASP MARKERS AND THE GENOMIC REGIONS ASSOCIATED WITH CAROTENOID CONCENTRATION. 102

TABLE 3.2 CONCENTRATION OF LUTEIN AND ZEAXANTHIN FOR PARENTS AND DISTRIBUTION OF $F_{2:3}$ PROGENIES 107

TABLE 3.3 DESCRIPTION OF KASP MARKERS EVALUATED AND THEIR ASSOCIATION WITH LUTEIN AND ZEAXANTHIN CONTENT. 111

TABLE 4.1 A PRIORI CANDIDATE GENES EXPRESSED IN SORGHUM GRAIN.... 132

TABLE 4.2 CORRELATIONS BETWEEN *A PRIORI* CANDIDATE GENE EXPRESSION AND CONCENTRATIONS OF LUTEIN, ZEAXANTHIN, AND B-CAROTENE AT 42 DAP. 147

LIST OF FIGURES

FIGURE 1.1 SORGHUM BIOFORTIFICATION PIPELINE AND STATUS. 4

FIGURE 1.2 CAROTENOID PRECURSOR MEP PATHWAY, CAROTENOID BIOSYNTHESIS AND CAROTENOID DEGRADATION PATHWAYS. 6

FIGURE 2.1 GENOMICS-ASSISTED SELECTION METHOD BASED ON GENETIC ARCHITECTURE OF A TRAIT... 58

FIGURE 2.2 GENOME-WIDE ASSOCIATION STUDY OF CAROTENOID BLUP ESTIMATES USING MLM. 67

FIGURE 2.3 DISTRIBUTION OF Β-CAROTENE BLUPS AMONG COUNTRIES IN THE SAP/CAP COLLECTION. 69

FIGURE 2.4 GENETIC DIVERSITY OF SAP/CAP ACCESSIONS BASED ON THE TOP 5% AND BOTTOM 5% RANKINGS FOR Β-CAROTENE BLUP ESTIMATES......... 71

FIGURE 2.5 CAROTENOIDS IN UNEXPLORED GERMPLASM COLLECTIONS. 75

FIGURE 2.6 GEOGRAPHIC DISTRIBUTION OF THE SNP S06_46717975 INSIDE THE ZEP GENE AND THE DISTRIBUTION OF ALLELIC CLASSES IN HIGH CAROTENOID GERMPLASM... 77

FIGURE 3.1 FAMILY DEVELOPMENT CROSSING SCHEME AND TIMELINE FOR GENOTYPING WITH KASP MARKERS AND PHENOTYPING. 105

FIGURE 3.2 LUTEIN DENSITY PLOT FOR THE SIX $F_{2:3}$ POPULATIONS AND THEIR PARENTS. 108

FIGURE 3.3 ZEAXANTHIN DENSITY PLOT FOR THE SIX $F_{2:3}$ POPULATIONS AND THEIR PARENTS. 109

FIGURE 3.4 LEAST SIGNIFICANT DIFFERENCE (LSD) FOR MARKERS ASSOCIATED WITH LUTEIN CONCENTRATION IN THE $F_{2:3}$ PROGENIES. 112

FIGURE 3.5 LEAST SIGNIFICANT DIFFERENCE (LSD) FOR MARKERS ASSOCIATED WITH ZEAXANTHIN CONCENTRATION IN THE $F_{2:3}$ PROGENIES. 113

FIGURE 4.1 SAMPLES HIERARCHICAL CLUSTERING AND HEAT MAP OF TOTAL CAROTENOID CONCENTRATION AT MATURITY FOR EACH ACCESSION. DARKER COLORS IN THE HEAT MAP REPRESENT HIGHER CAROTENOID CONCENTRATION. R1-R3 DENOTES EACH BIOLOGICAL REPLICATE. 133

FIGURE 4.2 PRECURSOR PATHWAY: DIFFERENTIALLY EXPRESSED GENES AND PATTERNS OF EXPRESSION ACROSS GRAIN DEVELOPMENTAL TIME POINTS (DAP). ... 135

FIGURE 4.3 BIOSYNTHESIS PATHWAY: DIFFERENTIALLY EXPRESSED GENES AND PATTERNS OF EXPRESSION ACROSS GRAIN DEVELOPMENTAL TIME POINTS (DAP). ... 137

FIGURE 4.4 DEGRADATION PATHWAYS: DIFFERENTIALLY EXPRESSED GENES AND PATTERNS OF EXPRESSION ACROSS GRAIN DEVELOPMENTAL TIME POINTS (DAP). ... 138

FIGURE 4.5 NORMALIZED COUNTS FOR DIFFERENTIALLY EXPRESSED GENES BETWEEN HIGH AND LOW CAROTENOID ACCESSIONS AT 14 DAP. A) CAROTENOID PRECURSORS (MEP PATHWAY); B) CAROTENOID DEGRADATION PATHWAYS ... 140

FIGURE 4.6 NORMALIZED COUNTS FOR DIFFERENTIALLY EXPRESSED GENES BETWEEN HIGH AND LOW CAROTENOID ACCESSIONS AT 28 DAP. A) CAROTENOID PRECURSORS (MEP PATHWAY); B) CAROTENOID DEGRADATION PATHWAYS ... 142

FIGURE 4.7 NORMALIZED COUNTS FOR DIFFERENTIALLY EXPRESSED GENES BETWEEN HIGH AND LOW CAROTENOID ACCESSIONS AT 42 DAP. A) CAROTENOID PRECURSORS (MEP PATHWAY); B) CAROTENOID DEGRADATION PATHWAYS ... 144

FIGURE 4.8. GENOTYPES NORMALIZED TRANSCRIPT COUNTS FOR CAROTENOID BIOSYNTHESIS GENES DIFFERENTIALLY EXPRESSED AT 42 DAP IN HIGH VERSUS LOW CAROTENOID SORGHUM ACCESSIONS. A) PHYTOENE SYNTHASE (PSY); B) PHYTOENE DESATURASE (PDS) ... 145

CHAPTER 1. BIOFORTIFICATION POTENTIAL OF SORGHUM GRAIN CAROTENOIDS USING MOLECULAR BREEDING

INTRODUCTION

Malnutrition continues to affect millions of people around the world and current global crises are exacerbating the problem further. Global trends showed a decrease of undernutrition until 2014 when the decline came to a halt (FAO et al., 2021). The COVID-19 pandemic, emerging conflicts, economic downturns, climate variability and extreme weather events are proposed as the main drivers underlying the changes in the global trend (Global Nutrition Report, 2021). In 2020, after an alarming increase of 1.5%, it is estimated that 768 million people are undernourished today (FAO et al., 2021). Undernutrition in the form of micronutrient deficiencies—often referred to as hidden hunger—affects 1 in 3 people globally. The most prevalent micronutrient deficiencies are iodine, iron, zinc, and vitamin A. Vitamin A deficiency (VAD) affects 190 million preschool age children and 19 million pregnant women worldwide, contributing to decreased immune response and increased morbidity and mortality from common infections (Global Nutrition Report, 2014). It is estimated that 46 countries have severe rates of VAD, predominantly in Africa and South-East Asia (World Health Organization, 2009). Plant carotenoids are the main source of vitamin A in these regions, as over 90% of the calories come from staple crops such as wheat, rice, maize, sorghum, cassava, and sweet potato (Global Nutrition Report, 2021). Staple crops provide a reliable source of calories for a low cost, however, the concentration of pro-vitamin A carotenoids in staple crops tends

to be low, thus these plant-based diets often fail to provide the necessary requirements, leaving these communities vulnerable to VAD (World Health Organization, 2009).

Public health measures to counteract VAD have been developed and implemented with varying success over the last decades. Pre-formed vitamin A supplementation programs and food fortification—the addition of micronutrients to processed food—alone or in combination, are the two most adopted intervention methods, as they can alleviate VAD and the associated disorders in a relatively short time. Since the 1990's, over 80 countries have implemented a vitamin A supplementation program for children 6 months to 5 years of age and pregnant or lactating women, reducing child mortality from major infections such as diarrhea, measles, and malaria (Wirth et al., 2017). Fortification of sugar (Allen et al., 2006; Palmer et al., 2012), edible oils (Garrett et al., 2015; Olson et al., 2021), wheat flour (Allen et al., 2006) and milk (Olson et al., 2021) have also been implemented successfully in multiple countries to improve the vitamin A status of a wider population. However, despite their immediate success, neither supplementation nor food fortification improves the underlying poor-quality diet.To be effective, continuous government, private, or non-profit action is required to reach the most susceptible households, which often are located in remote communities (Allen et al., 2006; Palmer et al., 2012). Less adopted measures, such as diet diversification—the consumption of food from multiple food groups—and crop biofortification—the improvement of the nutritional value of crops through breeding or genetic engineering—can have a long-term impact on the vitamin A status in a population by improving the quality of the diet. Diet diversification has an advantage over the other methods, because it can address unknown nutritional deficiencies that might be present in a population, and if implemented correctly, can target

multiple micronutrient deficiencies at the same time (Nair et al., 2016). However, it is a strategy that requires time to implement, has slow returns, relies on changing consumption habits that might be culturally rooted, and might not be economically feasible for the low income communities that tend to be at higher risk (Nair et al., 2016). Biofortification is a sustainable alternative or complementary method to address micronutrient deficiencies, and can impact the vitamin A status of a population long-term, with the potential over time to move millions of people over the malnourishment threshold to micronutrient sufficiency (Bouis et al., 2011). Unlike other measures, biofortification can reach the most susceptible households in poor or rural communities that rely on staple crops, and can be utilized in processed products accessible in urban areas. Biofortification is also an economically feasible alternative, because after initial investment to develop nutrient dense varieties, the biofortified germplasm can be grown year after year by farmers and distributed internationally utilizing existing seed distribution systems (Qaim et al., 2007). However, biofortification is a decades-long process as it requires multidisciplinary research and development (Figure 1.1).

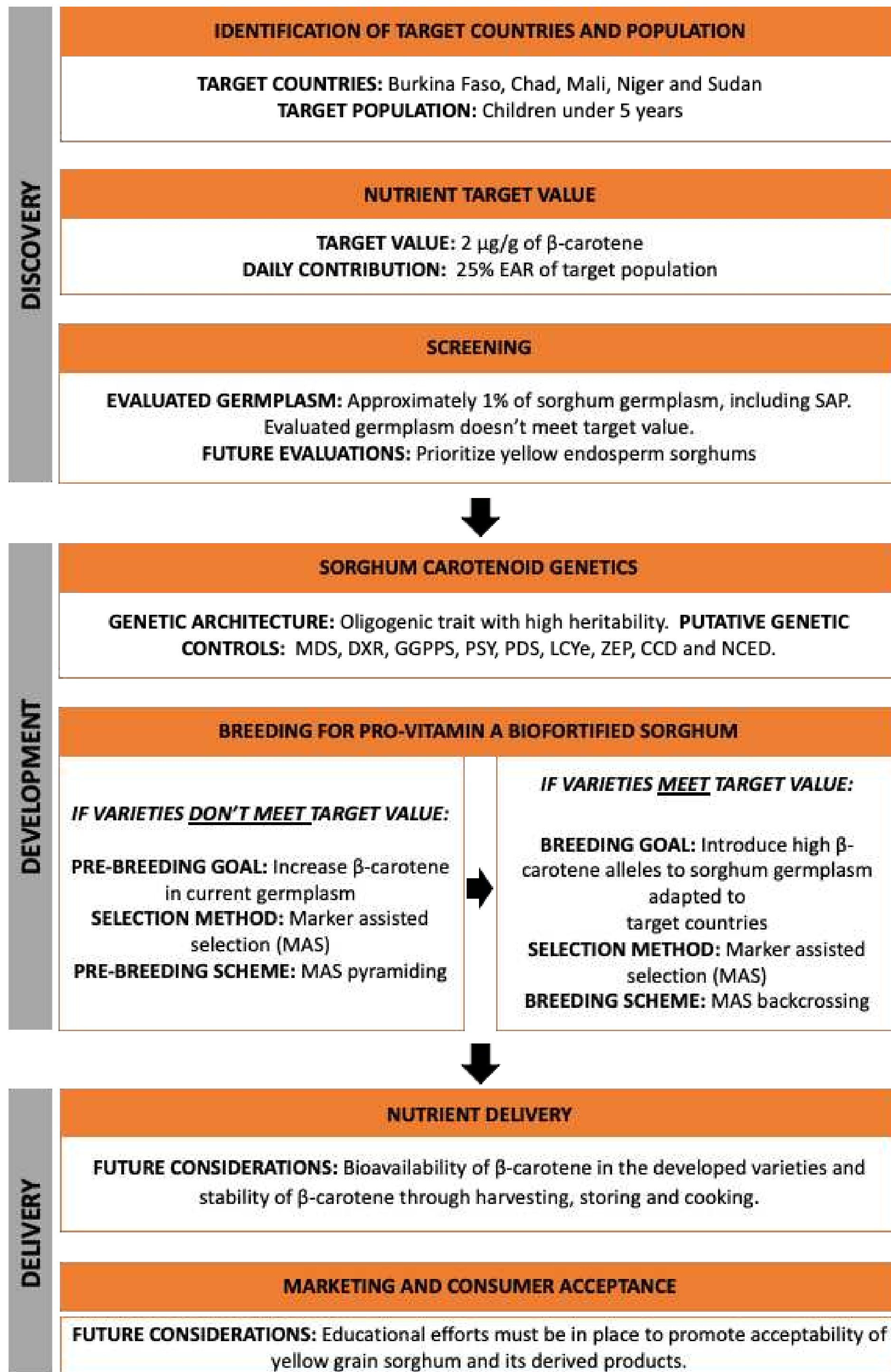


Figure 1.1 Sorghum biofortification pipeline and status.
Based on the workflow from HarvestPlus (Bouis et al., 2011))

Crop biofortification can be divided into three phases: discovery, development, and delivery (Bouis et al., 2011). Perhaps the most important factor for successful biofortification occurs at the beginning stages of the program when the target crops, target population, and target nutrient levels are established. This can be difficult because the consumption patterns in the target populations within a country are not reflected in national averages. For vitamin A biofortification, however, the main challenges tend to be in the development phases, because, even though the biosynthesis pathway of carotenoids is almost completely elucidated (Figure 1.2), there is still a lack of knowledge of the pathway regulation, carotenoid transport and storage between cells and tissues, as well as chemical stability of carotenoids and their bioavailability upon consumption (Zheng et al., 2020). This constrains the ability of researchers and breeders to identify target genes for biotechnological modification or gene variants for breeding, which would result in adequate accumulation of pro-vitamin A carotenoids in the tissues to be consumed (grain, flesh, leaves, roots, etc.) and that remain stable from the farm to the table. Non-profit research and development programs such as HarvestPlus under CGIAR lead global efforts to biofortify major staples crops (Saltzman et al., 2013). Today, joint efforts have led to the development of vitamin A biofortified varieties through conventional breeding for orange flesh sweet potato (OSP), cassava, maize, wheat and banana/plantains (Saltzman et al., 2013; Shwetha et al., 2020); and through genetic engineering for rice, maize, wheat, potato, sweet potato, banana/plantains, sorghum, and cassava (Saltzman et al., 2013; Shwetha et al., 2020; Zheng et al., 2020).

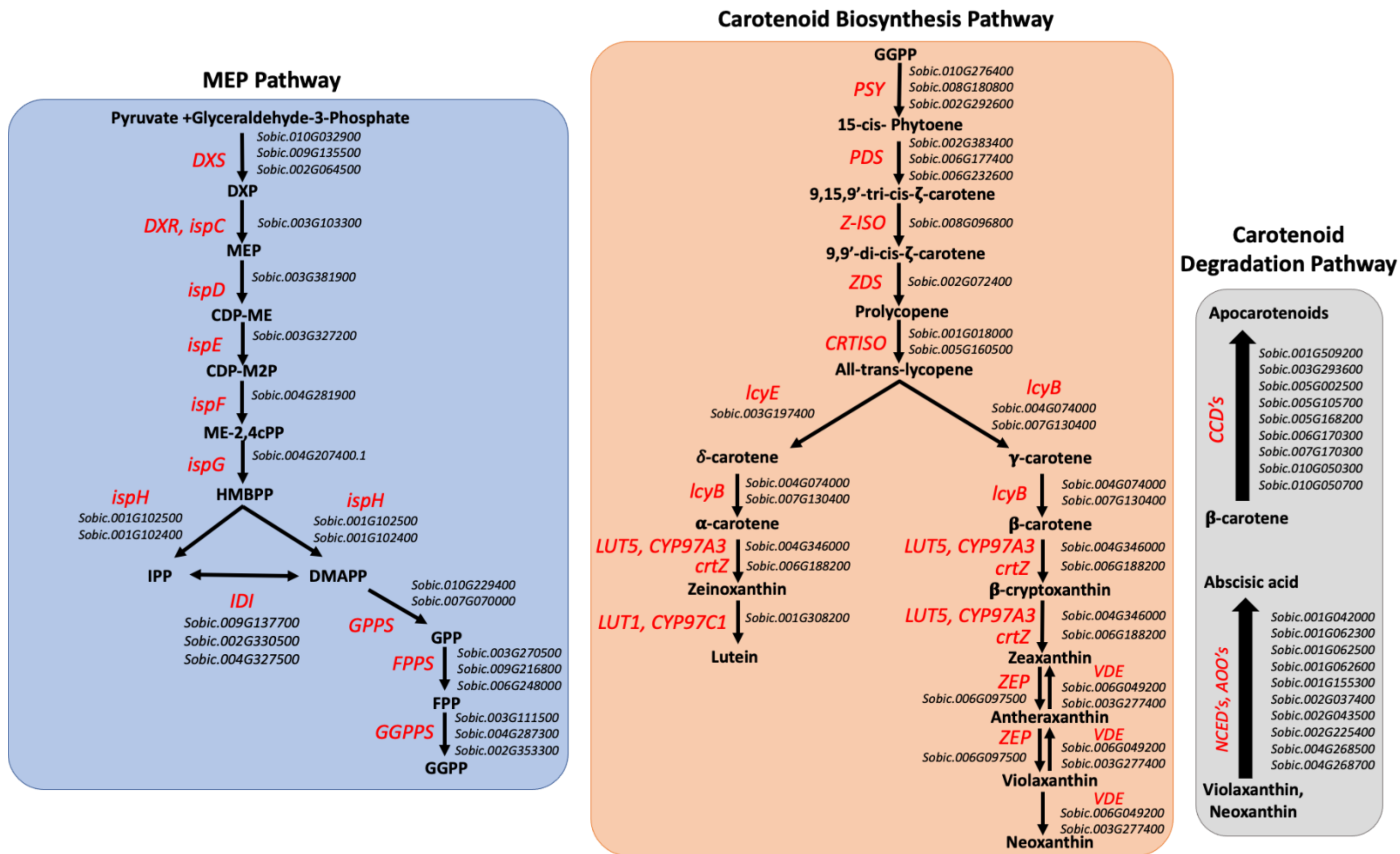


Figure 1.2 Carotenoid precursor MEP pathway, carotenoid biosynthesis and carotenoid degradation pathways. Sorghum putative genes are highlighted for each of the enzymatic reaction.

After development, however, several challenges remain to successfully deliver biofortified crops to farmer fields, and some of the vitamin A biofortified crops have failed in this phase. OSP delivery has been highly successful and can highlight some underlying factors that must be incorporated in every vitamin A biofortified crop delivery effort for it to be impactful (Hotz, Loechl, de Brauw, et al., 2012; Hotz, Loechl, Lubowa, et al., 2012). First, agronomic performance must be maintained or improved in the biofortified variety to ensure willingness in adoption by farmers. OSP relied on farmer participation through the breeding process and varietal selection. Second, consumer-preferred traits must be maintained in the biofortified variety to gain acceptance by end-users. Traditional processing and cooking of the biofortified variety must produce a product that is of the same quality as food made with traditional varieties, while also maintaining biologically relevant concentrations of pro-vitamin A carotenoids. OSP maintained consumer-preferred traits and was able to retain between 75-80% of β-carotene through the traditional drying and cooking processes. Lastly, vitamin A biofortification will always require behavioral changes from consumers, because the color of the biofortified crop will change, traditionally from white to yellow or orange, which can be met with consumer resistance. Therefore, delivery must be accompanied by marketing efforts, educational campaigns, extensionists and volunteer promoters. Due to the complexity of development and adoption of biofortified crops, a crop's potential for biofortification must be carefully considered at every phase of the process to ensure it can be successful.

Sorghum is a promising target for vitamin A biofortification as it is the fourth most consumed cereal among the countries with severe VAD in South East Asia and Africa (FAOSTAT, 2021; World Health Organization, 2009). Similar to other cereals, the sorghum

nutritional profile is characterized by high concentrations of starch, which can provide sufficient calories, but low concentrations of micronutrients, leaving the communities that rely on it as a staple crop susceptible to deficiencies (Boyles et al., 2017; Cruet-Burgos et al., 2020; Fernandez et al., 2008; Rhodes et al., 2017; Shakoor et al., 2016). In rural communities, sorghum is consumed in higher quantities than in urban communities (Bhagavatula et al., 2013), thus a biofortified sorghum can significantly impact the vitamin A status in these communities, where the prevalence of VAD tends to be higher and the accessibility to other public health measures is limited. A biofortified sorghum could also benefit specific target countries with severe VAD in the Sahel regions such as Niger, Sudan and Chad, as sorghum can represent as much as 50% of their caloric intake (Mundia et al., 2019). Most consumption methods rely on the production of sorghum flour (ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), 1982), which would suggest that flour fortification could be also beneficial. However, the benefits to the most vulnerable communities are limited, because rural and poor communities have limited access to fortified foods and fortified sorghum flour has a short shelf life (ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), 1982), thus vitamin A biofortified sorghum could have a higher impact than fortified sorghum products.

Sorghum biofortification through genetic modification has been carried out by the Africa Harvest Biotech Foundation International in the *African Biofortified Sorghum (ABS)* project funded by The Bill and Melinda gates foundation in 2005 (Zhao, 2007). The objectives were to improve the nutritional quality of sorghum by increasing limiting amino acids, pro-vitamin A carotenoids, iron, and zinc, as well as improving overall digestibility through biotechnology. To increase pro-vitamin A carotenoids, they co-expressed two

biosynthesis genes, *phytoene synthase 1 (PSY1)* from maize and the *carotenoid desaturase (CRT1)* from *Pantoea ananatis*, as well as the *homogentisate geranylgeranyltransferase (HGGT)* from barley to increase carotenoid stability, under a sorghum endosperm-specific promoter (Che et al., 2016). They were able to achieve a concentration of β-carotene; of 12.30 μg/g, which they estimated could provide 90% of the estimated average requirement (EAR) of vitamin A for children under 3 years when consumed fresh and 20% EAR after 6 months of storage at room temperature (Che et al., 2016). Unfortunately, there are no reports that this biofortified sorghum has yet to be utilized. Lack of continuous funding, as well as safety and acceptability concerns of transgenics in African countries have been the main limitations for the development and deployment of the ABS project. Safety risks, primarily the possibility of gene transfer to wild species given that Africa is the center of origin of sorghum, were assessed with no major concerns being raised (Hokanson et al., 2010; Magomere et al., 2016). Also, some progress has been made towards acceptability with four (Burkina Faso, Egypt, South Africa and Kenya) of the five initial target countries for the ABS project having some legislation that allows commercialization of genetically engineered crops (Wambugu & Kamanga, 2014). However, even if legal approval is given and safety concerns are addressed, anti-GM groups, such as the African Centre for Biosafety (ACB), have a strong public and media influence that might hinder farmer and consumer acceptability of a transgenic sorghum (Wambugu & Kamanga, 2014).

Biofortification of sorghum through conventional breeding has received limited attention and should be examined as an alternative or complementary strategy to address VAD. In this review, the current status of development efforts for a biofortified pro-vitamin

A sorghum through conventional breeding is explored (Figure 1.1). Topics discussed are (1) target populations and target concentrations based on consumption patterns, (2) which germplasm has been characterized and which remains unexplored, (3) the genetics underlying carotenoids in sorghum, (4) what breeding strategies can be adopted, (5) and considerations for consumer and farmer acceptability.

DISCOVERY PHASE FOR SORGHUM PRO-VITAMIN A BIOFORTIFICATION

Target countries

Biofortification has the potential to adress nutritional deficiencies long term in a sustainable manner. To be impactful, efforts must first identify the countries and populations within them that are the most susceptible and can benefit the most from a biofortified variety. Target countries or populations must have a high incidence of the micronutrient deficiency to be addressed, as well as produce and consume large quantities of the crop to be biofortified. For a pro-vitamin A biofortified sorghum, Burkina Faso, Chad, Mali, Niger and Sudan have been identified as the top five target countries (Table 1.1). VAD is a major public health concern in these five countries, particularly in children under five who are the most affected group suffering from severe deficiencies, followed by pregnant women with moderate deficiencies (Table 1.1). In these target countries, sorghum is among the top three consumed cereals, providing over 300 kcal/capita daily. Sorghum imports are low (Table 1.1), indicating the majority of the sorghum consumed comes from local production, so the introduction of a biofortified variety to these target countries will directly impact their vitamin A status.

Several public health measures are currently in place to reduce VAD in the target countries. Vitamin A supplementation is the most common intervention (Table 1.1). Food fortification of oil, sugar and cereal flours has also been widely adopted, with mandatory fortification implemented in several of the target countries. Biofortification, on the other hand, has been adopted by only two target countries: OSP has been released in Burkina Faso and yellow maize in Mali. Introducing a biofortified pro-vitamin A sorghum variety into target countries could have a significantly greater impact than current efforts alone. Sorghum is an important part of the diet in the target population and it has the potential to impact rural households that have not been reached by existing methods, so it can be used to provide a sustainable long-term vitamin A source. The identification of target countries, the target populations within them, and their consumption patterns, helps focus biofortification efforts to generate varieties that can address their specific nutritional needs.

Table 1.1 Sorghum consumption patterns, prevalence of vitamin A deficiency (VAD) and current public health measures in target countries for a vitamin A biofortified sorghum.

	Sorghum Consumption [1]				Prevalence of VAD [2] (percentage of the population)		Current public health measures
Target Country	Cereal Ranking (Human Consumption)	g/capita /day	kcal/capita /day	Annual Imports (ktonne)	Preschool age children	Pregnant Woman	
Burkina Faso	2nd	158	438	0	54.3	16.7	Vitamin A supplementation programs[3,4], Mandatory oil fortification[5], Biofortified Orange Sweet Potato[3]
Chad	1st	146	426	57	50	17.1	Vitamin A supplementation programs[3]
Mali	3rd	136	396	0	58.6	16.7	Vitamin A supplementation programs, Biofortified maize[3], Mandatory vegetable oil fortification[5]
Niger	2nd	131	340	16	67	14.7	Vitamin A supplementation programs, Mandatory fortification of sugar, wheat flour, maize flour [3] and vegetable oil [3,5]
Sudan	1st	201	588	99	27.8	16.1	Vitamin A supplementation programs, micronutrient powders [3]

[1] (FAOSTAT, 2021); [2](World Health Organization, 2009); [3] (Wirth et al., 2017); [4] (Kargougou et al., 2015); [5] (Grant et al., 2018)

Nutrient target value

The nutrient target value for a biofortified crop is the amount of micronutrient that must be present in the biofortified varieties to provide a biologically relevant amount of micronutrient to the target populations. Determination of this level is far from simple and it depends upon a number of factors, such as consumption patterns of the crop in the target populations, estimated nutritional requirements in the population, bioavailability or conversion rate of the nutrient, as well as stability of the nutrient through storage, processing and cooking.

Consumption patterns in target countries

Biofortification efforts must rely on availability of data to establish target values. Today, FAOSTAT is the most comprehensive tool harboring data for global agricultural and food systems. However, the available data is collected from annual country level surveys, which might not reflect consumption patterns in the communities within a country that are most susceptible to nutritional deficiencies. Country-level data might fail to account for patterns of consumption in countries with significant agroecological and climatic variation, resulting in regional variation in staple crops, with some populations relying on sorghum as a staple crop and other populations relying on other crops. Also, country-level data might underestimate consumption of staple crops in rural communities, due to the higher proportion of people living in urban areas (Asare-Marfo et al., 2013). Country-level data can be a good starting point to establish which countries could benefit the most from a biofortified crop, however, additional community level surveys can ensure that the efforts and resources are placed in the communities with higher prevalence and susceptibility to nutrient deficiencies.

Estimated average requirements (EARs)

The nutrient target value of a biofortified crop is intended to satisfy at least a part of the nutritional requirements of the target population; however, determining the exact nutritional requirements is very hard to establish. Therefore, efforts to calculate a target value rely on theoretical estimates of nutrient requirements for the target population, and the EAR is commonly used. EARs are estimates provided by the Food and Nutrition Board of the Institute of Medicine of the amount of a nutrient needed to satisfy the daily requirement for half of the healthy individuals in an age and/or sex group. For children under 5 years, the age group with the highest incidence and severity of VAD in the target countries, the EAR for vitamin A ranges from 210 to 275 µg of vitamin A equivalents per day, depending on their specific age group (Institute of Medicine, 2001). For crop biofortification, the nutrient target values are established to satisfy a percentage of the EAR as it is assumed that other foods in the diet will also provide the nutrient. Taking into consideration current efforts to alleviate VAD implemented in the target countries, it can be hypothesized that a pro-vitamin A biofortified sorghum that provides 25% EARs (68.75 µg of vitamin A equivalents per day) for children under 5 years, can be sufficient to close the gap and achieve nutrient sufficiency. Determining the percent EAR is the first step in calculating the target nutrient value in the crop, but other factors such as stability and bioavailability must be considered as well.

Stability of sorghum carotenoids

The target value for biofortification must take into consideration potential post-harvest carotenoid degradation, therefore, variation in carotenoid stability during storage

and food preparation must be examined. The carotenoid backbone structure is composed of conjugated double bonds, which are susceptible to oxidation induced by heat, light, oxygen, and acids. Isomerization of carotenoids has also been found to affect their stability. Carotenoids naturally occur as trans-isomers in plants, but can be isomerized to cis-isomers during food processing, resulting in reduction of stability for all carotenoids (Honda, 2021). Oftentimes, this means that the amount of carotenoids present in the biofortified crop will be significantly reduced by the time it reaches the table. For the determination of target values, then, stability of carotenoids in the biofortified varieties must be taken into consideration. Sorghum, as with most cereals, is often dried to reduce grain moisture, stored for long periods of time and cooked by multiple methods (soaked, fermented, baked, boiled, etc.); exposing the grain or flour to the same factors that degrade carotenoids. There is still a lot to learn about the stability of sorghum carotenoids through processing and storage, but studies have shown that soaking, dry heating (baking or as a method to reduce grain moisture), and wet heating (boiling) significantly reduce the concentration of carotenoids (Afify et al., 2012; Cardoso et al., 2014, 2015). Storage of biofortified transgenic sorghum under room temperature and light exposure was also shown to reduce β-carotene concentrations (Che et al., 2016). Through these studies it was also observed that stability of sorghum carotenoids is genotype dependent. Therefore, at the end of the development process it is imperative that the biofortified varieties must be tested under culturally relevant conditions to ensure the target value results in sufficient pro-vitamin A carotenoids in the prepared meals.

Sorghum carotenoids bioavailability

Vitamin A bioavailability is an important consideration when establishing the target value for a biofortified variety. In humans, retinol is the primary form of vitamin A, however, plants can not directly produce it. Instead, plants produce pro-vitamin A carotenoids such as β-carotene, β-cryptoxanthin and α-carotene that can be absorbed and converted to retinol by the human body. Bioavailability—the fraction of carotenoid that is absorbed and available for utilization—is one of the main factors that determines the amount of carotenoids available for retinol conversion. Carotenoid bioavailability is often estimated from animal models and is dependent on the amount of carotenoid present, the type of carotenoid, localization within the food/plant matrix, the type and amount of lipids consumed, and the food processing and preparation methods adopted (Kean et al., 2011). Studies in sorghum are limited, but they have shown that sorghum carotenoids are bio-accessible in *in-vitro* models (Kean et al., 2011; Lipkie et al., 2013; You, 2016). These findings support that a pro-vitamin A biofortified sorghum could provide an important source of vitamin A for the target populations. A recent estimate for the bioconversion of sorghum derived β-carotene to retinol is from the transgenically biofortified sorghum with conversion values of 4.5 μg of β-carotene to 1 unit of retinol (You, 2016). As a starting point for determining the target value of a pro-vitamin A biofortified variety, a conversion rate of β-carotene to retinol of 4.5 : 1 can be assumed, under the hypothesis that a pro-vitamin A biofortified variety developed through breeding will have a similar conversion rate as the biofortified variety developed through transgenics. Although previous studies are a good starting point for determining a biofortification target value, bioavailability and bioconversion rates can vary considerably between genotypes, therefore, a pro-vitamin

A biofortified sorghum variety will have to undergo its own testing to determine accurate estimates. However, due to the complexity and cost of these *in-vitro* studies, they are often conducted at the end of the development phase when only a few candidate varieties remain. Sorghum carotenoid bioavailability of biofortified varieties must be assessed after development to ensure the target value satisfies the nutritional requirements of the target population.

Based on currently available information on sorghum consumption patterns and the bioavailability and stability of sorghum carotenoids, an initial target value can be determined, to be revised as field, lab, and clinical studies provide more information. To satisfy 25% of the EAR for children under 5 years of age in the target countries, a biofortified sorghum should provide ~69 ug of vitamin A equivalents per day. Using the sorghum estimated bioconversion rate of 4.5 : 1 (You, 2016), and taking into consideration the average daily per capita sorghum consumption among the target countries of ~150 g, a biofortified sorghum should contain 2 μg of β-carotene per gram of sorghum. After reaching the 25% target, the next breeding target is 50% of the EAR, which would be a biofortified sorghum containing 4 μg of β-carotene per gram of sorghum.

Sorghum germplasm screening

With the determination of the target population and target values for a vitamin A biofortified sorghum, concentrations of carotenoids in sorghum germplasm must be examined to determine if any variety meets the target value or if crop improvement is necessary. For a pro-vitamin A biofortified sorghum for the target countries, a sorghum variety must contain at least 2 μg/g of β-carotene. Germplasm evaluations for carotenoid

content in sorghum date back to almost a century ago when corn started to be supplemented or replaced by grain sorghum in animal feeds in the southwest region of the United States. Yellow grain sorghum quickly gained attention as studies suggested that their feeds had higher vitamin A content, resulting in better performance of the animals (Smith, 1930, p. 19). Over the next few decades, it was determined that carotenoids were indeed present in sorghum grain, although their concentrations were below commonly grown yellow maize varieties, and that lutein, zeaxanthin, and β-carotene were the primary carotenoid compounds. Across studies, the total carotenoid ranged from 1.1 to 8 μg/g and β-carotene concentrations ranged from 0.3 to 3.23 μg/g (Blessin et al., 1958, 1962; Fu, 1960; Gross & Heller, 1943; Suryanarayana Rao et al., 1967; Worzella et al., 1965). Based on the concentrations reported in these studies, there were sorghum varieties that already contain the target value of 2 μg/g of β-carotene needed in a biofortified variety for the target countries. Unfortunately, tracing back varieties evaluated in these studies in the 50's is not possible due to differences in naming conventions and lack of a standardized naming convention at the time, or because the historic records are no longer available. However, this early research can be utilized to establish four main points of sorghum carotenoids: 1) carotenoids are present in sorghum grain; 2) lutein, zeaxanthin and β-carotene are the main carotenoids; 3) there is variation in concentration of carotenoids among varieties; and 4) there is a positive correlation between yellow grain and carotenoid concentration.

More recently, a handful of studies have evaluated sorghum germplasm with the goal of exploring its potential for carotenoid biofortification. The first of these studies continued exploring carotenoid content in sorghum with emphasis on yellow endosperm

varieties. Across these varieties, twenty six yellow endosperm sorghum varieties were evaluated for total carotenoids and/or β-carotene content, with total carotenoids ranging from 0.173 to 1.70 μg/g (Fernandez et al., 2008; Kean et al., 2008) and β-carotene from 0.01 to 1.13 μg/g (Fernandez et al., 2008; Kean et al., 2008; Reddy et al., 2005). Even though the varieties evaluated are yellow endosperm, as in earlier studies, the carotenoid content is notably lower, with β-carotene content below the estimated target value. Sorghum is a diverse crop with collections in the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the U.S. National Plant Germplasm System (NPGS) listing over 40,000 accessions. The limited number of accessions evaluated in the study might mean that carotenoid diversity in sorghum has not been fully captured, thus germplasm with higher carotenoid content might be possible. Supporting this, the most recent study on sorghum carotenoids evaluated the largest and most diverse sorghum germplasm to date and detected higher total carotenoid content in sorghum than has previously been found (Cruet-Burgos et al., 2020). The germplasm consisted of 403 accession, of which 321 accessions were from the sorghum association panel (SAP), which contains geographic and genotypic diversity of all African regions (Casa et al., 2008), as well as 82 photoperiod insensitive yellow grain or yellow endosperm varieties. Total carotenoid content ranged from 0.50 μg/g to 19.3 μg/g, with β-carotene concentrations ranging from 0.00 to 0.80 μg/g (Cruet-Burgos et al., 2020). Interestingly, despite higher total carotenoid content, the β-carotene concentration detected is similar to those reported in other recent germplasm evaluations (Fernandez et al., 2008; Kean et al., 2008; Reddy et al., 2005), which is insufficient to provide the necessary vitamin A to target populations. The discrepancy in β-carotene content between current germplasm evaluations (Cruet-Burgos et al., 2020; Fernandez

et al., 2008; Kean et al., 2008; Reddy et al., 2005) and earlier evaluations (Blessin et al., 1958, p. 19, 1962, p. 19; Fu, 1960; Gross & Heller, 1943; Suryanarayana Rao et al., 1967; Worzella et al., 1965) could be due to multiple factors, including differences in experimental designs such as location where the accessions were grown, as well as the methodology used for carotenoid quantification. However, this could also mean that recent studies have not been able to capture the entirety of carotenoid variation present in sorghum germplasm.

Based on recent studies, current β-carotene concentrations in sorghum grain do not meet target values, therefore two alternatives remain: (1) further explore germplasm for accession with higher concentrations, and/or (2) employ breeding to increase the concentration. Expanding germplasm evaluations could be the most feasible alternative, as only about 1% of sorghum accessions present in the collections have been phenotyped for carotenoids,and it is a faster alternative than breeding, which could take decades. However, large-scale evaluations can be impractical because phenotyping carotenoids remains a complex and expensive process, using high-performance liquid chromatography (HPLC) for accurate measurements. Yellow endosperm varieties have higher carotenoid content than their white or mixed endosperm counterparts (Blessin et al., 1958, p. 19, 1962, p. 19; Fernandez et al., 2008; Fu, 1960; Gross & Heller, 1943; Kean et al., 2008; Suryanarayana Rao et al., 1967; Worzella et al., 1965), therefore, future germplasm evaluations can focus on accessions that contain yellow endosperm. In the NPGS, there are 381 sorghum accessions listed as yellow endosperms that could be evaluated for carotenoid content. Unfortunately, this approach has the limitation that most of the accessions available in genbank collections have not been characterized for

endosperm color, thus the prevalence of yellow endosperm sorghum in the collection is unknown.

Alternatively, accessions can be narrowed down by focusing on those originating from a region where yellow endosperm is a valued trait. For example, there are a few lines of evidence supporting a relationship between yellow endosperm varieties of Nigerian origin and higher carotenoid content. In early research the high carotenoid germplasm consisted of yellow endosperm lines derived from crosses between USA lines and Nigerian yellow endosperm lines (Blessin et al., 1958; Fu, 1960; Worzella et al., 1965). Nigerian yellow endosperm varieties were also found to contain higher concentrations of β-carotene when compared to those of different origins (Suryanarayana Rao et al., 1967). The high incidence of Nigerian germplasm and the high concentrations reported could be a result of positive selective pressure for yellow endosperm as they are preferred in the region due to their high yields and drought tolerance (Sun & Hu, 2013). Supporting selective pressure, genetic diversity studies in yellow endosperm varieties from Niger and Nigeria showed that they were genetically different from yellow endosperm accessions from other origins (Salas Fernandez et al., 2009). This could mean that they harbor unexplored genetic and phenotypic diversity thus representing promising candidates for evaluations. Is important to note that although in the literature there has been a high incidence of yellow endosperm germplasm from Nigerian and West African origin, there might be other regions where yellow endosperm is a valued trait leading to additional genetic diversity that is worth exploring. Certainly, germplasm evaluations could be the fastest alternative to reach the target values if they

already exist in the species. If current carotenoid content in sorghum grain does not meet the target values, breeding must be used to increase concentrations.

THE DEVELOPMENT PHASE FOR A VITAMIN A BIOFORTIFIED SORGHUM

Sorghum carotenoid genetics

Sorghum carotenoid inheritance

Breeding has been used to successfully develop staple crop varieties that have high pro-vitamin A concentrations and acceptable agronomic characteristics (Low et al., 2017; Zhang et al., 2012; Zunjare et al., 2018). For breeding to be possible in sorghum carotenoid biofortification, carotenoid variation must have a genetic component that can be inherited. Progeny studies derived from yellow endosperm and non-yellow endosperm parents demonstrated that both yellow endosperm and β-carotene content where inherited in progenies, indicating they had both a genetic component and that they were correlated (Fernandez et al., 2008; Gorbet, 1971; Worzella et al., 1965). Furthermore, broad sense heritability estimates for endosperm color, total carotenoids, lutein, zeaxanthin and β-carotene indicate that carotenoids are highly heritable traits, with recombinant inbred line (RIL)-based estimates of 91%, 91%, 84%, 89% and 91%, respectively (Fernandez et al., 2008). High heritability estimates for carotenoid compounds in sorghum are consistent with findings in other cereals crops, such as maize (Diepenbrock et al., 2021) and wheat (Blanco et al., 2011), suggesting there might be a conserved genetic component. Phenotypic distribution of progenies from yellow endosperm and non-yellow endosperm parents, indicates that carotenoid content and endosperm color follows a quantitative inheritance with more than one gene contributing

to the traits (Fernandez et al., 2008; Gorbet, 1971; Worzella et al., 1965). Quantitative trait loci (QTL) identified in recombinant inbred lines as well as in a diverse germplasm suggest that endosperm color, total carotenoids, lutein, zeaxanthin and β-carotene content in sorghum grain are oligogenic in nature—traits controlled by a moderate number of moderate effect genes—(Cruet-Burgos et al., 2020; Fernandez et al., 2008). Carotenoid inheritance in maize kernels has also been identified as an oligogenic trait (Suwarno et al., 2015; Venado et al., 2017; Wong et al., 2004, p. 20; Xu et al., 2019), with a recent report identifying eleven genes that primarily control carotenoid variation in maize (Diepenbrock et al., 2021). The oligogenic inheritance of carotenoids in sorghum implies that marker-assisted selection (MAS) is a viable method to accelerate biofortification breeding efforts. The alleles contributing to high carotenoid content in sorghum, therefore, must be identified and characterized in order to develop markers with good predictive values to be used in breeding programs.

Genetic control of carotenoids in sorghum: Precursors genes

Genetic controls for carotenoid biosynthesis could be genes involved in preceding or competing pathways, as they could determine the amount and rate of precursor molecules entering the carotenoid pathway. The methylerythritol phosphate (MEP) pathway might be the most influential preceding metabolic route, as it produces geranyl geranyl diphosphate (GGPP), the immediate precursor compound for the carotenoid biosynthetic pathway. Genomic studies of sorghum grain carotenoid have identified associations between zeaxanthin and β-carotene concentration and genes involved in the MEP pathway (Cruet-Burgos et al., 2020). Zeaxanthin concentration was associated

with a distal region of chromosome 4 approximately 46 kb away from Sobic.004G281900, annotated as 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase (MDS, ispF). MDS is an enzyme within the MEP pathway catalyzing the conversion of 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-M2P) to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) (Cruet-Burgos et al., 2020). This precursor gene has not been previously identified in genomic studies in other species, which could mean its contribution to carotenoid variation might be exclusive to sorghum. Zeaxanthin concentration was also associated with a region in chromosome 3, 476 kb away from Sobic.003G103300, annotated as a deoxyxylulose reductoisomerase (DXR). DXR is also an enzyme in the MEP pathway, catalyzing the first committed step of the MEP pathway toward isoprenoid synthesis by rearrangement and reduction of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP). DXR has been associated with carotenoid concentration in genomic studies in wheat (Zhai et al., 2018) and Arabidopsis (Gonzalez-Jorge et al., 2014), and its expression in Arabidopsis has demonstrated a direct impact on total carotenoid concentrations (Carretero-Paulet et al., 2006; Estévez et al., 2001). In chromosome 2, an association was identified for β-carotene concentration and a region approximately 98 kb from Sobic.002G353300, a geranyl geranyl diphosphate synthase (GGPPS). The GGPPS catalyzes the condensation of two molecules of geranyl geranyl diphosphate (GGPP) to GGPP, the precursor molecule for the carotenoid pathway. GGPPS has not been identified in other cereal crops, but it has been identified as associated with carotenoid variation in Arabidopsis (Gonzalez-Jorge et al., 2016). Although further confirmation of these associations of the MEP pathway genes and zeaxanthin and β-carotene concentration is

needed, results suggest that increasing the flux through the MEP pathway could increase the accumulation of β-carotene in sorghum grain. Competing pathways that use the same precursor molecules as the carotenoid pathway, such as the terpenoid biosynthesis pathways, could also be influencing carotenoid concentrations, however, no associations in competing pathways have been identified in genomic studies yet. Targeting precursor pathways for carotenoid biofortification could result in an increased accumulation of total carotenoids rather than specific pro-vitamin A carotenoids such as β-carotene. However, perhaps a combination of positive alleles from precursor pathways and positive alleles from carotenoid pathway genes could be used to increase total carotenoids, as well as to increase pro-vitamin A carotenoids.

Genetic control of carotenoids in sorghum: Biosynthesis genes

Genes directly involved in carotenoid biosynthesis might be the best prospects to not only increase total carotenoid concentrations, but to also specifically target an increase of the pro-vitamin A carotenoids β-carotene, β-cryptoxanthin and α-carotene (Table 1.2). PSY, the enzyme catalyzing the first committed step in the carotenoid biosynthesis pathway, has been associated with variation observed for lutein, zeaxanthin and β-carotene in a RIL population (Fernandez et al., 2008). Based on sequence similarity, sorghum possesses three PSY genes, Sobic.002G292600, Sobic.010G276400, Sobic.008G180800, which have also been identified in rice (Welsch et al., 2008) and maize (Li et al., 2008). The first sorghum PSY gene was associated with the concentration of the three main carotenoids—lutein, zeaxanthin and β-carotene—the second PSY gene was only associated with lutein and the third PSY gene has not been

identified in any genomic studies to date. PSY has also been identified in genomic studies in maize (Diepenbrock et al., 2021). Interestingly, PSY genes might be expressed in a tissue specific manner, as in other cereals (Gallagher et al., 2004; Paterson et al., 2009), making an endosperm-specific PSY a potentially good target to increase total carotenoid concentration in the grain. Phytoene desaturase (PDS) catalyzes the introduction of two double bonds to phytoene to generate ζ-carotene via phytofluene as an intermediate. Genomic studies have identified PDS (Sobic.006G232600) as underlying variation in concentrations of zeaxanthin (Fernandez et al., 2008) and β-carotene (Cruet-Burgos et al., 2020; Fernandez et al., 2008). Identification of PDS in these two independent studies suggests that it might be a major regulator of sorghum β-branch carotenoids, particularly β-carotene. This association with β-branch carotenoids was also observed in maize, where PDS was associated with the natural variation in β-carotene and β-cryptoxanthin (Diepenbrock et al., 2021), further supporting the potential of PDS as breeding target to increase pro-vitamin A carotenoids. After PDS, δ-carotene desaturase (ZDS) mediates the next steps in the carotenoid biosynthesis pathway, catalyzing two desaturations of δ-carotene, with δ-neurosporene as an intermediate, to generate δ-lycopene. In sorghum genomic studies, this single copy gene (Sobic.002G072400) has been associated with zeaxanthin variation (Cruet-Burgos et al., 2020). ZDS has also been found to contribute to the natural variation of lutein, zeaxanthin and β-carotene in maize F2:3 populations (Venado et al., 2017; Wong et al., 2004). As an early pathway enzyme, it is interesting that in sorghum it has only been associated with the concentration of zeaxanthin and not the other carotenoids. Lycopene epsilon cyclase (lcyE) is perhaps one of the most promising enzymes in the vitamin A biofortification efforts as it catalyzes the conversion

of lycopene to α-carotene. Despite α-carotene having some pro-vitamin A activity, shifting the pathway towards the β-branch to increase β-carotene accumulation is generally accepted as the main target for biofortification, therefore, lcyE could play an important role. In sorghum, lcyE (Sobic.003G197400) has been associated with variation in lutein concentration in a RIL population (Fernandez et al., 2008). In maize however, lcyE has been identified as one of the main genes controlling carotenoid variation and it has been associated with all the main carotenoids, as well as the ratios between the α and β branch (Azmach et al., 2018; Diepenbrock et al., 2021; Harjes et al., 2008; Owens et al., 2014; Suwarno et al., 2015; Venado et al., 2017; Xu et al., 2019). Perhaps the lack of association with other carotenoids in sorghum could be explained with lack of genetic diversity in the gene among the sorghum germplasm evaluated to date. Therefore, despite the lack of association of other carotenoids, lcyE in sorghum remains a candidate for biofortification as a low activity allele has the potential to shift the pathway toward the β branch where the most important pro-vitamin A carotenoids are synthesized. Zeaxanthin epoxidase (ZEP) catabolizes the last two steps in the β branch of the carotenoid pathway by converting zeaxanthin to violaxanthin with antheraxanthin as an intermediate. In a genome wide association study (GWAS) in sorghum it was associated with the natural variation in zeaxanthin concentration (Cruet-Burgos et al., 2020). Interestingly, this was the most significant association and the marker was located inside the ZEP gene (Sobic.006G097500). Associations of ZEP and variation in zeaxanthin have also been detected in maize (Azmach et al., 2018; Diepenbrock et al., 2021; Owens et al., 2014; Suwarno et al., 2015). Taken together, there is strong evidence suggesting that ZEP is the main controller of zeaxanthin concentration in both sorghum and maize.

Despite this strong association ZEP might not be the best target for biofortification approaches because, ZEP is a single copy gene in the sorghum genome and its products are precursors of abscisic acid (ABA) biosynthesis. ABA together with gibberellins (GA) control the equilibrium between dormancy and germination therefore, function must be maintained to ensure a timely germination of seeds (Steinbach et al., 1997). However, other genetic controls involved in the degradation of carotenoids to various other metabolic products could be good targets.

Genetic control of carotenoids in sorghum: Catabolism genes

Genes involved in the catabolism and degradation of carotenoids are potential targets to increase carotenoid concentrations. A QTL mapping study in a sorghum RIL population identified two such genes that were associated with lutein variation (Fernandez et al., 2008). Sobic.001G509200 and Sobic.001G155300 have been annotated as a *carotenoid cleavage dioxygenase (CCD)* and a *9-cis-epoxycarotenoid dioxygenase (NCED)*, respectively. These two enzymes are members of the family of carotenoid cleavage oxygenases (COO), which break the conjugated double bonds in carotenoids to produce various apocarotenoids (Priya et al., 2019). Although much is yet to be learned about the differences in catabolism between the CCD and NCED across species, studies in Arabidopsis suggest that they work in a tissue specific manner. NCEDs catabolize carotenoids for the production of ABA, while CCDs catabolize carotenoids to produce a wider range of apocarotenoids, including strigolactones and compounds involved in aroma and color (Ohmiya, 2009). Based on the presence of multiple CCDs and NCEDs in other plants (Ohmiya, 2009; Priya et al., 2019), it could be hypothesized that sorghum

also contains multiple CCDs and NCEDs. If this is true, and they have low substrate specificity, then they would unlikely be a good target for biofortification.

MAS may be a good breeding strategy for developing pro-vitamin A biofortified sorghum varieties. Many association studies have identified markers near genes in the carotenoid precursor MEP pathway, carotenoid biosynthesis and degradation pathways that underlie carotenoid variation. Validity of these genetic controls must still be confirmed, however, they are promising candidates to target in breeding programs.

BREEDING FOR A VITAMIN A BIOFORTIFIED SORGHUM

The current concentration of β-carotene, the main pro-vitamin A carotenoid in sorghum grain, does not meet the 25% EAR target value of 2 µg/g.. Breeding has the potential to increase concentrations of β-carotene in sorghum grain, but only if genetic diversity exists in the species. However, modern sorghum germplasm evaluations so far have failed to discover an accession that has sufficient β-carotene concentration (Cruet-Burgos et al., 2020; Fernandez et al., 2008; Kean et al., 2007; Reddy et al., 2005). This represents a challenge that other vitamin A biofortification efforts in crops such as maize and cassava, which have been successful, have not faced. Biofortification efforts in maize and yellow cassava had a target value of 15 µg/g of pro-vitamin A carotenoids and germplasm evaluations in both species were able to identify accessions with pro-vitamin A content of up to 19 µg/g (Harjes et al., 2008; Menkir et al., 2015; Ortiz-Monasterio et al., 2007; Patil et al., 2017; Virk et al., 2021). The absence of accessions with the target β-carotene concentration in sorghum grain raises the concern that a lack of genetic diversity will make reaching the target values through breeding impossible. However, the

sorghum germplasm that has been evaluated for carotenoids represent a very small percentage of the total diversity in the species, suggesting that there may be untapped carotenoid diversity and higher concentrations. Expanding germplasm evaluations, therefore, must be prioritized to facilitate breeding efforts. On the other hand, if germplasm evaluations do not lead to higher concentrations of β-carotene, genetic diversity in the highest β-carotene concentration accessions must be explored. It is possible that current high β-carotene lines do not possess all the positive alleles in genes contributing to carotenoid variation, and pre-breeding efforts could increase the concentrations further.

Breeding for pro-vitamin A has the added challenge of phenotyping complexity for carotenoids, thus breeding efforts have used different methods of selection with varying success. OSP biofortification efforts selected for darker orange color on the flesh because high correlations with β-carotene content exist (Virk et al., 2021). This simple phenotypic selection made the development of a biofortified OSP that exceeded target levels easier. This approach also meant that accelerated breeding could be possible without added effort in infrastructure or training, reducing the breeding cycle from 7-8 years to 4-5 years (Low et al., 2017). Phenotypic selection, however, was not successful in maize where yellow color correlates with higher total carotenoid concentrations, but not with higher pro-vitamin A concentrations (Zunjare et al., 2018). Based on current knowledge of sorghum carotenoids, it seems that phenotypic selection will not be a good strategy in sorghum either, because only intermediate correlations of yellow color and β-carotene concentrations have been observed (Fernandez et al., 2008). In maize, several alternatives to color-based phenotypic selection have been used, such as HPLC, near

infrared reflectance (NIR) and ultra performance liquid chromatography (UPLC). HPLC methods offer the most reliable phenotype, but have the downside of high cost, need for specialized equipment and training, which results in the inability to use large-scale selection in early generations, thus extending breeding cycles. NIR and UPLC, are high-throughput methods that could be used in the future for accelerated breeding, but they still lack the precision, compared to current HPLC technologies, needed to ensure successful breeding (Zhang et al., 2012), therefore alternate methods must be used.

MAS in maize has been one of the most promising phenotyping strategies for pro-vitamin A biofortification, because of its low cost and ability to handle large-scale evaluations, when compared to other phenotyping methods (Andersson et al., 2017). Markers targeting the β-Carotene Hydroxylase 1 (crtRB1) and the lcyE genes were used for marker-assisted backcrossing with an approximate 4.5 fold increase in pro-vitamin A carotenoid compared to the recurrent parent, and a 89-93% of the recurrent genome was recovered (Zunjare et al., 2018). MAS has not been widely used in sorghum yet, but it has demonstrated potential in introducing resistance to striga in grain sorghum (Afolayan et al., 2019), as well as bmr6 for decreased lignin content in biomass sorghum (da Silva et al., 2020). If carotenoids in sorghum are truly oligogenic, marker-assisted backcrossing and/or marker-assisted pyramiding targeting genes identified in genomic studies (Table 1.2) can be used to develop high pro-vitamin A sorghum lines adapted to the target countries. For MAS to be successful in accelerating breeding efforts to develop a pro-vitamin A biofortified sorghum variety, further studies must evaluate markers and their linkage to the causative gene, false positive rates, predictive ability in different genotypes and locations, among other metrics (Platten et al., 2019).

Efforts in breeding a pro-vitamin A sorghum must also consider important agronomic and quality traits that must be maintained in the developed varieties to be accepted by farmers and consumers. A participatory breeding approach can help ensure that the varieties selected meet the preferences of farmers and consumers, as well as contain the necessary pro-vitamin A content. Participatory farmer trials were adopted in the development of the biofortified OSP, and have been highlighted as one of the reasons underlying their success in adoption (Hotz, Loechl, Lubowa, et al., 2012; Hotz & McClafferty, 2007). In sorghum, this technique for varietal development and release has been employed in African countries and has served to determine the most important traits for farmers and end-users. Agronomic traits such as panicle compactness, height and resistance to lodging, striga and grain mold resistance, drought and heat tolerance, flowering time, and grain yield, have been identified by farmers as the most valued traits. Grain quality traits such as hardness, ease of polish, vitreousness, size and color have also been noted as important for farmers' acceptance. Consumer preference traits were also identified with flour quality, injera quality and porridge quality emphasized (Mulatu & Belete, 2001; Nkongolo, n.d.; vom Brocke et al., 2010). Grain color and flour quality can be problematic in the development of a biofortified sorghum because in the studies “whiteness” was noted as the preferred trait, and increases in β-carotene concentrations will result in a change of color to yellow or orange. Educational efforts alongside farmer participatory trials can help farmers and consumers value the yellow color as an added nutritional trait. MAS breeding schemes using back crossing or pyramiding could help to maintain the farmer and end-user preferred traits, as these strategies will allow breeders

to introduce alleles for increased pro-vitamin A into locally adapted and preferred varieties.

Table 1.2 Sorghum a priori candidate genes associated with lutein, zeaxanthin or β-carotene in genomic studies and their maize homologs.

Chromosome	Gene Identifier (Sorghum bicolor v3.1.1)	Putative Function	Gene Name	Association	Maize Homolog[3] (Percentage Similarity)
1	Sobic.001G509200*	Carotenoid cleavage dioxygenase	Ccd	Lutein[1]	Zm00001d048373 (95%)
1	Sobic.001G155300*	9-cis-epoxycarotenoid dioxygenase	NCED	Lutein[1]	Zm00001d033222 (93%)
2	Sobic.002G292600*	Phytoene synthase	PSY	Lutein[1], Zeaxanthin[1], β-carotene [1]	Zm00001d021410 (86%)
2	Sobic.002G072400	δ-carotene desaturase	ZDS	Zeaxanthin[2]	Zm00001d019124 (98%)
2	Sobic.002G353300	Geranylgeranyl diphosphate synthase	GGPPS	β-carotene [2]	Zm00001d006678 (95%)
3	Sobic.003G197400*	Lycopene epsilon cyclase	lcyE	Lutein[1]	Zm00001d011210 (94%)
3	Sobic.003G103300	Deoxyxylulose reductoisomerase	DXR	Zeaxanthin[2]	Zm00001d008427 (99%)
4	Sobic.004G281900	2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase	MDS, ispF	Zeaxanthin[2]	Zm00001d051458 (97%)
6	Sobic.006G232600	Phytoene desaturase	PDS	Zeaxanthin[2], β-carotene [1, 2]	Zm00001d027936 (96%)
6	Sobic.006G097500	Zeaxanthin epoxidase	ZEP	Zeaxanthin[2]	Zm00001d025545 and Zm00001d003512 (93%)
10	Sobic.010G276400*	Phytoene synthase	PSY	Lutein[1]	Zm00001d036345 (91%)

[1](Fernandez et al., 2008); [2](Cruet-Burgos et al., 2020); [3]Gene identificator based on maize reference genome Z.mays RefGen_V4; *Gene was inferred from primers sequence, gene function, chromosome location and maize genes with highest homology

FUTURE CONSIDERATIONS

Biofortification efforts do not end after varietal development. Post-breeding studies must be done to ensure a sufficient amount of nutrients reach consumer tables. A vitamin A biofortified sorghum variety must undergo studies for conversion rates of β-carotene to retinol as there is evidence the conversion efficiency is genotype dependent (Lipkie et al., 2013), and the 4.5 µg of β-carotene to 1 unit of retinol rate, used in this review, might not be true for all varieties. Additionally, stability of carotenoids through harvesting, processing, storing, and cooking must be evaluated. If a substantial portion of carotenoids are lost during this process, the target values must be adjusted to account for this loss. Alternatively, other methods such as increasing the vitamin E concentration in sorghum grain could be adopted, as there is evidence that vitamin E helps stabilize carotenoids in sorghum (Debelo et al., 2020). Other nutritional traits such as protein digestibility should be studied post-development in order to guarantee that no trade-offs exist with increased carotenoids. Results of these post-breeding studies might mean that breeding must be resumed to develop a more suitable variety.

CONCLUSION

Sorghum pro-vitamin A biofortification could positively impact millions of people currently suffering from severe VAD. Burkina Faso, Chad, Mali, Niger and Sudan are the five countries that could benefit the most due to their high sorghum consumption and the prevalence and severity of VAD. Current studies have demonstrated that carotenoids are present in sorghum grain and that they have underlying genetic controls, suggesting that biofortification through breeding is possible. However, current β-carotene concentrations

in the germplasm that has been evaluated is not sufficient to satisfy the 25% EAR target value of 2 μg/g. Expanding germplasm evaluations could facilitate breeding efforts by providing donor lines with sufficient β-carotene concentrations. Due to the oligogenic nature of sorghum grain carotenoid variation, MAS presents an opportunity to maximize genetic gains in a cost effective manner. MAS breeding schemes would also allow the maintenance of farmer and consumer quality traits that are needed for acceptability. Despite current research, sorghum carotenoids remain relatively unexplored and more genomic and breeding studies are needed to develop molecular breeding tools to increase provitamin A carotenoids in sorghum grain.

LITERATURE CITED

Afify, A. E.-M. M., El-Beltagi, H. S., El-Salam, S. M. A., & Omran, A. A. (2012). Biochemical changes in phenols, flavonoids, tannins, vitamin E, β–carotene and antioxidant activity during soaking of three white sorghum varieties. *Asian Pacific Journal of Tropical Biomedicine*, *2*(3), 203–209. https://doi.org/10.1016/S2221-1691(12)60042-2

Afolayan, G., Aladele, S. E., Deshpande, S. P., Oduoye, O. T., Nwosu, D. J., Michael, C., Blay, E. T., & Danquah, E. Y. (2019). Marker assisted foreground selection for identification of striga resistant backcross lines in Sorghum Bicolor. *Covenant Journal of Physical and Life Sciences*, *7*(1), 29–36. http://oar.icrisat.org/11384/1/1360-2360-1-SM.pdf

Allen, L., World Health Organization, & Food and Agriculture Organization of the United Nations. (2006). *Guidelines on food fortification with micronutrients*. World Health Organization ; Food and Agriculture Organization of the United Nations. http://catalog.hathitrust.org/api/volumes/oclc/152582146.html

Andersson, M., Saltzman, A., & Pfeiffer, W. H. (2017). Progress update: Crop development of biofortified staple food crops under HarvestPlus. *African Journal of Food, Agriculture, Nutrition and Development*, *17*(02), 11905–11935. https://doi.org/10.18697/ajfand.78.HarvestPlus05

Asare-Marfo, D., Birol, E., Gonzalez, C., Moursi, M., Perez, S., Schwarz, J., & Zeller, M. (2013). *Prioritizing Countries for Biofortification Interventions Using Country-Level Data*. 48.

Azmach, G., Menkir, A., Spillane, C., & Gedil, M. (2018). Genetic Loci Controlling

Carotenoid Biosynthesis in Diverse Tropical Maize Lines. *G3 Genes|Genomes|Genetics*, *8*(3), 1049–1065. https://doi.org/10.1534/g3.117.300511

Bhagavatula, S., Parthasarathy Rao, P., Basavaraj, G., & Nagaraj, N. (2013). *Sorghum and Millet Economies in Asia – Facts, Trends and Outlook.* India: International Crops Research Institute for the Semi-Arid Tropics.

Blanco, A., Colasuonno, P., Gadaleta, A., Mangini, G., Schiavulli, A., Simeone, R., Digesù, A. M., De Vita, P., Mastrangelo, A. M., & Cattivelli, L. (2011). Quantitative trait loci for yellow pigment concentration and individual carotenoid compounds in durum wheat. *Journal of Cereal Science*, *54*(2), 255–264. https://doi.org/10.1016/j.jcs.2011.07.002

Blessin, C. W., Dimler, R. J., & Webster, O. J. (1962). Carotenoids of corn and sorghum II. Carotenoid Loss in Yellow-Endosperm Sorghum Grain during Weathering. *Cereal Chemistry*, *39*, 389–392.

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

Bouis, H., Hotz, C., McClafferty, B., Meenakshi, J. V., & Pfeiffer, W. H. (2011). Biofortification: A New Tool to Reduce Micronutrient Malnutrition. *Food and Nutrition Bulletin*, *32*(1), S31–S40.

Boyles, R. E., Pfeiffer, B. K., Cooper, E. A., Rauh, B. L., Zielinski, K. J., Myers, M. T., Brenton, Z., Rooney, W. L., & Kresovich, S. (2017). Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theoretical and Applied Genetics*, *130*(4), 697–716.

https://doi.org/10.1007/s00122-016-2844-6

Cardoso, L. de M., Montini, T. A., Pinheiro, S. S., Pinheiro-Sant'Ana, H. M., Martino, H. S. D., & Moreira, A. V. B. (2014). Effects of processing with dry heat and wet heat on the antioxidant profile of sorghum. *Food Chemistry*, *152*, 210–217. https://doi.org/10.1016/j.foodchem.2013.11.106

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

Carretero-Paulet, L., Cairó, A., Botella-Pavía, P., Besumbes, O., Campos, N., Boronat, A., & Rodríguez-Concepción, M. (2006). Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Molecular Biology*, *62*(4–5), 683–695. https://doi.org/10.1007/s11103-006-9051-9

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., Franks, C. D., & Kresovich, S. (2008). Community Resources and Strategies for Association Mapping in Sorghum. *Crop Science*, *48*(1), 30–40. https://doi.org/10.2135/cropsci2007.02.0080

Che, P., Zhao, Z.-Y., Glassman, K., Dolde, D., Hu, T. X., Jones, T. J., Gruis, D. F., Obukosia, S., Wambugu, F., & Albertsen, M. C. (2016). Elevated vitamin E content improves all- *trans* β-carotene accumulation and stability in biofortified sorghum. *Proceedings of the National Academy of Sciences*, *113*(39), 11040–

11045. https://doi.org/10.1073/pnas.1605689113

Cruet-Burgos, C., Cox, S., Ioerger, B. P., Perumal, R., Hu, Z., Herald, T. J., Bean, S. R., & Rhodes, D. H. (2020). Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *The Plant Genome*, *13*(1). https://doi.org/10.1002/tpg2.20013

da Silva, M. J., Damasceno, C. M. B., Guimarães, C. T., de Oliveira Pinto, M., de Almeida Barros, B., de Souza Carneiro, J. E., Schaffert, R. E., & da Costa Parrella, R. A. (2020). Introgression of the bmr6 allele in biomass sorghum lines for bioenergy production. *Euphytica*, *216*(6), 95. https://doi.org/10.1007/s10681-020-02635-5

Debelo, H., Albertsen, M., Simon, M., Che, P., & Ferruzzi, M. (2020). Identification and Characterization of Carotenoids, Vitamin E and Minerals of Biofortified Sorghum. *Current Developments in Nutrition*, *4*(Supplement_2), 1792–1792. https://doi.org/10.1093/cdn/nzaa067_019

Diepenbrock, C. H., Ilut, D. C., Magallanes-Lundback, M., Kandianis, C. B., Lipka, A. E., Bradbury, P. J., Holland, J. B., Hamilton, J. P., Wooldridge, E., Vaillancourt, B., Góngora-Castillo, E., Wallace, J. G., Cepela, J., Mateos-Hernandez, M., Owens, B. F., Tiede, T., Buckler, E. S., Rocheford, T., Buell, C. R., … DellaPenna, D. (2021). Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *The Plant Cell*, *33*(4), 882–900. https://doi.org/10.1093/plcell/koab032

Estévez, J. M., Cantero, A., Reindl, A., Reichler, S., & León, P. (2001). 1-Deoxy-d-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid

Biosynthesis in Plants. *Journal of Biological Chemistry*, *276*(25), 22901–22909. https://doi.org/10.1074/jbc.M100854200

FAOSTAT. (2021). *Food and Agriculture Organization of the United Nations*. http://www.fao.org/faostat/en/#data/QC

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., & Kresovich, S. (2008). Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Science*, *48*(5), 1732–1743. https://doi.org/10.2135/cropsci2007.12.0684

Fu, W.-N. (1960). *AGRONOMIC CHARACTERISTICS, PROTEIN, AND CAROTENOID COMPOSITION OF SOME GRAIN SORGHUM VARIETIES, STRAINS, AND HYBRIDS---WITH EMPHASIS ON YELLOW ENDOSPERM TYPES* [Master of Science]. Oklahoma State University.

Gallagher, C. E., Matthews, P. D., Li, F., & Wurtzel, E. T. (2004). Gene Duplication in the Carotenoid Biosynthetic Pathway Preceded Evolution of the Grasses. *Plant Physiology*, *135*(3), 1776–1783. https://doi.org/10.1104/pp.104.039818

Garrett, G., Spohrer, R., & Keefe, M. (2015, November 1). *Fortifying our Future: A Snapshot on Food Fortification*.

Global Nutrition Report. (2014). *2014 Global Hunger Index: The challenge of hidden hunger* (p. 56).

Global Nutrition Report. (2021). *2021 Global Nutrition Report: The state of global nutrition*. https://globalnutritionreport.org/

Gonzalez-Jorge, S., Ha, S.-H., Magallanes-Lundback, M., Gilliland, L. U., Zhou, A., Lipka, A. E., Nguyen, Y.-N., Angelovici, R., Lin, H., Cepela, J., Little, H., Buell, C.

R., Gore, M. A., & DellaPenna, D. (2014). CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds. *The Plant Cell*, *25*(12), 4812–4826. https://doi.org/10.1105/tpc.113.119677

Gonzalez-Jorge, S., Mehrshahi, P., Magallanes-Lundback, M., Lipka, A. E., Angelovici, R., Gore, M. A., & DellaPenna, D. (2016). ZEAXANTHIN EPOXIDASE Activity Potentiates Carotenoid Degradation in Maturing Seed. *Plant Physiology*, *171*(3), 1837–1851. https://doi.org/10.1104/pp.16.00604

Gorbet, D. W. (1971). *INHERITANCE OF SOME ENDOSPERM TYPES IN SORGHUM, Sorghum bicolor (L.) Moench* [Doctor of Philosophy]. Oklahoma State University.

Grant, F., Tsang, B. L., & Garrett, G. S. (2018). Food Fortification in West Africa. Progress and lessons learned. *Scaling up Rice Fortification in West Africa*, 25.

Gross, W., & Heller, V. G. (1943). The Carotene Content of the Grain Sorghums. *Academy of Science for 1943*.

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., Stapleton, A. E., Vallabhaneni, R., Williams, M., Wurtzel, E. T., Yan, J., & Buckler, E. S. (2008). Natural Genetic Variation in *Lycopene Epsilon Cyclase* Tapped for Maize Biofortification. *Science*, *319*(5861), 330–333. https://doi.org/10.1126/science.1150255

Hokanson, K. E., Ellstrand, N. C., Ouedraogo, J. T., Olweny, P. A., Schaal, B. A., & Raybould, A. F. (2010). Biofortified sorghum in Africa: Using problem formulation to inform risk assessment. *Nature Biotechnology*, *28*(9), 900–903. https://doi.org/10.1038/nbt0910-900

Honda, M. (2021). Carotenoid isomers: A systematic review of the analysis, biological activity, physicochemical property, and methods for isomerization. In *Studies in Natural Products Chemistry* (Vol. 68, pp. 173–220). Elsevier. https://doi.org/10.1016/B978-0-12-819485-0.00002-5

Hotz, C., Loechl, C., de Brauw, A., Eozenou, P., Gilligan, D., Moursi, M., Munhaua, B., van Jaarsveld, P., Carriquiry, A., & Meenakshi, J. V. (2012). A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *British Journal of Nutrition*, *108*(1), 163–176. https://doi.org/10.1017/S0007114511005174

Hotz, C., Loechl, C., Lubowa, A., Tumwine, J. K., Ndeezi, G., Nandutu Masawi, A., Baingana, R., Carriquiry, A., de Brauw, A., Meenakshi, J. V., & Gilligan, D. O. (2012). Introduction of β-Carotene–Rich Orange Sweet Potato in Rural Uganda Resulted in Increased Vitamin A Intakes among Children and Women and Improved Vitamin A Status among Children. *The Journal of Nutrition*, *142*(10), 1871–1880. https://doi.org/10.3945/jn.111.151829

Hotz, C., & McClafferty, B. (2007). From Harvest to Health: Challenges for Developing Biofortified Staple Foods and Determining Their Impact on Micronutrient Status. *Food and Nutrition Bulletin*, *28*(2_suppl2), S271–S279. https://doi.org/10.1177/15648265070282S206

ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). (1982). *Proceedings of the International Symposium on Sorghum Grain QualityQuality, 28-31 October 1981, Patancheru, A.P., India* (pp. 28–31).

Institute of Medicine. (2001). Vitamin A. In *Dietary Reference Intakes for Vitamin A,*

Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc. (p. 10026). National Academies Press. https://doi.org/10.17226/10026

Kargougou, R., Ouaro, B., N'Diaye, B., Garnier, D., & Cisse, D. (2015). Vitamin a Supplementation Program in Burkina Faso: Coverage Performance during the Last Ten Years and Ways Forward. *European Journal of Nutrition & Food Safety*, *5*(5), 467–468. https://doi.org/10.9734/EJNFS/2015/20915

Kean, E. G., Bordenave, N., Ejeta, G., Hamaker, B. R., & Ferruzzi, M. G. (2011). Carotenoid bioaccessibility from whole grain and decorticated yellow endosperm sorghum porridge. *Journal of Cereal Science*, *54*(3), 450–459. https://doi.org/10.1016/j.jcs.2011.08.010

Kean, E. G., Ejeta, G., Hamaker, B. R., & Ferruzzi, M. G. (2007). Characterization of Carotenoid Pigments in Mature and Developing Kernels of Selected Yellow-Endosperm Sorghum Varieties. *Journal of Agricultural and Food Chemistry*, *55*(7), 2619–2626. https://doi.org/10.1021/jf062939v

Kean, E. G., Hamaker, B. R., & Ferruzzi, M. G. (2008). Carotenoid Bioaccessibility from Whole Grain and Degermed Maize Meal Products. *Journal of Agricultural and Food Chemistry*, *56*(21), 9918–9926. https://doi.org/10.1021/jf8018613

Li, F., Vallabhaneni, R., & Wurtzel, E. T. (2008). PSY3, a New Member of the Phytoene Synthase Gene Family Conserved in the Poaceae and Regulator of Abiotic Stress-Induced Root Carotenogenesis. *Plant Physiology*, *146*(3), 1333–1345. https://doi.org/10.1104/pp.107.111120

Lipkie, T. E., De Moura, F. F., Zhao, Z.-Y., Albertsen, M. C., Che, P., Glassman, K., &

Ferruzzi, M. G. (2013). Bioaccessibility of Carotenoids from Transgenic Provitamin A Biofortified Sorghum. *Journal of Agricultural and Food Chemistry*, *61*(24), 5764–5771. https://doi.org/10.1021/jf305361s

Low, J., Ball, A., Magezi, S., Njoku, J., Mwanga, R., Andrade, M., Tomlins, K., Dove, R., & van Mourik, T. (2017). Sweet potato development and delivery in sub-Saharan Africa. *African Journal of Food, Agriculture, Nutrition and Development*, *17*(02), 11955–11972. https://doi.org/10.18697/ajfand.78.HarvestPlus07

Magomere, T., Obukosia, S., Albertsen, M., Wambugu, F., Kamanga, D., Njuguna, M., Gaffney, J., Zhao, Z.-Y., Che, P., Aseta, A., Kimani, E., & Mwasame, E. (2016). Evaluation of fitness in F 2 generations of Africa Biofortified Sorghum event 188 and weedy Sorghum bicolor ssp. Drummondii. *Electronic Journal of Biotechnology*, *22*, 52–61. https://doi.org/10.1016/j.ejbt.2016.06.001

Menkir, A., Rocheford, T., Maziya-Dixon, B., & Tanumihardjo, S. (2015). Exploiting natural variation in exotic germplasm for increasing provitamin-A carotenoids in tropical maize. *Euphytica*, *205*(1), 203–217. https://doi.org/10.1007/s10681-015-1426-z

Mulatu, E., & Belete, K. (2001). PARTICIPATORY VARIETAL SELECTION IN LOWLAND SORGHUM IN EASTERN ETHIOPIA: IMPACT ON ADOPTION AND GENETIC DIVERSITY. *Experimental Agriculture*, *37*(2), 211–229. https://doi.org/10.1017/S0014479701002034

Mundia, C. W., Secchi, S., Akamani, K., & Wang, G. (2019). A Regional Comparison of Factors Affecting Global Sorghum Production: The Case of North America, Asia and Africa's Sahel. *Sustainability*, *11*(7), 2135.

https://doi.org/10.3390/su11072135

Nair, M. K., Augustine, L. F., & Konapur, A. (2016). Food-Based Interventions to Modify Diet Quality and Diversity to Address Multiple Micronutrient Deficiency. *Frontiers in Public Health*, *3*. https://doi.org/10.3389/fpubh.2015.00277

Nkongolo, K. K. (n.d.). *Participatory variety selection and characterization of Sorghum (Sorghum bicolor (L.) Moench) elite accessions from Malawian gene pool using farmer and breeder knowledge*. 11.

Ohmiya, A. (2009). Carotenoid cleavage dioxygenases and their apocarotenoid products in plants. *Plant Biotechnology*, *26*(4), 351–358. https://doi.org/10.5511/plantbiotechnology.26.351

Olson, R., Gavin-Smith, B., Ferraboschi, C., & Kraemer, K. (2021). Food Fortification: The Advantages, Disadvantages and Lessons from Sight and Life Programs. *Nutrients*, *13*(4), 1118. https://doi.org/10.3390/nu13041118

Ortiz-Monasterio, J. I., Palacios-Rojas, N., Meng, E., Pixley, K., Trethowan, R., & Peña, R. J. (2007). Enhancing the mineral and vitamin content of wheat and maize through plant breeding. *Journal of Cereal Science*, *46*(3), 293–307. https://doi.org/10.1016/j.jcs.2007.06.005

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C. R., Buckler, E. S., DellaPenna, D., Gore, M. A., & Rocheford, T. (2014). A Foundation for Provitamin A Biofortification of Maize: Genome-Wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics*, *198*(4), 1699–1716. https://doi.org/10.1534/genetics.114.169979

Palmer, A. C., West, K. P., Dalmiya, N., & Schultink, W. (2012). The use and interpretation of serum retinol distributions in evaluating the public health impact of vitamin A programmes. *Public Health Nutrition*, *15*(7), 1201–1215. https://doi.org/10.1017/S1368980012000560

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., … Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, *457*(7229), 551–556. https://doi.org/10.1038/nature07723

Patil, N. Y., Klein, R. R., Williams, C. L., Collins, S. D., Knoll, J. E., Burrell, A. M., Anderson, W. F., Rooney, W. L., & Klein, P. E. (2017). Quantitative Trait Loci Associated with Anthracnose Resistance in Sorghum. *Crop Science*, *57*(2), 877–890. https://doi.org/10.2135/cropsci2016.09.0793

Platten, J. D., Cobb, J. N., & Zantua, R. E. (2019). Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. *PLOS ONE*, *14*(1), e0210529. https://doi.org/10.1371/journal.pone.0210529

Priya, R., Sneha, P., Dass, J. F. P., Doss C, G. P., Manickavasagam, M., & Siva, R. (2019). Exploring the codon patterns between CCD and NCED genes among different plant species. *Computers in Biology and Medicine*, *114*, 103449. https://doi.org/10.1016/j.compbiomed.2019.103449

Qaim, M., Stein, A. J., & Meenakshi, J. V. (2007). Economics of biofortification: Economics of biofortification. *Agricultural Economics*, *37*, 119–133.

https://doi.org/10.1111/j.1574-0862.2007.00239.x

Reddy, B. V., Ramesh, S., & Longvah, T. (2005). *Prospects of Breeding for Micronutrients and β-Carotene-Dense Sorghums*. 5.

Rhodes, D. H., Hoffmann, L., Rooney, W. L., Herald, T. J., Bean, S., Boyles, R., Brenton, Z. W., & Kresovich, S. (2017). Genetic architecture of kernel composition in global sorghum germplasm. *BMC Genomics*, *18*(1), 15. https://doi.org/10.1186/s12864-016-3403-x

Salas Fernandez, M. G., Kapran, I., Souley, S., Abdou, M., Maiga, I. H., Acharya, C. B., Hamblin, M. T., & Kresovich, S. (2009). Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. *Genetic Resources and Crop Evolution*, *56*(7), 991–1000. https://doi.org/10.1007/s10722-009-9417-3

Saltzman, A., Birol, E., Bouis, H. E., Boy, E., De Moura, F. F., Islam, Y., & Pfeiffer, W. H. (2013). Biofortification: Progress toward a more nourishing future. *Global Food Security*, *2*(1), 9–17. https://doi.org/10.1016/j.gfs.2012.12.003

Shakoor, N., Ziegler, G., Dilkes, B. P., Brenton, Z., Boyles, R., Connolly, E. L., Kresovich, S., & Baxter, I. (2016). Integration of Experiments across Diverse Environments Identifies the Genetic Determinants of Variation in *Sorghum bicolor* Seed Element Composition. *Plant Physiology*, *170*(4), 1989–1998. https://doi.org/10.1104/pp.15.01971

Shwetha, H. J., Shilpa, S., Arathi, B. P., Raju, M., & Lakshminarayana, R. (2020). Biofortification of Carotenoids in Agricultural and Horticultural Crops: A Promising Strategy to Target Vitamin A Malnutrition. In N. Benkeblia (Ed.), *Vitamins and Minerals Biofortification of Edible Plants* (1st ed., pp. 123–161). Wiley.

https://doi.org/10.1002/9781119511144.ch7

Smith, M. C. (1930). The comparative nutritive value of yellow corn and the grain sorghums hegari and yellow milo. *Journal of Agricultural Research*, *40*(12), 1129–1145.

Steinbach, H. S., Benech-Arnold, R. L., & Sanchez, R. A. (1997). Hormonal Regulation of Dormancy in Developing Sorghum Seeds. *Plant Physiology*, *113*(1), 149–154. https://doi.org/10.1104/pp.113.1.149

Sun, W., & Hu, Y. (2013). EQTL Mapping Using RNA-seq Data. *Statistics in Biosciences*, *5*(1), 198–219. https://doi.org/10.1007/s12561-012-9068-3

Suryanarayana Rao, R., Rukmini, C., & Mohan, V. S. (1967). β-Carotene Context of Some Yellow-Endosperm Varieties of Sorghum. *The Indian Journal of Agricultural Science*, *38*(4), 368–372.

Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, *128*(5), 851–864. https://doi.org/10.1007/s00122-015-2475-3

The State of Food Security and Nutrition in the World 2021. (2021). FAO, IFAD, UNICEF, WFP and WHO. https://doi.org/10.4060/cb4474en

Venado, R. E., Owens, B. F., Ortiz, D., Lawson, T., Mateos-Hernandez, M., Ferruzzi, M. G., & Rocheford, T. R. (2017). Genetic analysis of provitamin A carotenoid β-cryptoxanthin concentration and relationship with other carotenoids in maize grain (Zea mays L.). *Molecular Breeding*, *37*(10), 127. https://doi.org/10.1007/s11032-017-0723-8

Virk, P. S., Andersson, M. S., Arcos, J., Govindaraj, M., & Pfeiffer, W. H. (2021). Transition From Targeted Breeding to Mainstreaming of Biofortification Traits in Crop Improvement Programs. *Frontiers in Plant Science*, *12*, 703990. https://doi.org/10.3389/fpls.2021.703990

vom Brocke, K., Trouche, G., Weltzien, E., Barro-Kondombo, C. P., Gozé, E., & Chantereau, J. (2010). Participatory variety development for sorghum in Burkina Faso: Farmers' selection and farmers' criteria. *Field Crops Research*, *119*(1), 183–194. https://doi.org/10.1016/j.fcr.2010.07.005

Wambugu, F., & Kamanga, D. (Eds.). (2014). *Biotechnology in Africa* (Vol. 7). Springer International Publishing. https://doi.org/10.1007/978-3-319-04001-1

Welsch, R., Wüst, F., Bär, C., Al-Babili, S., & Beyer, P. (2008). A Third Phytoene Synthase Is Devoted to Abiotic Stress-Induced Abscisic Acid Formation in Rice and Defines Functional Diversification of Phytoene Synthase Genes. *Plant Physiology*, *147*(1), 367–380. https://doi.org/10.1104/pp.108.117028

Wirth, J., Petry, N., Tanumihardjo, S., Rogers, L., McLean, E., Greig, A., Garrett, G., Klemm, R., & Rohner, F. (2017). Vitamin A Supplementation Programs and Country-Level Evidence of Vitamin A Deficiency. *Nutrients*, *9*(3), 190. https://doi.org/10.3390/nu9030190

Wong, J. C., Lambert, R. J., Wurtzel, E. T., & Rocheford, T. R. (2004). QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. *Theoretical and Applied Genetics*, *108*(2), 349–359. https://doi.org/10.1007/s00122-003-1436-4

World Health Organization. (2009). *Global prevalence of vitamin A deficiency in*

populations at risk 1995–2005. World Health Organization. https://apps.who.int/iris/handle/10665/101008

Worzella, W. W., Khalidy, R., Badawi, Y., & Daghir, S. (1965). Inheritance of Beta-Carotene in Grain Sorghum Hybrids. *Crop Science*, *5*(6), 591–592.

Xu, J., Fu, X., Cai, Y., Wang, M., Xu, S., Li, J., & Yang, X. (2019). Uncovering the genetic basis of carotenoid variations in maize kernels using two segregating populations. *Molecular Breeding*, *39*(6), 78. https://doi.org/10.1007/s11032-019-0982-7

You, H. (2016). *Quantifying the bioefficacy of β-carotene-biofortified sorghum* (p. 8943363) [Doctor of Philosophy, Iowa State University, Digital Repository]. https://doi.org/10.31274/etd-180810-4827

Zhai, S., Liu, J., Xu, D., Wen, W., Yan, J., Zhang, P., Wan, Y., Cao, S., Hao, Y., Xia, X., Ma, W., & He, Z. (2018). A Genome-Wide Association Study Reveals a Rich Genetic Architecture of Flour Color-Related Traits in Bread Wheat. *Frontiers in Plant Science*, *9*, 1136. https://doi.org/10.3389/fpls.2018.01136

Zhang, X., Pfeiffer, W. H., Palacios-Rojas, N., Babu, R., Bouis, H., & Wang, J. (2012). Probability of success of breeding strategies for improving pro-vitamin A content in maize. *Theoretical and Applied Genetics*, *125*(2), 235–246. https://doi.org/10.1007/s00122-012-1828-4

Zhao, Z. (2007). The Africa Biofortified Sorghum Project– Applying Biotechnology to Develop Nutritionally Improved Sorghum for Africa. In Z. Xu, J. Li, Y. Xue, & W. Yang (Eds.), *Biotechnology and Sustainable Agriculture 2006 and Beyond* (pp. 273–277). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6635-1_41

Zheng, X., Giuliano, G., & Al-Babili, S. (2020). Carotenoid biofortification in crop plants: Citius, altius, fortius. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1865*(11), 158664. https://doi.org/10.1016/j.bbalip.2020.158664

Zunjare, R. U., Hossain, F., Muthusamy, V., Baveja, A., Chauhan, H. S., Bhat, J. S., Thirunavukkarasu, N., Saha, S., & Gupta, H. S. (2018). Development of Biofortified Maize Hybrids through Marker-Assisted Stacking of β-Carotene Hydroxylase, Lycopene-ε-Cyclase and Opaque2 Genes. *Frontiers in Plant Science*, *9*, 178. https://doi.org/10.3389/fpls.2018.00178

CHAPTER 2 CHARACTERIZATION OF GRAIN CAROTENOIDS IN GLOBAL SORGHUM GERMPLASM TO GUIDE GENOMICS-ASSISTED BREEDING STRATEGIES[1]

INTRODUCTION

Carotenoids are the main source of vitamin A in many developing countries where diets are primarily plant based. Cereals provide the majority of calories in developing countries, and most cereal grains accumulate carotenoids —particularly lutein, zeaxanthin, β-carotene, and β-cryptoxanthin (Trono, 2019). However, the concentration of provitamin A carotenoids —β-carotene, β-cryptoxanthin and α-carotene— is low in cereals compared to fruits, vegetables, and animal derived products. For example, among cereals, maize contains the highest concentrations of carotenoids (Trono, 2019), however, the majority of yellow maize accessions accumulate less than 2µg/g of provitamin A carotenoids (Tanumihardjo, 2013), although accessions with higher concentrations have been identified (Menkir et al., 2015). This concentration is low compared to carotenoid-containing fruits and vegetables, such as carrots (50 µg/g of β-carotene) (Jeffery et al., 2012), melon (1 µg/g of β-carotene) (Bouis, 2018; Jeffery et al., 2012) and kale (6 µg/g of β-carotene) (Sikora & Bodziarczyk, 2012).

Globally, vitamin A deficiency affects an estimated 190 million preschool age children and 19 million pregnant women, contributing to poor growth, intellectual impairment, vision loss, perinatal complications, and increased mortality (Global Nutrition

[1] Submitted to BMC Plant Biology, June 19, 2022, Authors: Clara Cruet-Burgos, Geoffrey P Morris, Davina H. Rhodes

Report, 2014; WHO, 2009). Cereal biofortification is o ne of the most sustainable strategies to combat vitamin A deficiency in developing countries (Bouis et al., 2011). HarvestPlus has accomplished successful maize vitamin A biofortification through traditional breeding, with current releases in several countries containing β-carotene ranging from 4-16 μg/g (Harvest Plus, 2022). Given the prevalence of vitamin A deficiency in many developing countries and proven success of current biofortified crops, expanding biofortification efforts to other staple crops could significantly reduce global vitamin A deficiencies.

Sorghum [Sorghum bicolor (L.) Moench] is a good candidate for biofortification as it is a staple food in regions with high prevalence of vitamin A deficiency, such as in South East Asia and sub-Saharan Africa (FAOSTAT, 2021). Studies have demonstrated that carotenoids are present in sorghum grains, and β-carotene is the main provitamin A carotenoid, with concentrations up to 3.23 μg/g (Afify et al., 2012; Blessin et al., 1962, 1962; Cardoso et al., 2015; Cruet-Burgos et al., 2020; Fernandez et al., 2008; W.-N. Fu, 1960, p. 19; Kean et al., 2011; Shen et al., 2017, p. 200; Worzella et al., 1965). We estimate a sorghum biofortification target value of approximately 12 μg/g β-carotene, although that value will vary depending on the sorghum intake of the target population. Sorghum biofortification using genetic engineering has developed sorghum grain with β-carotene concentrations as high as 12 μg/g (Che et al., 2016; Zhao, 2007), but genetically modified sorghum has not been adopted by farmers due to limitations on use of transgenic crops in Africa (Tembo, 2021; Wambugu & Kamanga, 2014). However, progress in sorghum carotenoid research suggests that biofortification through breeding is feasible. Genetic studies have demonstrated that there is natural phenotypic variation in sorghum

grain carotenoids and that there are genetic components controlling the trait (Cruet-Burgos et al., 2020; Fernandez et al., 2008; Worzella et al., 1965). Genomics-assisted breeding, via marker-assisted selection (MAS) or genomic selection (GS), has the potential to accelerate biofortification efforts by removing the need to employ complex phenotyping methods. Therefore, due to the potential impact of biofortified sorghum in developing countries, the feasibility of sorghum vitamin A biofortification through breeding needs to be tested.

To use genomics-assisted breeding to develop a carotenoid biofortified sorghum variety, genomic regions associated with variation of provitamin A carotenoids must be identified, as well as efficient selection methods and parental donors. The carotenoid biosynthetic pathway is well understood and conserved in plants (Hirschberg, 2001), facilitating identification of carotenoid genes in sorghum. For β-carotene—the most abundant provitamin A carotenoid in sorghum grain—marker-trait associations have been identified in proximity to phytoene synthase (PSY), phytoene desaturase (PDS), and geranylgeranyl diphosphate synthase (GGPS) genes (Cruet-Burgos et al., 2020; Fernandez et al., 2008). For zeaxanthin, marker-trait associations have been identified within zeaxanthin epoxidase (ZEP), a gene that has also been identified in several other crops as underlying carotenoid variation (Azmach et al., 2018; Ikoma et al., 2016; Jourdan et al., 2015; Zhou et al., 2011). Previous research in sorghum suggests that carotenoids are oligogenic traits, meaning a moderate number of genes with moderate effect underly the majority of the phenotypic variation detected (Cruet-Burgos et al., 2020; Fernandez et al., 2008; Worzella et al., 1965), which is consistent with observations in maize

(Diepenbrock et al., 2021) and wheat (Kumar et al., 2018). The oligogenic variation of carotenoids in cereal grains suggests that MAS may be an effective biofortification method. Alternatively, due to the interconnectedness of carotenoid biosynthesis to other biochemical pathways, a combination of both oligogenic and polygenic models might more accurately explain carotenoid variation. In this instance, GS, or MAS followed by GS, could be employed to simultaneously select for both large and small effect genes (Figure 2.1).

Identifying germplasm that harbor alleles for high β-carotene concentrations in sorghum grains is also essential for biofortification breeding. This aspect is perhaps the most challenging, because even though carotenoids are naturally present in sorghum grains, studies have shown that the majority of sorghum varieties have low β-carotene concentrations (Abdel-Aal et al., 2013; Cardoso et al., 2015; Cruet-Burgos et al., 2020; Fernandez et al., 2008; Kean et al., 2011). The limited number of high β-carotene varieties could imply that there is limited genetic diversity for carotenoid concentrations, which would impede gains in breeding efforts. However, the accessions that have been phenotyped for β-carotene concentration is limited, suggesting there could be unexplored germplasm harboring genetic variation. No direct assessment of diversity in high carotenoid sorghum accessions has been conducted, but the high incidence of yellow endosperm from Nigeria in collections and previous studies (Blessin et al., 1958; Fernandez et al., 2008; W. Sun & Hu, 2013; Suryanarayana Rao et al., 1967) supports that limited diversity is a possibility. Expanding germplasm evaluations and genetic studies could therefore highlight new or conserved genomic regions associated with β-

carotene that can be used for MAS, as well as identify a set of parental lines with sufficient genetic diversity to employ in breeding efforts.

Given the potential impact of a high vitamin A sorghum developed through biofortification breeding, and the current gaps in knowledge for sorghum biofortification, in this study we further explored the potential of genomics-enabled breeding tools for increasing carotenoids in sorghum. We hypothesize that both oligogenic and polygenic components of variation exist for sorghum carotenoids (Figure 2.1), such that both MAS and GS could accelerate breeding efforts. We expanded the number of sorghum accessions phenotyped for carotenoids, identifying additional high carotenoid accessions, and found evidence for both an oligogenic and polygenic component of variation in sorghum grain carotenoids. We also found that the limited number of known high carotenoid accessions have low genetic diversity among them, but genomic predictions identified new potential donor lines that could harbor novel genetic variation for carotenoids. Lastly, we examined allelic diversity in the *ZEP* gene and found evidence of selection for a high carotenoid allele found in only a few countries.

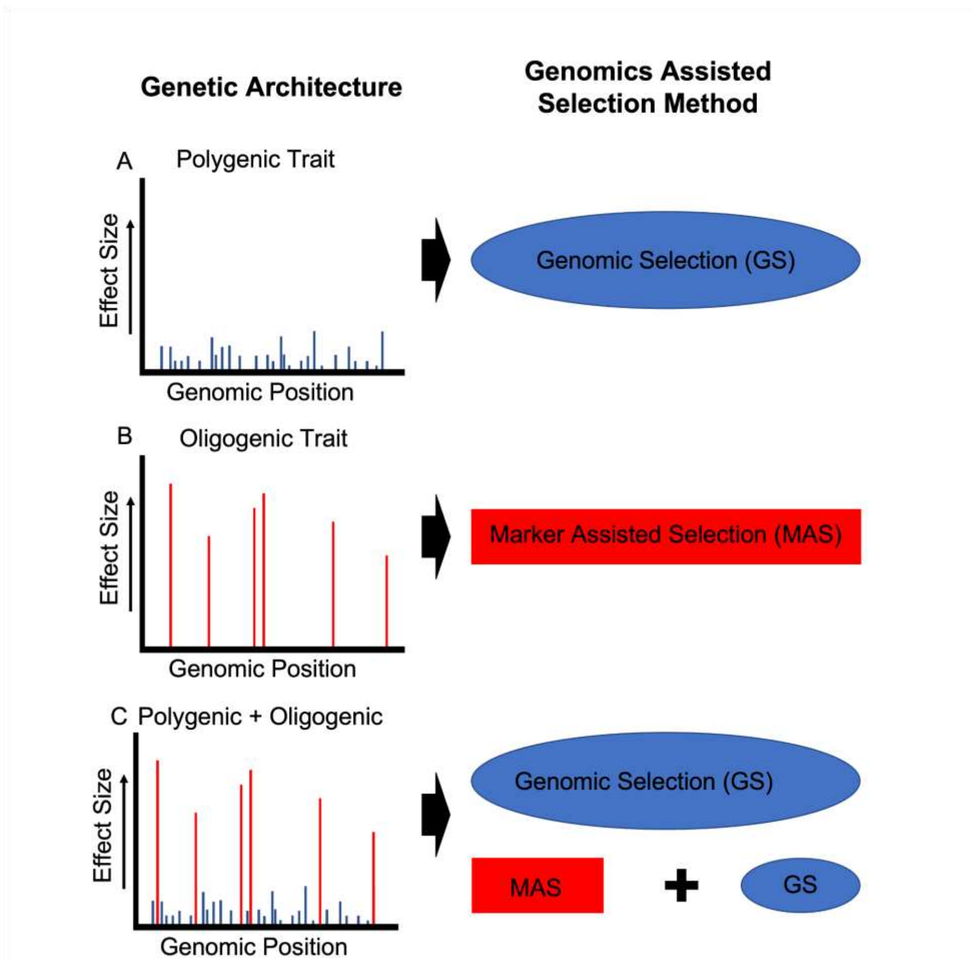


Figure 2.1 Genomics-assisted selection method based on genetic architecture of a trait. Genomics-assisted selection method for a trait under A) polygenic; B) oligogenic; and C) both polygenic and oligogenic control. Effect size is represented by height of vertical lines; number of genes associated with the trait is represented by number of vertical lines; blue vertical lines represent the polygenic component and red vertical lines represent the oligogenic component contributing to the trait.

METHODS

Plant material

Grain carotenoid concentration was evaluated for a total of 447 sorghum accessions, which included 316 from the sorghum association panel (SAP) (Casa et al., 2008) and 130 accessions from the carotenoid panel (CAP), a set of accessions chosen for presence of yellow endosperm and/or yellow grain (Cruet-Burgos et al., 2020; Salas Fernandez et al., 2009). The two panels were grown, selfed, and harvested by the authors at Kansas State University Agronomy North Farm in Manhattan, Kansas with a randomized complete block design (RCBD) with two replications during the summer of 2019. At maturity, grain was harvested, dried and stored at -80 °C until carotenoid quantification.

Carotenoid quantification

Carotenoid extractions were performed following a modified solid phase method (Irakli et al., 2016). All steps of the extraction were carried out under yellow light to avoid photodegradation. Briefly, approximately 5 seeds were ground to flour using a Bead Ruptor Elite (Omni International, Kennesaw, GA) and 20 mg of the sorghum flour were transferred to a 1.5mL eppendorf tube. Next, 20 mg of ascorbic acid and 400 μl of absolute ethanol with a 1 mg/mL concentration of butylated hydroxytoluene (BHT) were added. The tubes were vortexed for 1 minute and placed in an 80 °C water bath for 5 minutes. Following the incubation, 20 μl of a solution of potassium hydroxide (80% w/v, in water) was added and tubes vortexed for 1 minute. Next, samples were returned to the water bath for 15 minutes and mixed every 5 minutes. Samples were then brought to room temperature and centrifuged for 5 minutes at 1800 rcf. Supernatant was transferred

to a new 1.5 mL eppendorf tube. An additional 400 µl of absolute ethanol with 1 mg/mL of BHT was added to the residue, vortexed for 1 minute and centrifuged for 5 minutes at 1800 rcf. The supernatant was combined with the above extract, vortexed for 30 seconds and centrifuged for 5 minutes at 5000 rcf. The supernatant was then transferred to a new 1.5mL eppendorf tube and evaporated to dryness with a gentle N_2 stream at room temperature. Finally, the residue was reconstituted in 100 µl of Methanol:Ethyl Acetate (1:1) and centrifuge for 5 minutes at 5000 rcf. An aliquot of 40 µl of the clear supernatant was utilized for the HPLC analysis. Resolution of lutein, zeaxanthin, β-carotene, α-carotene and β-cryptoxanthin was conducted using an Agilent 1290 Infinity UHPLC with Eclipse Plus C18 column (Agilent Technologies, Santa Clara, California, U.S).

Statistical analysis of carotenoids and heritability estimation

Concentrations for lutein, zeaxanthin and β-carotene were analyzed with asREML R Package (Butler et al., 2018), which accounts for missing data. For the three carotenoids, we implemented a randomized complete block design model with genotype and block as random effects. The gamma parameterization with a maximum iteration number of 100 was used for the analysis. Best linear unbiased predictors (BLUPs) were obtained as predictors of genetic merit for lutein, zeaxanthin and β-carotene and were used for subsequent analysis. Broad sense heritability on an entry-mean was also calculated for lutein, zeaxanthin and β-carotene as followed:

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

where H2 represents the broad sense heritability on a entry-mean basis, σ_Genotype^2 represents the genotypic variance, σ_Phenotype^2 the phenotypic variance and

σ_{error}^2 the residual or error variance, and r represents the blocks. Pearson correlations were calculated between carotenoid pairs for lutein, zeaxanthin, and β-carotene.

Genome-wide association study

The genetic architecture and the genomic regions underlying carotenoid variation in sorghum grain were investigated through a genome wide association study (GWAS) implemented in GAPIT (Lipka et al., 2012), version "2022.4.16, GAPIT 3.1" . The genotype information for the SAP and the carotenoid panel was obtained from previous studies (Cruet-Burgos et al., 2020; Hu et al., 2019). The SNP datasets are available for download from the Dryad Data Repository (doi:10.5061/dryad.63h8fd4). After filtering the single nucleotide polymorphism (SNP) data set (Sorghum bicolor v3.1 genome version) by a minimum allele frequency of 0.05, 348,181 biallelic SNPs remained. A total of 345 accessions had genotype information and BLUP estimates for lutein, β-carotene and zeaxanthin. A mixed linear model (MLM) (model="MLM") was used with a marker derived kinship and ten principal components (PCA.total=10) to control for relationship and population structure, respectively. To account for multiple comparisons, the Bonferroni correction with P = 0.05 was used to identify significant SNPs. Significant associations were compared with candidate genes that are annotated as enzymes involved in the carotenoid pathway in Phytozome or that have been identified in other carotenoid association studies (Supplemental Table 1).

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

Genetic diversity among the high carotenoid lines identified was examined. We prioritized assessing the genetic diversity among the accessions with the highest β-

carotene BLUP estimates, because β-carotene is the most abundant provitamin A carotenoid in sorghum. Accessions were ranked as top 5% and bottom 5% based on their BLUP estimate for β-carotene. A marker-derived additive relationship matrix or kinship, was estimated with the 'A.mat' function in the rrBLUP R package (Endelman, 2011). The eigenvalues for the first two principal components were estimated with R function 'eigen' for the additive relationship matrix and the grouping of the top 5% was examined. Genetic diversity of regions associated with β-carotene variation was determined using a window of 1 megabase (Mb) upstream and downstream of significant SNPs identified by GWAS in proximity to *a priori* candidate genes. Nucleotide diversity (π) per base pair was calculated with rTASSEL (Monier et al., 2020) using a step size of 100 and a window size of 500.

Genotype from unexplored germplasm collections

Publicly available genotype data from unexplored germplasm collections were gathered. Accessions and their corresponding genotype information from Ethiopia (Cuevas et al., 2017), Haiti (Muleta et al., 2022), Niger (Maina et al., 2018), Nigeria (Hu et al., 2019; Lasky et al., 2015; Olatoye et al., 2018), Senegal (Faye et al., 2019) and Sudan (Cuevas & Prom, 2020) were obtained from published data or by contacting the authors. Common SNPs between the unexplored germplasm collections and the SAP/CAP global collection that had at least 80% of the data present were identified. Missing SNP data were then imputed using Beagle (Browning et al., 2018) with the default parameters. To assess the genetic relationships among the accessions in the unexplored germplasm and the SAP/CAP collection a realized additive relationship matrix was calculated first using the 'A.mat' function in rrBLUP R package (Endelman, 2011). The

additive relationship matrix was then used to perform a principal component analysis (PCA) in R.

Genomic prediction of GEBV for carotenoids in unexplored germplasm

Predictions of the genomic estimated breeding values (GEBV) and genomic heritability were conducted using the genomic data from unexplored germplasm, representing country collections of Ethiopia, Haiti, Niger, Nigeria, Senegal, Sudan, and the SAP/CAP collection. The accessions in the SAP/CAP collection for which we had BLUPs estimates for lutein, zeaxanthin and β-carotene were used as the training population. GEBV were estimated for lutein, zeaxanthin and β-carotene using the G-BLUP model with the additive relationship matrix or kinship as implemented in the rrBLUP package in R (Endelman, 2011). A 5 k fold cross validation approach was used for each carotenoid to determine prediction accuracy. The cross validation was repeated for 100 cycles. Genomic heritability for lutein, zeaxanthin and β-carotene was estimated during each k-fold and cycle. Prediction accuracy was also estimated by calculating the correlation between the genomic prediction and the validation values divided by square root of heritability (Dekkers, 2007). The unexplored germplasm was ranked in the top 5% for lutein, zeaxanthin and β-carotene based on the GEBV estimates. A marker-derived kinship was estimated with the 'A.mat' function in the rrBLUP R package (Endelman, 2011) for the unexplored germplasm. The eigenvalues for the first two principal components were estimated with R function 'eigen' for the additive relationship matrix and the grouping of the top 5% for each of the carotenoids was examined. Additionally, the distribution of the GEBVs for β-carotene was evaluated by country. Lastly, we compared the GEBV for the three carotenoids in the SAP/CAP collection and unexplored germplasm collections.

Allelic diversity and geographic distribution of ZEP allele

Distribution of the *ZEP* allele for the SAP/CAP collection and the unexplored germplasm collections was examined using country of origins and the allelic classes for the SNP S06_46717975. Countries that had less than 3 accessions were discarded from the analysis. We also aggregated the allelic variants present in the high β-carotene accessions from the SAP/CAP collection and the unexplored germplasm collections, as defined by the top 5% of BLUP or GEBV for β-carotene.

RESULTS

Phenotypic variation of carotenoids in the SAP/CAP collection

To characterize phenotypic variation of carotenoids in sorghum grain, and to confirm previously published phenotype data on one year of samples, we quantified lutein, zeaxanthin, β-carotene, and α-carotene for 447 accessions in the SAP/CAP global collection using HPLC. Lutein was the most abundant carotenoid, followed by zeaxanthin, β-carotene, and then α-carotene. Raw concentrations for lutein ranged from 0.02-4.61 µg/g, for zeaxanthin from 0.01-2.40 µg/g and for β-carotene from 0.03-1.19 µg/g, with means of 0.58 µg/g, 0.18 µg/g, and 0.17 µg/g, respectively. α-carotene was detected in only 31 accessions, with values ranging from 0.02-0.11 µg/g. Due to the limited number of accessions with detectable concentrations, α-carotene was omitted from subsequent genetic analysis. High correlations were found between β-carotene and zeaxanthin (r = 0.74; $p < 10^{-16}$), β-carotene and lutein (r = 0.78; $p < 10^{-16}$), and lutein and zeaxanthin (r = 0.75; $p < 10^{-16}$). Four accessions, two of which had not previously been phenotyped, had

higher concentrations of β-carotene than any accessions previously phenotyped in the SAP/CAP collection.

To account for unbalanced data and accurately predict the genetic merit for carotenoids of the SAP/CAP accessions, BLUPs and heritabilities were calculated for each of the carotenoid traits (Table 2.1, Supplemental Table 2). Due to the expected shrinkage effect, lower ranges were obtained for the BLUPs than for the raw concentrations. However, entry-mean basis heritability estimates (Table 2.1) were high, ranging from 0.78 for β-carotene to 0.92 for zeaxanthin.

Table 2.1. Range, mean, and entry-mean basis heritability (H^2) for the BLUPs of lutein, zeaxanthin, and β-carotene for the SAP/CAP collection.

	Lutein (μg/g)	Zeaxanthin (μg/g)	β-carotene (μg/g)
Minimum	0.15	0.04	0.07
Maximum	3.09	1.83	0.80
Mean	0.58	0.18	0.17
H^2	0.80	0.92	0.78

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

Next, we sought to characterize the genetic architecture of sorghum carotenoids. A previous study (Cruet-Burgos et al., 2020) suggested that global sorghum grain carotenoid variation is oligogenic, so to further test this hypothesis, we conducted a GWAS using more accessions and replicates. To maximize the number of accessions included, we used BLUPs rather than raw data in order to account for unbalanced data.

Marker-trait associations were identified for the BLUPs of the three carotenoid traits evaluated (Figure 2.2, Supplemental Table 3).

For lutein, only 1 SNP, on chromosome 4 (S04_275231), was above the Bonferroni threshold of significance (Figure 2.2A, Supplemental Table 3). To identify candidates that may not be found using the stricter Bonferroni multiple comparison corrections, we also considered a more liberal False Discovery Rate (FDR) criteria. Under the FDR < 0.05 threshold, 7 significant SNPs were identified, corresponding to four regions of association on chromosomes 3, 4, 6 and 9. Three of these SNPs were located in a region around 2.17 Mb on chromosome 9, which is not near any a priori candidate genes. The only association in proximity to an a priori candidate gene was at SNP S6_47123508, near Sobic.006G097500 (401 kb away), an a priori candidate gene that is annotated as a putative ortholog of the maize ZEP gene.

Zeaxanthin had the highest number of marker-trait associations above the Bonferroni significance threshold, with 39 significant SNPs in 17 regions across all chromosomes except chromosome 3 (Figure 2.2B, Supplemental Table 2). The most prominent association was on chromosome 6 between 45.9-48.6 Mb, with six significant SNPs, three of which were among the top ten most significant associations. The most significant association for zeaxanthin was the SNP near the *ZEP* gene (S6_47123508; 401 kb away) that was also associated with lutein. There was also an association on chromosome 2 (S2_61694864), which is in proximity to Sobic.002G225400 (42 kb away), an *a priori* candidate gene annotated as an *abscisic acid 8'-hydroxylase 3* (*CYP707A*).

β-carotene had ten significant marker-trait associations for a total of six regions of association across chromosomes 2, 6 and 10 (Figure 2.2C, Supplemental Table 3). Chromosome 10 had the highest number of marker-trait associations, particularly within

a region around 7.48 Mb. There was also a SNP on chromosome 10 (S10_14377366) that was significantly associated with both β-carotene and zeaxanthin, which is not in proximity to any *a priori* candidate genes. Among the ten markers associated with β-carotene, only SNP S06_47123508, 401 kb from the *ZEP* gene, was in proximity to an *a priori* candidate gene.

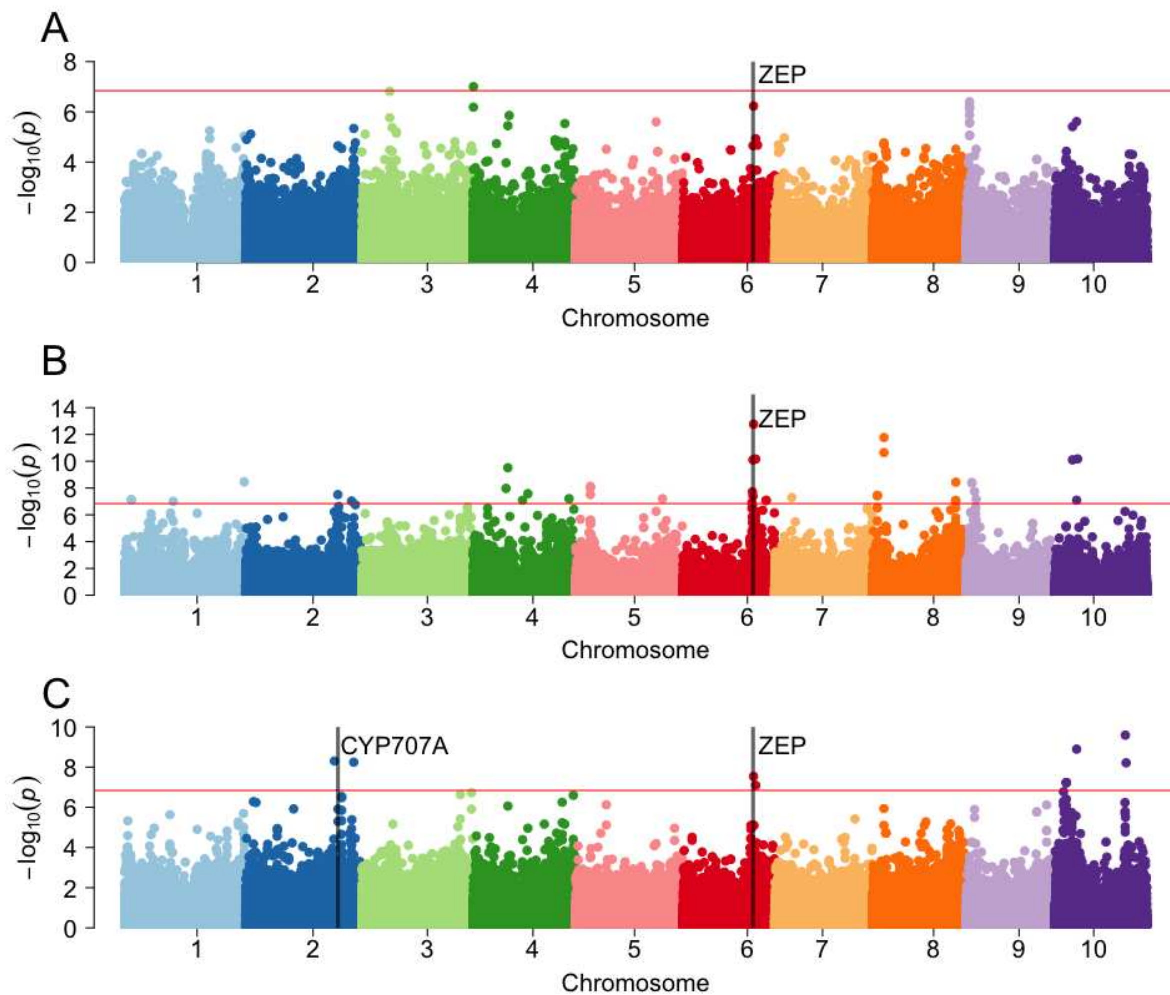


Figure 2.2 Genome-wide association study of carotenoid BLUP estimates using MLM. Manhattan plot of BLUPs for A) lutein, B) zeaxanthin, and C) β-carotene. The red horizontal line represents the genome wide significance threshold for the Bonferroni multiple comparisons correction at P = 0.05.

Population structure of sorghum carotenoids in SAP/CAP global collection

Next, we tested if carotenoid variation is structured by population and geographic origin. Since provitamin A carotenoids are our primary target, we focused on β-carotene concentrations for this analysis. Country of origin was obtained from the USDA NPGS GRIN database for the accessions in the SAP/CAP collection. Countries that had less than eight accessions were discarded from the analysis. In the SAP/CAP collection there were nine countries represented by more than eight accessions: Botswana, Ethiopia, India, Lebanon, Nigeria, South Africa, Sudan, Uganda, and the United States. β-carotene BLUP estimates among this subset of 309 accessions had the same range and average as the full set of SAP/CAP global collection (Table 2.1, Figure 2.3). Interestingly, most of the countries had average values below the global average for β-carotene of 0.17 μg/g (Figure 2.3). Furthermore, β-carotene distribution for the accessions of Sudan, South Africa, India, and Botswana were almost completely below the global average. In contrast, Lebanon had the highest β-carotene BLUPs estimates with the majority of their accessions above the global average. Notably, Nigeria, had the widest range of variation as well as the highest β-carotene concentrations among the countries.

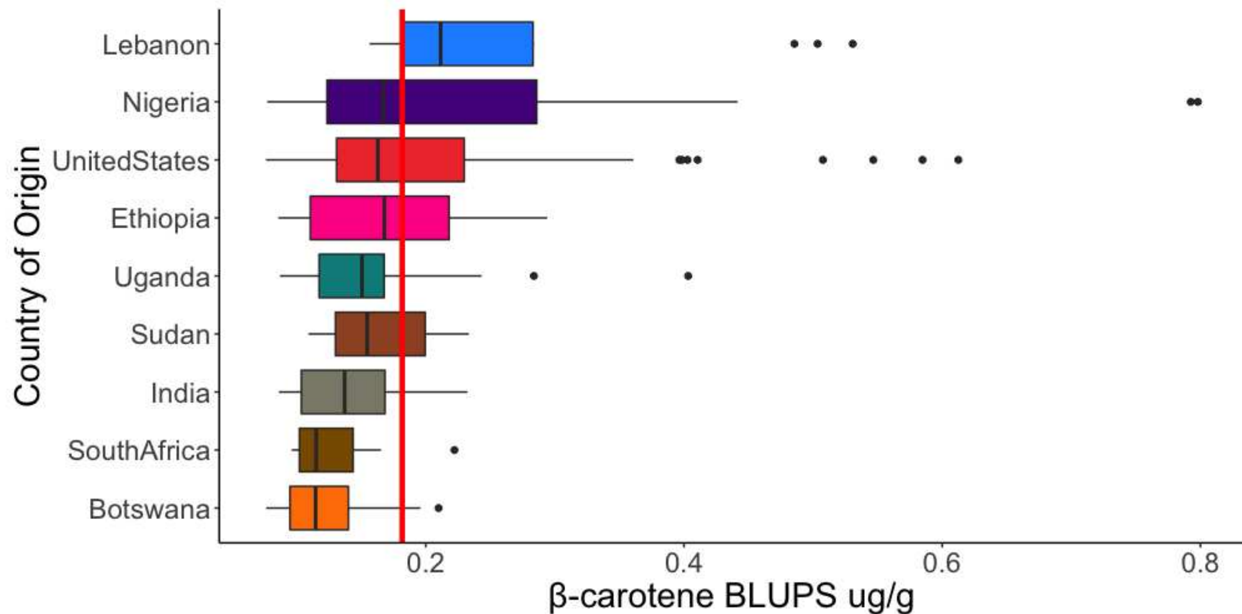


Figure 2.3 Distribution of β-carotene BLUPs among countries in the SAP/CAP collection.
The red vertical line represents the SAP/CAP average across all accessions. Countries with less than 9 accessions were excluded.

Based on the limited geographic distribution of high carotenoid sorghums, we hypothesized that the high carotenoid germplasm originates from a narrow genetic pool. To test this hypothesis, we conducted a PCA to evaluate the genetic diversity of the high carotenoid accessions identified in the SAP/CAP collection. The high carotenoid accessions were defined as those within the top 5% for β-carotene BLUP estimates, which consisted of 17 accessions ranging from 0.40 to 0.80 µg/g β-carotene. The majority of the high carotenoid accessions originated in the United States (6 accessions), followed by Nigeria (3 accessions) and Lebanon (3 accessions) (Figure 2.4A). Interestingly, the three high carotenoid accessions from Nigeria grouped together and were clustered separately from most of the other high carotenoid accessions, and another high accession of unknown origin did not group with any other high carotenoid accessions, suggesting three genetically distinct high carotenoid groups (Figure 2.4A).

To further test our hypothesis on a narrow genetic pool for high carotenoid lines, we evaluated the genetic diversity surrounding the most prominent SNP identified by GWAS for all three carotenoids (S06_47123508). We analyzed a window of 1 Mb upstream and downstream of S06_47123508, which encompassed 1,665 SNPs. Nucleotide diversity was decreased in the high carotenoid accessions, but not in the low carotenoid accessions (defined as the lowest 5% for β-carotene BLUP estimates) or in the complete set of SAP/CAP collection accessions. The most prominent region of low nucleotide diversity was surrounding SNP S06_47123508, a region which includes the *a priori* candidate gene encoding *ZEP* (Figure 2.4B).

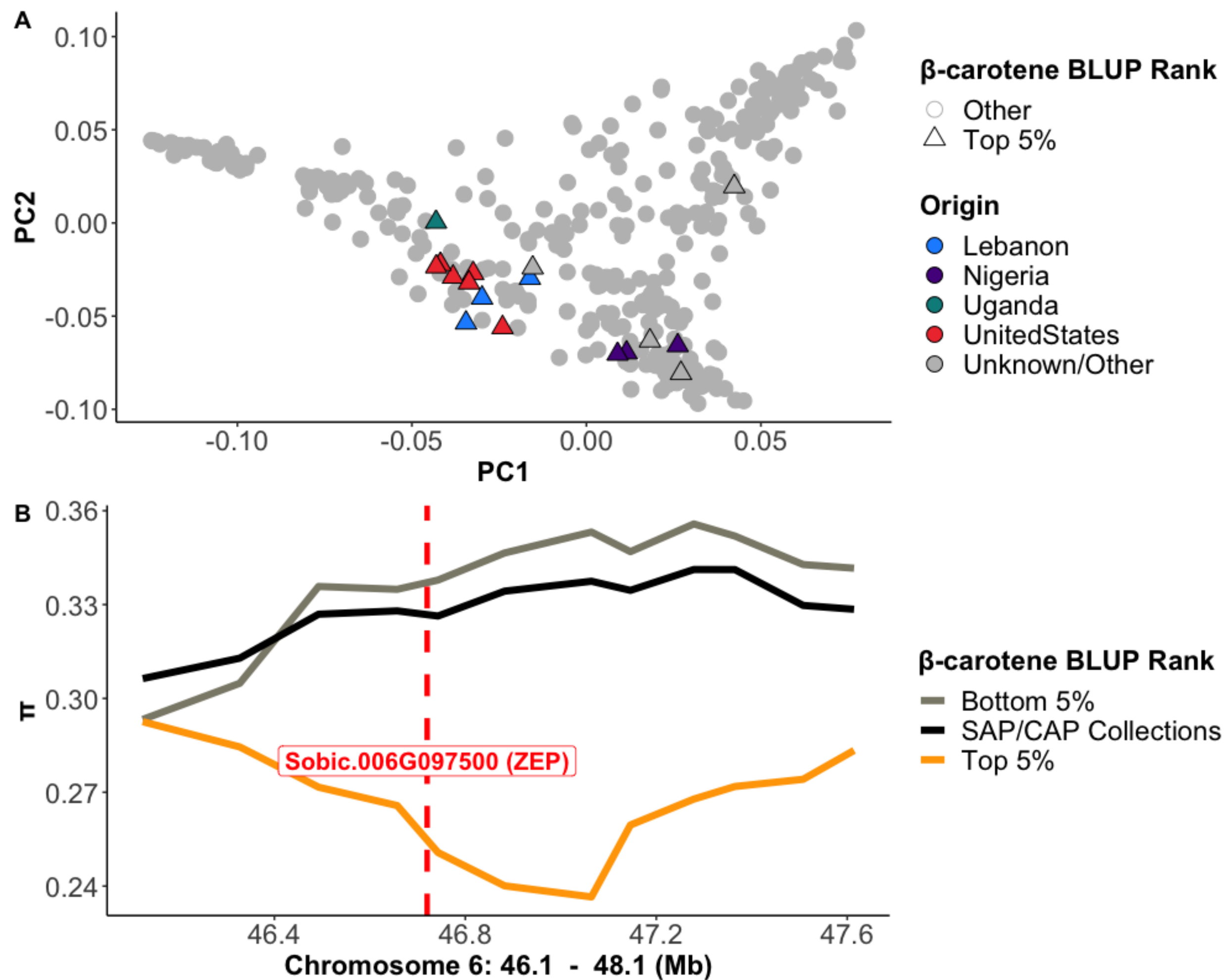


Figure 2.4 Genetic diversity of SAP/CAP accessions based on the top 5% and bottom 5% rankings for β-carotene BLUP estimates.
A) Accessions plotted according to the first two principal components for sorghum kinships coded by country of origin and ranking for β-carotene BLUP estimates. B) Nucleotide diversity of the region 1 Mb upstream and downstream of marker S06_47123508. The gray and orange lines represent nucleotide diversity for the bottom and top 5% rankings for β-carotene BLUP estimates, respectively. The black line represents the nucleotide diversity for all of the accessions in the SAP/CAP collections. The red dashed line represents the start position for the ZEP gene.

Prediction of carotenoid breeding values in unexplored germplasm collection

Next, we sought to explore if there exists unidentified high carotenoid germplasm in additional germplasm collections. Publicly available genotype data was obtained for germplasm collections from six countries: Ethiopia, Haiti, Niger, Nigeria, Senegal, and Sudan. Together with the SAP/CAP collection, the dataset consisted of 60,129 common

SNPs with less than 20% of data missing for 2,495 accessions. There were 367 accessions from Ethiopia, 296 from Haiti, 516 from Niger, 180 from Nigeria, 421 from Senegal, 319 from Sudan, and 396 from the SAP/CAP collection. Most of this germplasm is photoperiod sensitive making it difficult to phenotype in temperate regions such as the United States.

Genomic prediction has the potential to guide resource allocations by identifying the most promising germplasm to test in future work. We first explored the feasibility of the SAP/CAP collection as a training population for the unexplored germplasm collections. For this, the genetic relationship among the unexplored germplasm collections and the SAP/CAP collection was tested with a PCA, highlighting the country of origin for each accession (Figure 2.5A). Germplasm from Haiti, Ethiopia, Niger, Nigeria, and Senegal formed independent clusters, indicating genetic similarities within but not between countries. Haiti segregated the most distantly, followed by more sparsely grouped germplasm from Senegal and Nigeria. The distant genetic relationship of Hatian germplasm with the other countries was expected as these materials are from a breeding program that went through a recent bottleneck after a sugarcane aphid infestation (Muleta et al., 2022). Germplasm from Niger and Ethiopia clustered very close together, but separate from the other countries. As expected based on previous studies (Cuevas & Prom, 2020; Morris et al., 2013), accessions from the Sudan collection and SAP/CAP collection were scattered across all clusters, rather than clustering together, indicating high genetic diversity. The scattered distribution of the SAP/CAP collection confirms that it is an appropriate training population for genomic predictions in the unexplored germplasm.

Next, we estimated GEBV from the BLUPs of β-carotene, lutein, and zeaxanthin in the unexplored germplasm collections and the SAP/CAP collection (Supplemental Table 4). Lutein GEBV ranged from -0.37 to 2.16 µg/g, with an average prediction accuracy of 0.62 and a genomic heritability of 0.96 (Table 2.2). For zeaxanthin, GEBV ranged from -0.20 to 1.44 µg/g, with a prediction accuracy of 0.69 and a genomic heritability of 1.00 (Table 2.2). Lastly, for β-carotene, GEBV values ranged from -0.08 to 0.46 µg/g, with a prediction accuracy of 0.67 and a genomic heritability of 0.75 (Table 2.2). Interestingly, there were no accessions in the unexplored germplasm that had predicted GEBV for β-carotene higher than the highest values in the SAP/CAP collection, however there were some accessions that had values among the highest in all the collections (Supplemental Figure 4 and Supplemental Table 4). Finally, as seen in the SAP/CAP accessions, high correlations were identified for GEBV between β-carotene and zeaxanthin ($r = 0.89$; $p < 10^{-16}$), β-carotene and lutein ($r = 0.87$; $p < 10^{-16}$), and lutein and zeaxanthin ($r = 0.85$; $p < 10^{-16}$)

Table 2.2 Range of GEBV, average prediction accuracy and genomic heritability (H^2) for the lutein, zeaxanthin, and β-carotene for the unexplored germplasm collections and SAP/CAP collection.

	GEBV Lutein (µg/g)	GEBV Zeaxanthin (µg/g)	GEBV β-carotene (µg/g)
Minimum	-0.37	-0.20	-0.08
Maximum	2.16	1.44	0.46
Average Prediction Accuracy	0.62	0.69	0.67
H^2	0.96	1.00	0.75

To further explore geographic patterns of sorghum carotenoid distribution beyond the SAP/CAP collection, we aggregated GEBV by country using the unexplored germplasm collections (Figure 2.5B-D). Nigeria had the highest GEBV and range of values for all three carotenoids, followed by Niger. In contrast, Haiti had some of the smallest carotenoid GEBV values, as well as the smallest range of values. Interestingly, Ethiopia had several accessions with high GEBV for lutein, but only three high accessions for β-carotene, and no high accessions for zeaxanthin. Similarly, Senegal had one accession with a high GEBV for β-carotene and zeaxanthin, but not for lutein. These differences suggest that although the three carotenoids are highly correlated—consistent with common genetic controls—there are independent genetic controls, as well.

Next, we looked at the genetic relationships among the predicted top 5% for each carotenoid using a PCA for the GEBV (Figure 2.5E-G, Supplemental Table 3). The pattern of distribution differed by carotenoids, but the majority of the accessions were clustered by country. Lutein (Figure 2.5E) had two major clusters corresponding to Ethiopian accessions and a combination of accessions mostly from Nigeria and Niger. For zeaxanthin (Figure 2.5F) and β-carotene (Figure 2.5G), the clustering was similar, with the accessions of Nigeria and Niger forming the tightest cluster. Accessions from Sudan and the SAP/CAP germplasm were scattered for the three carotenoids, suggesting they are genetically distinct. Taken together, a proportion of accessions predominantly from Nigeria and Niger formed the most distinct cluster in the PCA for the three carotenoids, indicating they are genetically similar. The accessions with the highest GEBV for β-carotene were also part of this cluster.

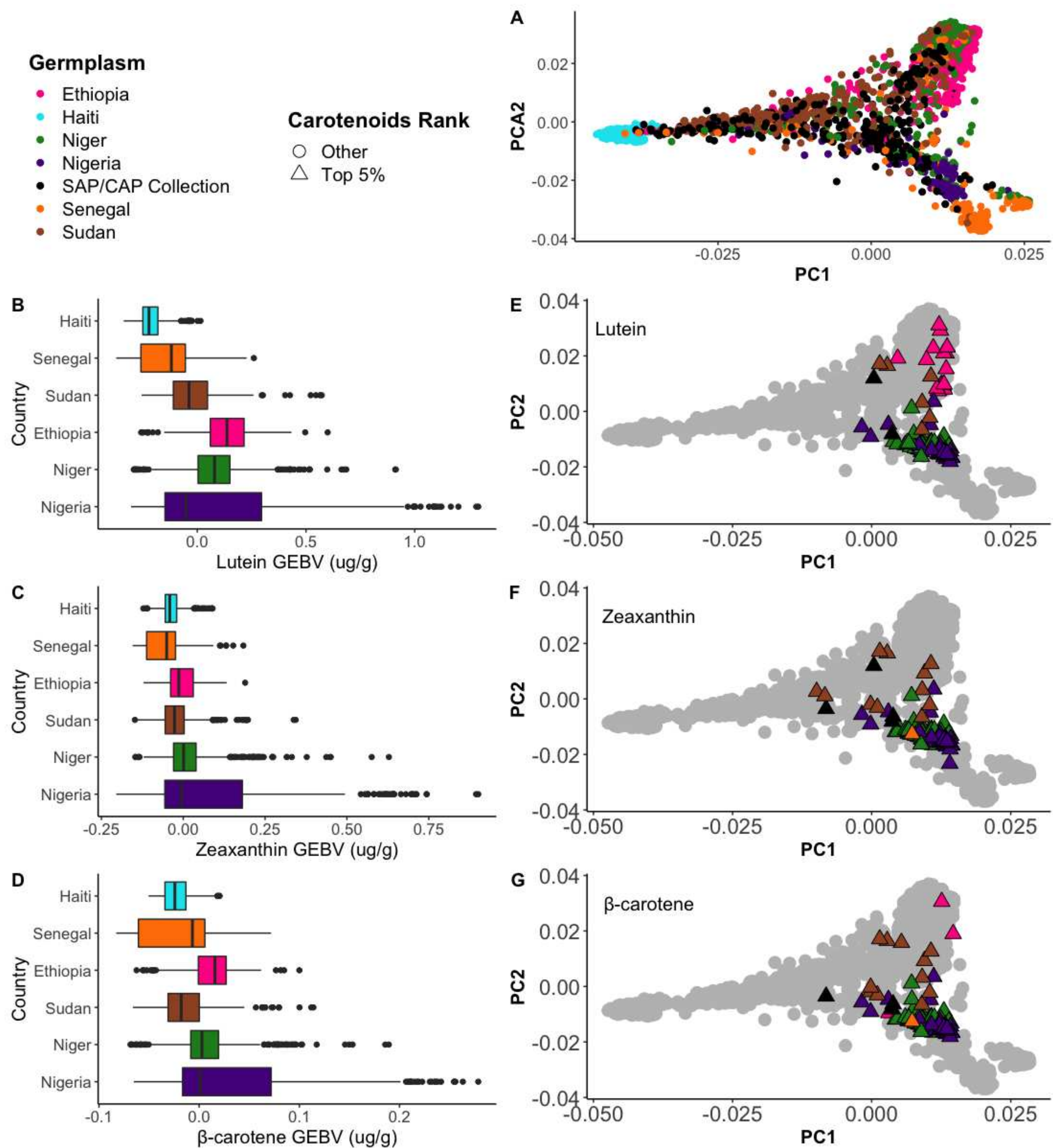


Figure 2.5 Carotenoids in unexplored germplasm collections.
A) PCA of genome-wide SNP variation for unexplored germplasm collections and SAP/CAP collection; boxplot of distribution of GEBVs aggregated by country and ordered by lowest to highest carotenoid for B) lutein, C) zeaxanthin, and D) β-carotene; PCA of the genetic relationships of the top 5% of E) lutein, F) zeaxanthin, and G) β-carotene GEBVs in the unexplored germplasm collections.

Allelic diversity and geographic distribution of ZEP allele

To further test the hypothesis that high carotenoid lines originate from a narrow genetic pool, we analyzed the SNPs inside the ZEP gene in the SAP/CAP collection and unexplored germplasm collections. In the SAP/CAP collection, we identified 14 SNPs in the ZEP gene with MAF > 0.05. Due to low marker density, the majority of these SNPs were absent in the unexplored germplasm collections. However, SNP S06_46717975 was present in the SAP/CAP collection and the SNP data set for Haiti, Niger, and Nigeria germplasm. This SNP is found within the ZEP gene and was previously identified by our group as associated with zeaxanthin variation (Cruet-Burgos et al., 2020). S06_46717975 was found to be bi-allelic with A/G variants present among the germplasms. The minor allele ‘A’ is moderately common globally, with 10% presence in SAP/CAP collection. However, among countries there are striking differences in the allele frequency; for instance, 24% in Nigeria versus 2% in Niger and 0% in Haiti germplasm.

We next explored if there were any patterns between allelic variant, geographic distribution, and carotenoid content (Figure 2.6). In the SAP/CAP collection, there was a correlation between allelic type and country of origin with the United States, Lebanon, and Nigeria, the countries with the highest prevalence of the ‘A’ allele (Figure 2.6A). Among the high carotenoid accessions in the SAP/CAP collection (defined as the top 5% for β-carotene concentration), the ‘A’ allele was present in 85% of them (Figure 2.6B). We then analyzed the alleles in the unexplored germplasm collections and the SAP/CAP accessions that were not phenotyped. Similar patterns were observed for the geographic distribution of the ‘A’ allele with the highest prevalence in the United States and Nigeria (Figure 2.6C). Surprisingly, the difference in the distribution of the ‘A’ and ‘G’ alleles was

not nearly as pronounced in the predicted high carotenoid lines based on β-carotene GEBV (Figure 2.6D).

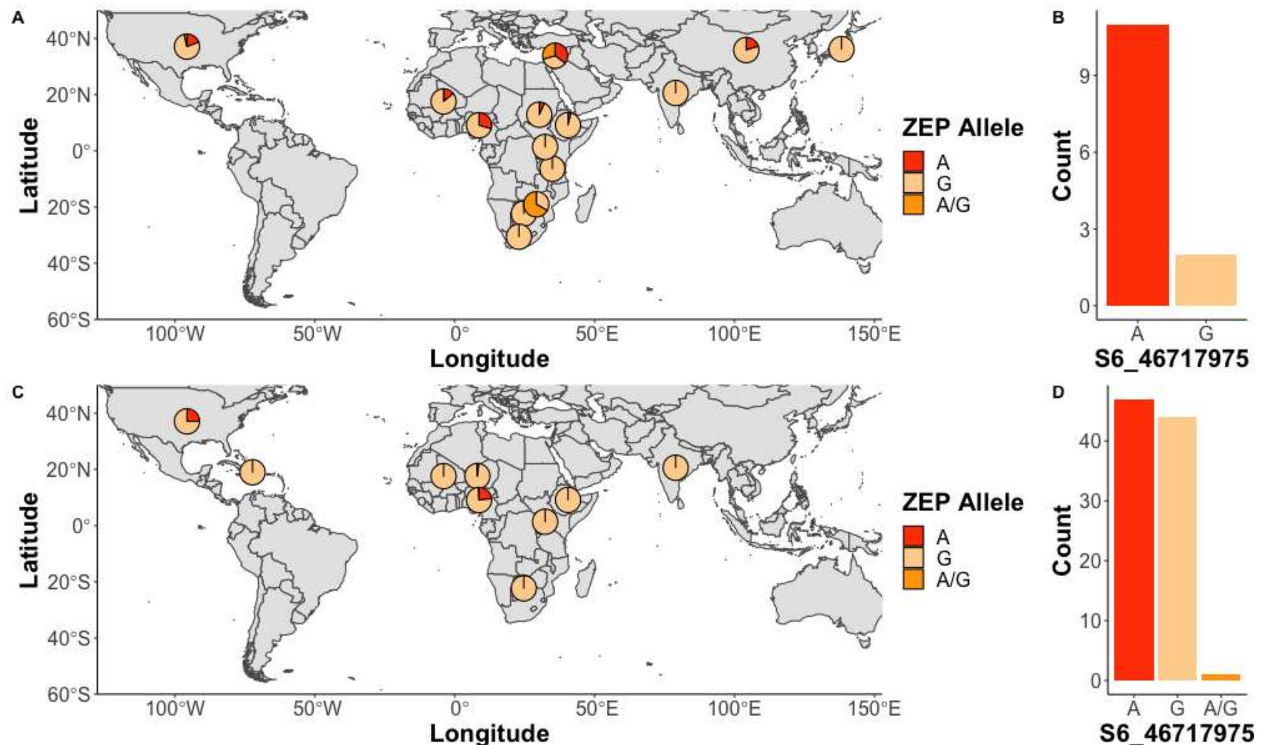


Figure 2.6 Geographic distribution of the SNP S06_46717975 inside the ZEP gene and the distribution of allelic classes in high carotenoid germplasm.
A) Geographic distribution for accessions in the SAP/CAP collection. B) Distribution of allelic classes for the top 5% rankings for β-carotene BLUP estimates. C) Geographic distribution for accessions in unexplored germplasm collections and the SAP/CAP accessions without phenotype; D) Distribution of allelic classes for the top 5% rankings for β-carotene GEBV where A is the minor allele.

DISCUSSION

Genetic diversity among high carotenoid lines

A vitamin biofortified sorghum variety has the potential to positively impact the livelihood of millions that rely on it as a dietary staple. We estimate a β-carotene target value of 12 μg/g for biofortified sorghum grain. Although the highest β-carotene content measured in our study was 1.19 μg/g, a previous study reported a sorghum variety with β-carotene concentrations as high as 3.23 μg/g (Worzella et al., 1965). In that study, crosses with high carotenoid parents resulted in F2 progeny with β-carotene as high as 3.57 μg/g, suggesting that classical breeding can increase concentrations further. Overall, our findings, along with these previous studies, suggest that sorghum provitamin A carotenoid biofortification is feasible using breeding coupled with modern genomic breeding tools.

In order to ensure continuous genetic gains and trait improvement, genetic diversity is necessary. For carotenoid content in sorghum, however, this might be a limitation, because high carotenoid lines appear to be highly related (Figure 2.4A and Figure 2.5E-G). The tight clustering of countries (Figure 2.5E-G) and few countries with high carotenoid lines suggest that there has not been much exchange of germplasm among the countries and most germplasm might not have high carotenoid alleles. Interestingly, here we identified Nigeria, Lebanon, and the United States as the major origins with high carotenoid germplasm both in the SAP/CAP collection and the unexplored germplasm collections. However, it seems that the high carotenoid lines from Lebanon and the United States are from Nigerian origin. In the 1950s, a breeder named O.J. Webster collected yellow endosperm kaura sorghums from Nigeria, which were

subsequently used for breeding material in the United States (Maunder, 2000). Some of these kaura lines were then sent by another breeder, R.E. Karper, from the United States to the Arid Land Agricultural Development (ALAD) Program in Lebanon, which eventually became the International Center for Agricultural Research in the Dry Areas (ICARDA) (ICARDA, 2009; Mengesha & Rao, 1990). This relationship between the kaura accessions from the United States and Lebanon explains the close genetic similarity identified in our study between accessions from the United States and Lebanon (Figure 2.4A).

Selection for kaura types could also be the underlying driver of the limited genetic diversity and selection signals observed for the high carotenoid lines (Figure 2.4B). In Nigeria, kaura types are one of the most common sorghum landraces grown due to their high yield, drought resistance, and grain quality (Prasada Rao et al., 1985). They also have widely-sought agronomic traits as they are generally of short stature, have large yellow seeds, and are photoperiod insensitive. In the United States, selection for kaura types could also have contributed further to the limited genetic diversity. The first yellow hybrids developed in the United States, using Nigerian germplasm, had stronger root development, improved stay-green, and resistance to charcoal rot (Maunder, 2000), which could have led to the incorporation of high carotenoid alleles into multiple pedigrees. Also, among the high carotenoid accessions identified in our study, several of them are listed in GRIN as kaura (durra-caudatums), which could support the indirect selection for carotenoids among the kaura types. All together, these results suggest that kaura types from Nigeria are the main source of high carotenoid alleles and that efforts to increase diversity can focus on them.

Here, we identified 107 accessions from Nigeria, Niger, Ethiopia, Senegal, Sudan and the SAP/CAP collection (Supplemental Table 4) that have a high GEBV for β-carotene. These accessions need to be phenotyped with HPLC to test the hypothesis that they have high β-carotene concentrations and have potential as donor parents for breeding efforts. Importantly, some of the lines hypothesized to have high carotenoids based on GEBV are highly genetically divergent from the high-carotenoid Nigerian germplasm (e.g. Ethiopia and Sudanese lines in Figure 2.5, Supplemental Table 3), suggesting that previously untapped high-carotenoid germplasm exists. Given that sorghum germplasm does not meet current target values, breeding will be necessary to increase β-carotene concentrations. The high heritabilities for β-carotene (Cruet-Burgos et al., 2020; Fernandez et al., 2008) suggest that increasing concentrations through breeding is possible. Developing crosses among high β-carotene lines in the SAP/CAP collection as well as those identified by genomic prediction, can provide insights into if there is enough genetic diversity to reach target values. Genomics-assisted breeding via MAS or GS has the potential to accelerate efforts by simplifying selection methods.

Marker-assisted selection for sorghum carotenoids

MAS could be a viable alternative to select carotenoids in sorghum given that GWAS suggests an oligogenic architecture (Figure 2.2 and (Cruet-Burgos et al., 2020)). MAS for carotenoids has been tested in cassava with a marker linked to the PSY gene, which initiates the first committed step in the carotenoid biosynthesis pathway, demonstrating prediction accuracies above 0.8 (Gelli et al., 2017). MAS has also been

implemented in maize biofortification efforts with markers in linkage with the biosynthesis genes lcyE (Babu et al., 2013), crtRB1 (Babu et al., 2013; Gebremeskel et al., 2018; Muthusamy et al., 2014), ZEP (Gebremeskel et al., 2018), and opaque 2 (Gupta et al., 2013). For a successful implementation of MAS for sorghum carotenoids, breeder-friendly markers (i.e. convenient and low-cost) with tight linkage and high LD with target alleles must be developed (Cobb et al., 2019).

Chromosome 6 might be a good place to start for sorghum carotenoid marker development due to the high number of associations detected for β-branch carotenoids, where most of the provitamin A carotenoids are synthesized. Four regions of association on chromosome 6 have been identified: 46.7 Mb (Cruet-Burgos et al., 2020), 50.3-53.5 Mb (Fernandez et al., 2008), 57.4 Mb (Cruet-Burgos et al., 2020) and 47.1 Mb in this study. For β-carotene, associations have been detected in proximity to PDS (Sobic.006G232600) (Cruet-Burgos et al., 2020), the second enzyme in the carotenoid biosynthesis pathway. For zeaxanthin, associations have been detected near δ-carotene desaturase (ZDS, Sobic.006G177400) (Fernandez et al., 2008) and ZEP (Sobic.006G097500) (Cruet-Burgos et al., 2020). Interestingly, in this study we also identified significant associations near ZEP for lutein, zeaxanthin, and β-carotene (Figure 2.2, Supplemental Table 3). These genes have also been associated with natural variation of carotenoids in maize (Diepenbrock et al., 2021). If no linkage drag is present, the prevalence of associations on chromosome six could mean that positive alleles for multiple genes could be introduced simultaneously, reducing the generations needed. Understanding the allelic diversity of these genes in sorghum germplasm and the

expression profiles among high carotenoid germplasm could further demonstrate their potential for utilization in vitamin A biofortification.

The ZEP gene in sorghum could be a candidate to initiate such efforts. Based on this and our previous study (Cruet-Burgos et al., 2020) ZEP is a core gene controlling variation in the sorghum β-branch carotenoids, i.e. zeaxanthin and β-carotene. ZEP is also a good candidate for breeding efforts though MAS as it seems to have allelic variants with strong correlation with carotenoid concentrations. Marker S06_46717975 is a biallelic allele (A/G) in the ZEP coding sequence and is in proximity to S06_47123508, here associated with β-carotene and zeaxanthin. In the SAP/CAP collection, the allele was minor with only 10% of the accessions having the 'A' allele. The geographic distribution of the 'A' allele also correlates with Nigeria, United States, and Lebanon, the countries that had the highest observed or predicted β-carotene concentrations (Figure 2.6). Interestingly, among the unexplored germplasm collections, the top 5% had a more balanced distribution of the A/G variants. One hypothesis is that the top 5% of β-carotene GEBV might capture a wider carotenoid content and diversity than what is present in the SAP/CAP collection and therefore the allele has not been fixed (Supplemental Table 6). The higher prevalence of countries in the top 5% of GEBV for β-carotene, the lower GEBV when compared to the SAP/CAP collection (Supplemental Figure 1), and the more sparsely grouped cluster (Figure 2.5G) here observed supports this hypothesis. However, we hypothesize that due to the high correlation of allelic variant and carotenoid content observed and predicted, marker S06_46717975 could be used for MAS and to identify potential donor lines. Further germplasm evaluations are needed to assess this marker's predictive ability.

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

Genomic selection for sorghum carotenoids

Genomic selection is an alternative to MAS that is increasingly used for complex traits as genotypic costs decrease. For quality traits GS could potentially reduce the cost compared to phenotyping, and reduce the need for specialized equipment and training. In wheat, GS has been proven to be superior for quality traits over MAS as it allows for the selection for both small and large effect loci (Guzman et al., 2016; Plavšin et al., 2021; Yao et al., 2018). GS for carotenoids has yet to be implemented in breeding programs, but it has been tested in cassava (Esuma et al., 2021) and maize (Owens et al., 2014). Here we report the first study on genomic prediction for sorghum carotenoids. Genomic predictions are designed to capture polygenic variance and allow for selection on complex traits. GS accuracy and efficiency is dependent on several factors, particularly heritability of the trait, because it often directly translates into the prediction accuracy. The high heritability estimates (0.78, Table 2.1) and prediction accuracy (0.67, Table 2.2) here obtained for β-carotene would suggest that there is a polygenic component to sorghum carotenoids and GS can be an efficient method for biofortification. One hypothesis that explains why we see evidence for both oligogenic and polygenic variation is that sorghum

carotenoids are omnigenic traits, in which a small number of core genes directly regulate carotenoids and a large number of peripheral genes that are expressed in the grain indirectly regulate carotenoids (Boyle et al., 2017). This hypothesis could be tested with a genome-wide expression study in high carotenoid germplasm.

Despite its potential, there are several factors to consider in GS for sorghum carotenoids. First, the results of this study suggest that most countries' germplasm currently lack enough phenotypic and genotypic diversity for sorghum carotenoids. This suggests that the next step for provitamin A biofortification would be to introduce high carotenoid alleles into these breeding programs via prebreeding. Given that oligogenic variation for grain carotenoids exists, this initial introduction of alleles could be accomplished with MAS. Second, even though GS has the potential to reduce cost of phenotyping, simulation studies suggest that depending on population size, genotyping costs must be under $15 (U.S. Dollars) to be more cost-effective than simple phenotypic selection (Muleta et al., 2019). This genotyping cost can make GS unrealistic for breeding programs in developing countries, which would be the ones to benefit the most from a biofortified sorghum, as it is estimated that genotyping several hundred SNP markers remains at $14 (U.S. Dollars) (Bernardo, 2021).

Lastly, incorporation of a GS scheme for a young breeding program can be very challenging. GS has the biggest potential for genetic gain per unit of time when breeding cycles are closed rapidly and effectively (Bernardo, 2021; Heffner et al., 2010; Muleta et al., 2019). However, many breeding programs in developing countries have slow breeding cycles with recycling improved lines as a parent often taking well over 10 years (Atlin et al., 2017). Therefore, under these scenarios we suggest the direction for biofortification breeding will be to first introduce major genes through MAS in breeding programs. After

the introduction of these alleles, and as genotyping cost continues to decrease, MAS in tandem with GS can then be used for continuous improvement. If carotenoid variation in sorghum is in fact both oligogenic and polygenic, then the incorporation of MAS, GS, and rapid breeding cycles could substantially increase β-carotene to target values and ensure continuous genetic gains.

CONCLUSIONS

In this study we evaluated carotenoid concentrations in SAP/CAP collection, identifying the accessions with the highest β-carotene concentrations. Also, it was established that current concentrations of β-carotene are low and current known high β-carotene germplasm has a narrow genetic diversity. We used the SAP/CAP collection as a training population to predict the genetic merit or GEBV via genomic prediction for unexplored germplasm. Based on GEBV, we present 107 accessions with the highest predicted concentrations for β-carotene that potentially represent novel genetic variation for the trait. Finally, we proposed that MAS should be initially used to introduce high carotenoid alleles like S06_46717975 inside the *ZEP* gene into breeding programs followed by GS for continuous improvement.

LITERATURE CITED

Abdel-Aal, E.-S., Akhtar, H., Zaheer, K., & Ali, R. (2013). Dietary Sources of Lutein and Zeaxanthin Carotenoids and Their Role in Eye Health. *Nutrients*, *5*(4), 1169–1185. https://doi.org/10.3390/nu5041169

Afify, A. E.-M. M., El-Beltagi, H. S., El-Salam, S. M. A., & Omran, A. A. (2012). Biochemical changes in phenols, flavonoids, tannins, vitamin E, β–carotene and antioxidant activity during soaking of three white sorghum varieties. *Asian Pacific Journal of Tropical Biomedicine*, *2*(3), 203–209. https://doi.org/10.1016/S2221-1691(12)60042-2

Atlin, G. N., Cairns, J. E., & Das, B. (2017). Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Security*, *12*, 31–37. https://doi.org/10.1016/j.gfs.2017.01.008

Azmach, G., Menkir, A., Spillane, C., & Gedil, M. (2018). Genetic Loci Controlling Carotenoid Biosynthesis in Diverse Tropical Maize Lines. *G3 Genes|Genomes|Genetics*, *8*(3), 1049–1065. https://doi.org/10.1534/g3.117.300511

Babu, R., Rojas, N. P., Gao, S., Yan, J., & Pixley, K. (2013). Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theoretical and Applied Genetics*, *126*(2), 389–399. https://doi.org/10.1007/s00122-012-1987-3

Bernardo, R. (2021). Upgrading a maize breeding program via two-cycle genomewide selection: Same cost, same or less time, and larger gains. *Crop Science*, *61*(4), 2444–2455. https://doi.org/10.1002/csc2.20516

Blessin, C. W., Dimler, R. J., & Webster, O. J. (1962). Carotenoids of corn and sorghum II. Carotenoid Loss in Yellow-Endosperm Sorghum Grain during Weathering. *Cereal Chemistry*, *39*, 389–392.

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

Bouis, H. (2018). Reducing Mineral and Vitamin Deficiencies through Biofortification: Progress Under HarvestPlus. In H. K. Biesalski & R. Birner (Eds.), *World Review of Nutrition and Dietetics* (Vol. 118, pp. 112–122). S. Karger AG. https://doi.org/10.1159/000484342

Bouis, H., Hotz, C., McClafferty, B., Meenakshi, J. V., & Pfeiffer, W. H. (2011). Biofortification: A New Tool to Reduce Micronutrient Malnutrition. *Food and Nutrition Bulletin*, *32*(1), S31–S40.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, *169*(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2018). *ASReml estimates variance components under a general linear*.

Cardoso, L. de M., Pinheiro, S. S., da Silva, L. L., de Menezes, C. B., de Carvalho, C. W. P., Tardin, F. D., Queiroz, V. A. V., Martino, H. S. D., & Pinheiro-Sant’Ana, H. M. (2015). Tocochromanols and carotenoids in sorghum (Sorghum bicolor L.): Diversity and stability to the heat treatment. *Food Chemistry*, *172*, 900–908.

https://doi.org/10.1016/j.foodchem.2014.09.117

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., Franks, C. D., & Kresovich, S. (2008). Community Resources and Strategies for Association Mapping in Sorghum. *Crop Science*, *48*(1), 30–40. https://doi.org/10.2135/cropsci2007.02.0080

Che, P., Zhao, Z.-Y., Glassman, K., Dolde, D., Hu, T. X., Jones, T. J., Gruis, D. F., Obukosia, S., Wambugu, F., & Albertsen, M. C. (2016). Elevated vitamin E content improves all- *trans* β-carotene accumulation and stability in biofortified sorghum. *Proceedings of the National Academy of Sciences*, *113*(39), 11040–11045. https://doi.org/10.1073/pnas.1605689113

Cobb, J. N., Biswas, P. S., & Platten, J. D. (2019). Back to the future: Revisiting MAS as a tool for modern plant breeding. *Theoretical and Applied Genetics*, *132*(3), 647–667. https://doi.org/10.1007/s00122-018-3266-4

Cruet-Burgos, C., Cox, S., Ioerger, B. P., Perumal, R., Hu, Z., Herald, T. J., Bean, S. R., & Rhodes, D. H. (2020). Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *The Plant Genome*, *13*(1). https://doi.org/10.1002/tpg2.20013

Cuevas, H. E., & Prom, L. K. (2020). Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan Sorghum Core collection. *BMC Genomics*, *21*(1), 88. https://doi.org/10.1186/s12864-020-6489-0

Cuevas, H. E., Rosa-Valentin, G., Hayes, C. M., Rooney, W. L., & Hoffmann, L. (2017). Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: Implications for germplasm conservation, evaluation, and

utilization in crop improvement. *BMC Genomics*, *18*(1), 108. https://doi.org/10.1186/s12864-016-3475-7

Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory: Selection index theory for genomic selection. *Journal of Animal Breeding and Genetics*, *124*(6), 331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Diepenbrock, C. H., Ilut, D. C., Magallanes-Lundback, M., Kandianis, C. B., Lipka, A. E., Bradbury, P. J., Holland, J. B., Hamilton, J. P., Wooldridge, E., Vaillancourt, B., Góngora-Castillo, E., Wallace, J. G., Cepela, J., Mateos-Hernandez, M., Owens, B. F., Tiede, T., Buckler, E. S., Rocheford, T., Buell, C. R., … DellaPenna, D. (2021). Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *The Plant Cell*, *33*(4), 882–900. https://doi.org/10.1093/plcell/koab032

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Esuma, W., Ozimati, A., Kulakow, P., Gore, M. A., Wolfe, M. D., Nuwamanya, E., Egesi, C., & Kawuki, R. S. (2021). Effectiveness of genomic selection for improving provitamin A carotenoid content and associated traits in cassava. *G3 Genes|Genomes|Genetics*, *11*(9), jkab160. https://doi.org/10.1093/g3journal/jkab160

FAOSTAT. (2021). *Food and Agriculture Organization of the United Nations*. http://www.fao.org/faostat/en/#data/QC

Faye, J. M., Maina, F., Hu, Z., Fonceka, D., Cisse, N., & Morris, G. P. (2019). Genomic

signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. *Ecology and Evolution*, *9*(10), 6038–6051. https://doi.org/10.1002/ece3.5187

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., & Kresovich, S. (2008). Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Science*, *48*(5), 1732–1743. https://doi.org/10.2135/cropsci2007.12.0684

Fu, W.-N. (1960). *AGRONOMIC CHARACTERISTICS, PROTEIN, AND CAROTENOID COMPOSITION OF SOME GRAIN SORGHUM VARIETIES, STRAINS, AND HYBRIDS---WITH EMPHASIS ON YELLOW ENDOSPERM TYPES* [Master of Science]. Oklahoma State University.

Gebremeskel, S., Garcia-Oliveira, A. L., Menkir, A., Adetimirin, V., & Gedil, M. (2018). Effectiveness of predictive markers for marker assisted selection of pro-vitamin A carotenoids in medium-late maturing maize (Zea mays L.) inbred lines. *Journal of Cereal Science*, *79*, 27–34. https://doi.org/10.1016/j.jcs.2017.09.001

Gelli, M., Konda, A. R., Liu, K., Zhang, C., Clemente, T. E., Holding, D. R., & Dweikat, I. M. (2017). Validation of QTL mapping and transcriptome profiling for identification of candidate genes associated with nitrogen stress tolerance in sorghum. *BMC Plant Biology*, *17*(1), 123. https://doi.org/10.1186/s12870-017-1064-9

Global Nutrition Report. (2014). *2014 Global Hunger Index: The challenge of hidden hunger* (p. 56).

Gupta, H. S., Raman, B., Agrawal, P. K., Mahajan, V., Hossain, F., & Thirunavukkarasu, N. (2013). Accelerated development of quality protein maize hybrid through

marker-assisted introgression of *opaque-2* allele. *Plant Breeding*, *132*(1), 77–82. https://doi.org/10.1111/pbr.12009

Guzman, C., Peña, R. J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., Rutkoski, J., Poland, J., & Battenfield, S. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Applied & Translational Genomics*, *11*, 3–8. https://doi.org/10.1016/j.atg.2016.10.004

Harvest Plus. (2022). *Biofortied Crops Released*. https://bcr.harvestplus.org/varieties_released_list?crop=Vitamin_A_Maize

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., & Sorrells, M. E. (2010). Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science*, *50*(5), 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology*, *4*(3), 210–218. https://doi.org/10.1016/S1369-5266(00)00163-1

Hu, Z., Olatoye, M. O., Marla, S., & Morris, G. P. (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

ICARDA. (2009). *Lebanon and ICARDA: 32 years of partnership for sustainable agriculture*. ICARDA.

Ikoma, Y., Matsumoto, H., & Kato, M. (2016). Diversity in the carotenoid profiles and the expression of genes related to carotenoid accumulation among citrus genotypes. *Breeding Science*, *66*(1), 139–147. https://doi.org/10.1270/jsbbs.66.139

Irakli, M. N., Samanidou, V. F., Katsantonis, D. N., Biliaderis, C. G., & Papadoyannis, I. N. (2016). Phytochemical profiles and antioxidant capacity of pigmented and

non-pigmented genotypes of rice (*Oryza sativa* L.). *Cereal Research Communications*, *44*(1), 98–110. https://doi.org/10.1556/0806.43.2015.033

Jeffery, J. L., Turner, N. D., & King, S. R. (2012). Carotenoid bioaccessibility from nine raw carotenoid-storing fruits and vegetables using an in vitro model. *Journal of the Science of Food and Agriculture*, *92*(13), 2603–2610. https://doi.org/10.1002/jsfa.5768

Jourdan, M., Gagné, S., Dubois-Laurent, C., Maghraoui, M., Huet, S., Suel, A., Hamama, L., Briard, M., Peltier, D., & Geoffriau, E. (2015). Carotenoid Content and Root Color of Cultivated Carrot: A Candidate-Gene Association Study Using an Original Broad Unstructured Population. *PLOS ONE*, *10*(1), e0116674. https://doi.org/10.1371/journal.pone.0116674

Kean, E. G., Bordenave, N., Ejeta, G., Hamaker, B. R., & Ferruzzi, M. G. (2011). Carotenoid bioaccessibility from whole grain and decorticated yellow endosperm sorghum porridge. *Journal of Cereal Science*, *54*(3), 450–459. https://doi.org/10.1016/j.jcs.2011.08.010

Kumar, J., Saripalli, G., Gahlaut, V., Goel, N., Meher, P. K., Mishra, K. K., Mishra, P. C., Sehgal, D., Vikram, P., Sansaloni, C., Singh, S., Sharma, P. K., & Gupta, P. K. (2018). Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (Triticum aestivum L.) using multi-locus and multi-traits GWAS. *Euphytica*, *214*(11), 219. https://doi.org/10.1007/s10681-018-2284-2

Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma, K., Acharya, C., Mitchell, S. E., Buckler, E. S., Brenton, Z., Kresovich, S., & Morris, G. P. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, *1*(6), e1400218.

https://doi.org/10.1126/sciadv.1400218

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Maina, F., Bouchet, S., Marla, S. R., Hu, Z., Wang, J., Mamadou, A., Abdou, M., Saïdou, A.-A., & Morris, G. P. (2018). Population genomics of sorghum (*Sorghum bicolor*) across diverse agroclimatic zones of Niger. *Genome*, *61*(4), 223–232. https://doi.org/10.1139/gen-2017-0131

Maunder, A. B. (2000). History of Cultivar Development in the United States: From "Memoirs of A.B. Maunder-Sorghum Breeder." In C. W. Smith & R. A. Frederiksen (Eds.), *Sorghum: Origin, History, Technology, and Production*. John Wiley & Sons.

Mengesha, M. H., & Rao, K. E. P. (1990). World Sorghum Germplasm Collection And Conservation. *Plant Genetics and Breeding*, *150*, 90–104. http://oar.icrisat.org/5157/

Menkir, A., Rocheford, T., Maziya-Dixon, B., & Tanumihardjo, S. (2015). Exploiting natural variation in exotic germplasm for increasing provitamin-A carotenoids in tropical maize. *Euphytica*, *205*(1), 203–217. https://doi.org/10.1007/s10681-015-1426-z

Monier, B., Casstevens, T. M., Bradbury, P. J., & Buckler, E. S. (2020). *rTASSEL: An R interface to TASSEL for association mapping of complex traits* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.07.21.209114

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D.,

Riera-Lizarazu, O., Brown, P. J., Acharya, C. B., Mitchell, S. E., Harriman, J., Glaubitz, J. C., Buckler, E. S., & Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences*, *110*(2), 453–458. https://doi.org/10.1073/pnas.1215985110

Muleta, K. T., Felderhoff, T., Winans, N., Walstead, R., Charles, J. R., Armstrong, J. S., Mamidi, S., Plott, C., Vogel, J. P., Lemaux, P. G., Mockler, T. C., Grimwood, J., Schmutz, J., Pressoir, G., & Morris, G. P. (2022). *The recent evolutionary rescue of a staple crop depended on over half a century of global germplasm exchange* [Preprint]. Evolutionary Biology. https://doi.org/10.1101/2021.05.11.443651

Muleta, K. T., Pressoir, G., & Morris, G. P. (2019). Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. *G3 Genes|Genomes|Genetics*, *9*(2), 391–401. https://doi.org/10.1534/g3.118.200932

Muthusamy, V., Hossain, F., Thirunavukkarasu, N., Choudhary, M., Saha, S., Bhat, J. S., Prasanna, B. M., & Gupta, H. S. (2014). Development of β-Carotene Rich Maize Hybrids through Marker-Assisted Introgression of β-carotene hydroxylase Allele. *PLoS ONE*, *9*(12), e113583. https://doi.org/10.1371/journal.pone.0113583

Olatoye, M. O., Hu, Z., Maina, F., & Morris, G. P. (2018). Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. *G3 Genes|Genomes|Genetics*, *8*(10), 3269–3281. https://doi.org/10.1534/g3.118.200551

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C. R., Buckler, E. S., DellaPenna, D., Gore, M. A., & Rocheford, T. (2014). A

Foundation for Provitamin A Biofortification of Maize: Genome-Wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics*, *198*(4), 1699–1716. https://doi.org/10.1534/genetics.114.169979

Plavšin, I., Gunjača, J., Šatović, Z., Šarčević, H., Ivić, M., Dvojković, K., & Novoselović, D. (2021). An Overview of Key Factors Affecting Genomic Selection for Wheat Quality Traits. *Plants*, *10*(4), 745. https://doi.org/10.3390/plants10040745

Prasada Rao, K. E., Obilana, A. T., & Mengesha, M. H. (1985). Collection of Kaura, Fara-Fara and Guineense sorghums in Northern Nigeria. *Journal d'agriculture Traditionnelle et de Botanique Appliquée*, *32*(1), 73–81. https://doi.org/10.3406/jatba.1985.3930

Salas Fernandez, M. G., Kapran, I., Souley, S., Abdou, M., Maiga, I. H., Acharya, C. B., Hamblin, M. T., & Kresovich, S. (2009). Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. *Genetic Resources and Crop Evolution*, *56*(7), 991–1000. https://doi.org/10.1007/s10722-009-9417-3

Shen, Y., Su, X., Rhodes, D. H., Herald, T. J., Xu, J., Chen, X., Smith, J. S., & Wang, W. (2017). The pigments of Sorghum Pericarp Are Associated with the Contents of Carotenoids and Pro-vitamin A. *International Journal of Food and Nutritional Science*, *6*(3), 48–56.

Sikora, E., & Bodziarczyk, I. (2012). *COMPOSITION AND ANTIOXIDANT ACTIVITY OF KALE (BRASSICA OLERACEA L. VAR. ACEPHALA) RAW AND COOKED*. 10.

Sun, W., & Hu, Y. (2013). EQTL Mapping Using RNA-seq Data. *Statistics in Biosciences*, *5*(1), 198–219. https://doi.org/10.1007/s12561-012-9068-3

Suryanarayana Rao, R., Rukmini, C., & Mohan, V. S. (1967). β-Carotene Context of

Some Yellow-Endosperm Varieties of Sorghum. *The Indian Journal of Agricultural Science*, *38*(4), 368–372.

Tanumihardjo, S. A. (Ed.). (2013). *Carotenoids and Human Health*. Humana Press. https://doi.org/10.1007/978-1-62703-203-2

Tembo, L. (2021). Production and Adoption of Transgenic Crops in Sub-Saharan Africa. *Asian Research Journal of Agriculture*, 32–41. https://doi.org/10.9734/arja/2021/v14i230122

Trono, D. (2019). Carotenoids in Cereal Food Crops: Composition and Retention throughout Grain Storage and Food Processing. *Plants*, *8*(12), 551. https://doi.org/10.3390/plants8120551

Wambugu, F., & Kamanga, D. (Eds.). (2014). *Biotechnology in Africa* (Vol. 7). Springer International Publishing. https://doi.org/10.1007/978-3-319-04001-1

WHO. (2009). *Global prevalence of vitamin A deficiency in populations at risk: 1995-2005. WHO Global database on vitamin A deficiency*. https://apps.who.int/iris/bitstream/handle/10665/44110/9789241598019_eng.pdf

Worzella, W. W., Khalidy, R., Badawi, Y., & Daghir, S. (1965). Inheritance of Beta-Carotene in Grain Sorghum Hybrids. *Crop Science*, *5*(6), 591–592.

Yao, J., Zhao, D., Chen, X., Zhang, Y., & Wang, J. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (Triticum aestivum L.). *The Crop Journal*, *6*(4), 353–365. https://doi.org/10.1016/j.cj.2018.05.003

Zhao, Z. (2007). The Africa Biofortified Sorghum Project– Applying Biotechnology to Develop Nutritionally Improved Sorghum for Africa. In Z. Xu, J. Li, Y. Xue, & W. Yang (Eds.), *Biotechnology and Sustainable Agriculture 2006 and Beyond* (pp.

273–277). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6635-1_41

Zhou, X., Mcquinn, R., Fei, Z., Wolters, A.-M. A., Van Eck, J., Brown, C., Giovannoni, J. J., & Li, L. (2011). Regulatory control of high levels of carotenoid accumulation in potato tubers. *Plant, Cell & Environment*, *34*(6), 1020–1030. https://doi.org/10.1111/j.1365-3040.2011.02301.x

CHAPTER 3. DEVELOPMENT AND VALIDATION OF BREEDER-FRIENDLY KASP MARKERS FOR VITAMIN A BIOFORTIFICATION IN SORGHUM

INTRODUCTION

Starchy staples such as cereals and tubers are the main source of calories in developing countries where vitamin A deficiency (VAD) is a major public health problem (World Health Organization, 2009). Biofortification efforts in major staples such as maize, cassava and sweet potato are underway as a sustainable strategy to counteract VAD (Andersson et al., 2017). Studies have shown the potential impact of these biofortified crops in improving the nutritional status of these communities (Gannon et al., 2014; Hotz, Loechl, de Brauw, et al., 2012; Hotz, Loechl, Lubowa, et al., 2012), therefore expanding biofortification efforts to other important staples can be beneficial. Sorghum is a major staple crop in sub-Saharan Africa where vitamin A is prevalent (FAOSTAT, 2021; World Health Organization, 2009), so is a good target for vitamin A biofortification. Currently, sorghum accumulates low concentrations of carotenoids in the grain (Blessin et al., 1962; Cardoso et al., 2015; Cruet-Burgos et al., 2020; Fernandez et al., 2008; Shen et al., 2017; Worzella et al., 1965), however, an early study showed that increasing pro-vitamin A carotenoids through breeding is possible (Worzella et al., 1965).

Vitamin A biofortification breeding relies on increasing pro-vitamin A carotenoids in the plants. Selection of grain nutrients is challenging for breeders, because the nutrients can only be phenotyped on physiologically mature grain at the end of the

growing season, so either 1) many crosses must be carried out without knowing which lines will be good donor lines, or 2) donor lines need to be selected after phenotyping and can not be crossed until the following growing season. Additionally, phenotyping is time-consuming, labor-intensive, and highly technical. The most commonly used method for direct carotenoid selection requires quantification via analytical methods using absorption-based reverse-phase high-performance liquid chromatography (HPLC) (Amorim-Carrilho et al., 2014; S. M. Rivera & Canela-Garayoa, 2012). HPLC-based quantification provides the most accurate measurement of carotenoids, but it requires expensive and specialized equipment, extensive user training, and it has a high per sample cost. Biofortification breeding requires phenotyping of thousands of progeny, so phenotyping is a difficult task, at best, and not feasible for many research or breeding programs (Rosales et al., 2022).

Lower cost quantification methods like near infrared (NIR) spectroscopy, a technique based on the absorption of electromagnetic radiation, has been emerging as an alternative to HPLC (Abincha et al., 2020; Belalcazar et al., 2016; Rosales et al., 2022). When compared with HPLC, it is estimated that NIR can process samples 200X times faster (24 hours per sample versus 7 minutes per sample) for a tenth of the cost (Rosales et al., 2022). However, NIR is not an ideal replacement for HPLC yet. Current methods are reliable in maize and cassava during early stages of selection in which high trait variation exists, but are not reliable during later stages of trait selection when variation among genotypes is small (Abincha et al., 2020; Belalcazar et al., 2016; Rosales et al., 2022). For sorghum, carotenoid evaluations have been conducted using HPLC (Cardoso

et al., 2015; Cruet-Burgos et al., 2020; Fernandez et al., 2008; Shen et al., 2017), but NIR has yet to be implemented.

Biofortification breeding efforts have also employed indirect selection methods for carotenoids such as color selection and marker-assisted selection (MAS). Carotenoids are pigments ranging from yellow to red, so selection based on color intensity can be used for breeding if high correlations exist between color and carotenoid content. In orange sweet potato (OSP), for example, orange flesh color and β-carotene content are highly correlated, and selection on darker orange flesh has resulted in β-carotene concentrations above the biofortification target levels (Virk et al., 2021). In sorghum, however, color-based phenotypic selection might not be an effective selection strategy, because only intermediate correlations have been found for color and pro-vitamin A β-carotene (Fernandez et al., 2008).

MAS is another indirect selection strategy, in which the presence or absence of a DNA variant is used as the method of selection. MAS has been successful in vitamin A biofortification efforts in maize using genetic markers targeting the carotenoid biosynthesis pathway genes β-Carotene Hydroxylase 1 (crtRB1) and lycopene epsilon cyclase (lcyE) (Zunjare et al., 2018). MAS is an effective molecular breeding strategy for oligogenic traits—traits controlled by a only few genes— (Hasan et al., 2021), so is a promising indirect selection method for sorghum grain carotenoids, which have an oligogenic architecture component (Cruet-Burgos et al., 2020).

In order to use MAS for sorghum biofortification, DNA markers must be developed and tested. For MAS to be successful, markers must be 1) in the functional allele or co-segregate with the functional allele, 2) able to accurately predict the trait, 3) polymorphic, and 4) cost- effective (Hasan et al., 2021; Mohler & Singrün, 2004). Kompetitive allele specific PCR (KASP) is one of the current leading technologies for marker development for molecular breeding due to their low cost, low DNA requirements and their co-dominant nature (can distinguish between heterozygous and homozygous individuals) (Semagn et al., 2014). The objectives of this study are to 1) develop KASP markers that are linked to genes associated with carotenoid variation in sorghum (Cruet-Burgos et al., 2020; Fernandez et al., 2008), 2) test the predictive ability of the markers in F2:3 biparental families derived from parents with different carotenoids profiles, and 3) assess the potential for increasing carotenoid concentrations in sorghum through MAS. The results of this study provide publicly available markers that can be used for sorghum vitamin A biofortification breeding.

METHODS

Marker development

SNPs in genomic regions previously associated with carotenoid concentration (Cruet-Burgos et al., 2020) or in carotenoid biosynthesis genes, were used to develop eleven molecular markers (Table 3.1). SNP S4_62459432 in chromosome 4 was associated with zeaxanthin concentration and is 46kb away from Sobic.004G281900, annotated as a 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase (MDS) involved in the carotenoid precursor MEP pathway. Two SNPs—S6_46330663 and S6_46717975—

in chromosome 6 were associated with zeaxanthin and are in proximity to Sobic.006G097500, annotated as a zeaxanthin epoxidase (ZEP). Three SNPs—S6_54257267, S6_54257276, and S6_54257380—have not been associated with carotenoid concentration before, but they are within the coding sequence of Sobic.006G188200 annotated as a β-carotene 3-hydroxylase (β-OH). Four SNPs—and S6_47643430, S8_7569911, S8_7570056 and S10_57162947—were associated with zeaxanthin concentration (Cruet-Burgos et al., 2020), but were not in proximity to any gene known to be involved in biosynthesis biosynthesis or degradation. Sequences from 100 bp upstream and downstream of the SNP were sent to Intertek AgriTech (Alnarp, Sweden) for KASP marker development

Table 3.1 SNP used to develop KASP markers and the genomic regions associated with carotenoid concentration.

SNP	Marker Name	Chr	Alleles	Association	Nearby *a priori* candidates gene	Distance to *a priori* candidate gene	Genic region
S4_62459432	snpSB00266	4	C/T	Zeaxanthin	MDS, ispF	46 kb	Non-genic
S6_46330663	snpSB00264	6	C/T	Zeaxanthin	ZEP	384 kb	Intron
S6_46717975	snpSB00265	6	G/A	Zeaxanthin	ZEP	0 kb	Exon
S6_47643430	snpSB00277	6	C/T	Zeaxanthin	QTL	-	Exon
S6_54257267	snpSB00279	6	A/G	NA	β-OH	0 kb	Intron
S6_54257276	snpSB00280	6	G/A	NA	β-OH	0 kb	Intron
S6_54257355	snpSB00281	6	G/T	NA	β-OH	0 kb	Intron
S6_54257380	snpSB00282	6	C/G	NA	β-OH	0 kb	Intron
S8_7569911	snpSB00267	8	A/T	Zeaxanthin	QTL	-	Non-genic
S8_7570056	snpSB00276	8	G/T	Zeaxanthin	QTL	-	Non-genic
S10_57162947	snpSB00268	10	C/A	Zeaxanthin	QTL	-	Non-genic

Family development

Eight sorghum inbred accessions—PI585348, PI585369, PI569812, PI562777, PI563430, PI563447, PI585365 and PI510917—were used as parents to develop six biparental F2:3 populations (Figure 3.1). The initial crosses were developed using the plastic bag emasculation method (Laxman, 1997). The F1 plants were grown in a greenhouse and leaf tissue was collected and sent to Intertek AgriTech (Alnarp, Sweden) for KASP marker testing to identify true hybrids. The true hybrids were then self-pollinated. The F2 of the PI585348 x PI585369 was grown in the Kansas State University North Farm during Summer 2019, and the remaining F2's were grown during Summer 2020. Leaf tissue of the F2 plants was collected and sent to Intertek AgriTech (Alnarp, Sweden) for KASP marker testing. The F2 plants were self-pollinated and the resulting F2:3 grain was harvested and stored at -80 ℃ until carotenoid extraction.

Carotenoid Extraction

Carotenoid extractions of the F2:3 and parental lines were carried out under yellow light to avoid photodegradation using a modified solid phase method (Irakli et al., 2011). Due to the large number of samples (n=930), only one biological and one technical replicate was analyzed. Briefly, 20 mg of the sorghum flour was obtained from approximately 5 seeds using a Bead Ruptor Elite (Omni International, Kennesaw, GA). Next, the flour was transferred to a 1.5mL eppendorf tube along with 20 mg of ascorbic acid and 400 µl of absolute ethanol with a 1 mg/mL concentration of butylated hydroxytoluene (BHT). Samples were vortexed for 1 minute and incubated for 5 minutes in a 80 °C water bath. Next, 20 µl of a solution of potassium hydroxide (80% w/v, in water)

was added and tubes vortexed for 1 minute. Samples were incubated again at 80 °C for 15 minutes mixing every 5 minutes. Following incubation, samples were brought to room temperature and centrifuged for 5 minutes at 1900 rcf. Supernatant was transferred to a new 1.5 mL eppendorf tube and an additional 400 µl of absolute ethanol with 1 mg/mL of BHT was added to the residue. The tube with residue was then vortexed for 1 minute and centrifuged for 5 minutes at 1900 rcf. After centrifugation, the supernatant was combined with the previous extract, vortexed for 30 seconds and centrifuged for 5 minutes at 5000 rcf. The combined supernatant was then transferred to a new 1.5mL eppendorf tube and evaporated to dryness with a gentle N2 stream at room temperature. Finally, the residue was reconstituted in 100 µl of Ethanol:Ethyl Acetate (1:1) and centrifuge for 5 minutes at 5000 rcf. An aliquot of 40 µl of the clear supernatant was utilized for the HPLC analysis. Resolution of lutein, zeaxanthin, β-carotene, α-carotene and β-cryptoxanthin was conducted using a Perkin Elmer LC 300 UHPLC (Waltham, Massachusetts) with a Zorbax SB-CN (Agilent Technologies, Santa Clara, California, U.S).

Marker evaluation

Predictive ability of the eleven KASP markers (Table 3.1) development was assessed with an analysis of variance ANOVA in R with the 'aov' function. Lutein, zeaxanthin and β-carotene concentration for the $F_{2:3}$ progenies for the six families was independently tested for each of the eleven markers. Markers with a ($P < 0.05$) were considered significant and differences of the means of genotype classes for each marker was tested with a least significant difference (LSD) with the 'LSD.test' function in the agricolae R package

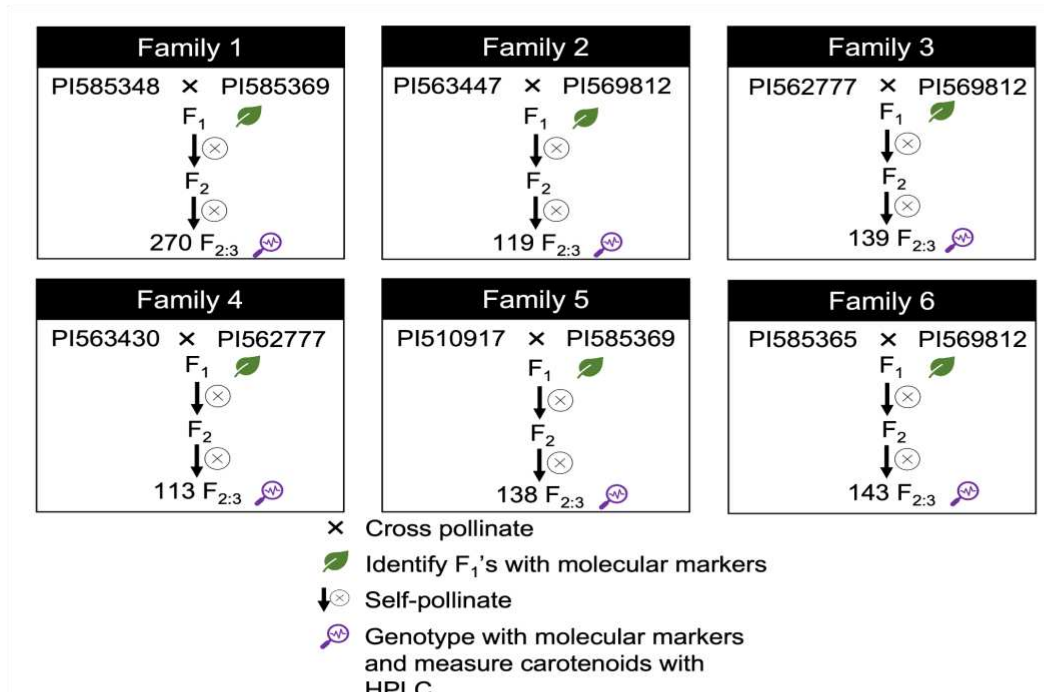


Figure 3.1 Family development crossing scheme and timeline for genotyping with KASP markers and phenotyping.

RESULTS

Carotenoid content of $F_{2:3}$ progenies

Carotenoid content of six $F_{2:3}$ biparental families and their parents were evaluated through HPLC to test the hypothesis that carotenoids can be increased in sorghum grain through breeding. Lutein and zeaxanthin were detected in the parental lines and in the progenies. β-carotene, β-cryptoxanthin and α-carotene were below the instrument detection limits. Lutein was the most abundant carotenoid for both the progenies and parents (Table 3.2). The average lutein concentration was 1.53 µg/g in the parental lines, and 1.64 µg/g in the progenies. For zeaxanthin, the average concentration was 1.25 µg/g in the parental lines and 2.08 µg/g in the progenies. Parental line PI585348 had the highest concentration for lutein whereas PI569812 had the highest concentration for zeaxanthin. Parental line PI510917 had the lowest concentrations for both lutein and

zeaxanthin (Table 3.2). Family 1 had the highest concentration for lutein (Figure 3.2A, Table 3.2), while Family 4 (Figure 3.3D, Table 3.2) had the highest concentration for zeaxanthin. Interestingly, for all the families there were progenies with concentrations both above and below the parental genotypes (Figure 3.2). For both zeaxanthin and lutein concentration, Family 1 (Figure 3.2A and 3.3A) had the highest range of distribution while Family 2 had the narrowest distribution (Figure 3.2B and 3.3B).

Table 3.2 Concentration of lutein and zeaxanthin for parents and distribution of $F_{2:3}$ progenies

Family	Compound (μg/g)	Parents		$F_{2:3}$ Progenies		
		PI585348	*PI585369*	*Mean*	*Max*	*Min*
Family 1	Lutein	2.63	1.92	2.11	6.42	0.33
	Zeaxanthin	1.74	1.42	1.53	3.62	0.35
		PI563447	*PI569812*	*Mean*	*Max*	*Min*
Family 2	Lutein	1.05	1.72	2.10	3.49	0.9
	Zeaxanthin	1.38	1.93	2.01	3.07	1.05
		PI562777	*PI569812*	*Mean*	*Max*	*Min*
Family 3	Lutein	1.25	1.72	2.36	4.18	1.02
	Zeaxanthin	1.16	1.93	2.15	3.97	1.12
		PI563430	*PI562777*	*Mean*	*Max*	*Min*
Family 4	Lutein	2.18	1.25	2.29	3.78	1.23
	Zeaxanthin	0.93	1.16	1.41	4.17	0.53
		PI510917	*PI585369*	*Mean*	*Max*	*Min*
Family 5	Lutein	0.68	1.92	1.66	2.93	0.96
	Zeaxanthin	0.52	1.42	1.21	2.59	0.58
		PI585365	*PI569812*	*Mean*	*Max*	*Min*
Family 6	Lutein	0.83	1.72	1.96	3.46	1.12
	Zeaxanthin	0.89	1.93	1.51	3.38	0.74
Average	Lutein	1.53		1.64		
	Zeaxanthin	1.25		2.08		

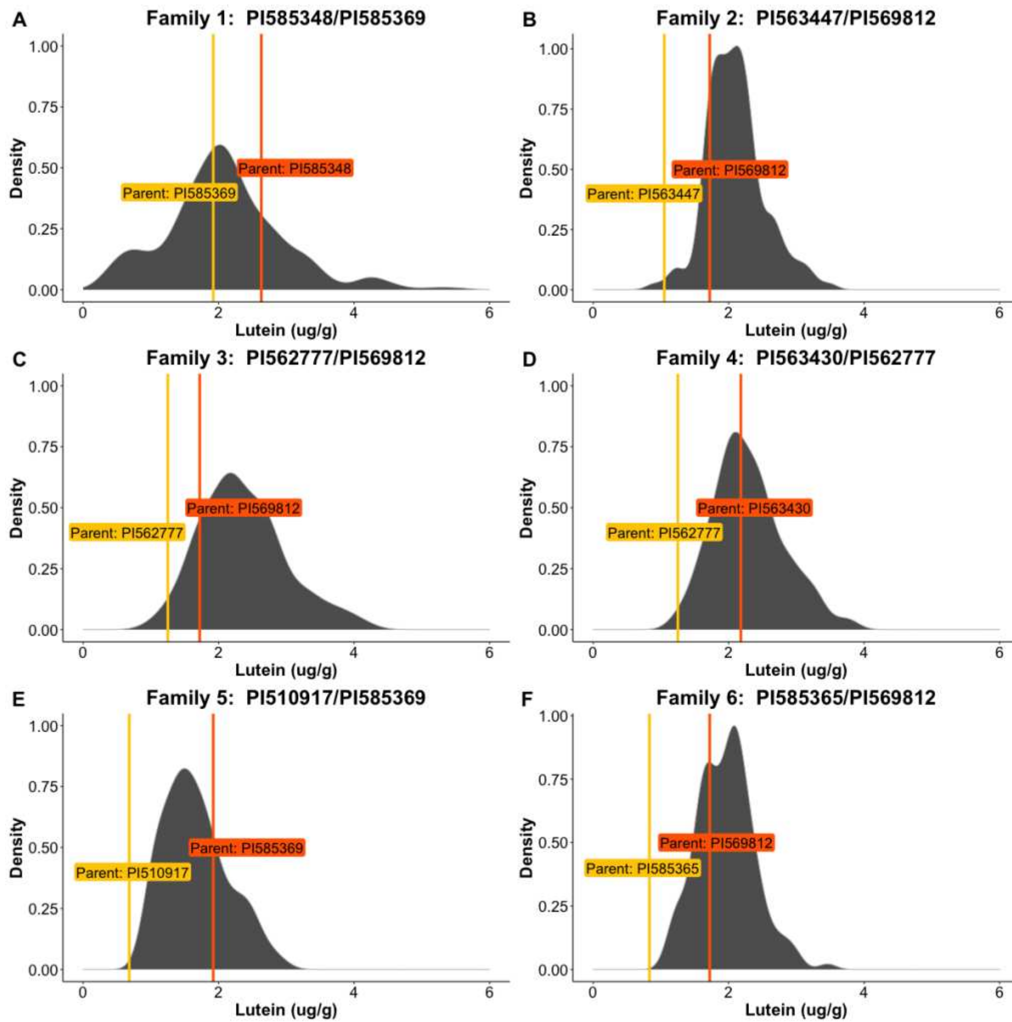


Figure 3.2 Lutein density plot for the six $F_{2:3}$ populations and their parents.
A) Distribution of family 1 and their parents PI585348 and PI585369; B) Distribution of family 2 and their parents PI563447 and PI569812; C) Distribution of family 3 and their parents PI562777 and PI569812; D) Distribution of family 4 and their parents PI563430 and PI562777; E) Distribution of family 5 and their parents PI510917 and PI585369; F) Distribution of family 6 and their parents PI585365 and PI569812. Horizontal line represents the parents' concentration.

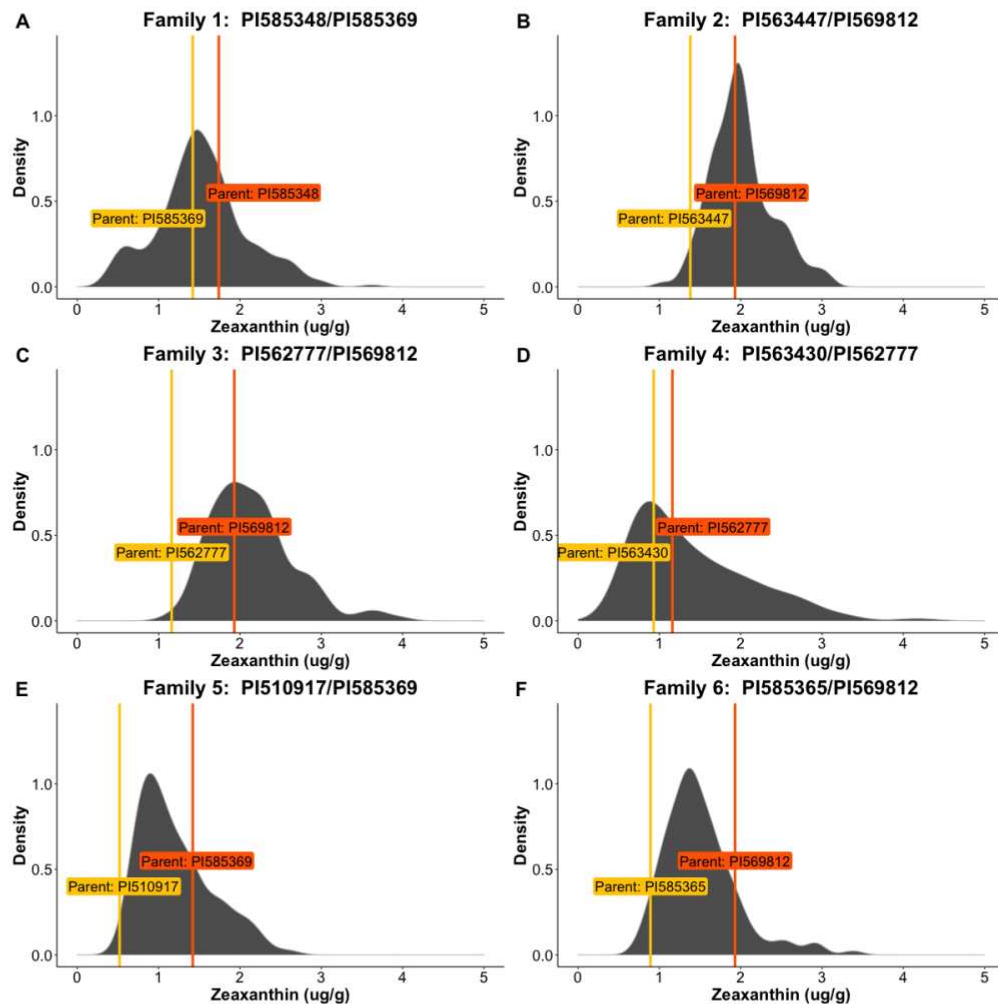


Figure 3.3 Zeaxanthin density plot for the six $F_{2:3}$ populations and their parents.
A) Distribution of family 1 and their parents PI585348 and PI585369 ; B)Distribution of family 2 and their parents PI563447 and PI569812; C) Distribution of family 3 and their parents PI562777 and PI569812; D) Distribution of family 4 and their parents PI563430 and PI562777; E) Distribution of family 5 and their parents PI510917 and PI585369; F) Distribution of family 6 and their parents PI585365 and PI569812. Horizontal line represents the parents' concentration.

Genotypic distribution of KASP

Eleven KASP markers were developed based on genomic regions associated with carotenoid concentration or genes involved in carotenoid biosynthesis. Among the progenies and parental lines, one marker, snpSB00281 (Table 3.1 and 3.2) was not

polymorphic and was thus discarded from subsequent analysis. Marker snpSB00282 had only two allelic classes present in the families and parental lines (Table 3.1 and 3.2). For the remaining nine KASP markers, three allelic classes were present among the parental lines and progenies (Supplemental Table 5). However, despite the presence of the allelic classes, all of the markers had a skewed distribution toward one of the genotypes for the progenies (Supplemental Table 5).

Association of KASP with carotenoid content

To test the hypothesis that marker-assisted selection can be used as a selection method for sorghum grain carotenoids, an ANOVA and LSD test between carotenoid content and genotype for ten KASP markers was performed (Table 3.3, Figure 3.4, Figure 3.5).

For lutein, seven markers were found to have significant differences between the means of the genotype groups (Table 3.3, Figure 3.4). Three markers—-snpSB00267, snpSB00276 and snpSB00277—-in proximity to QTLs for zeaxanthin had different means for the heterozygous groups, but were unable to differentiate between homozygous classes (Figure 3.4 A-D). Markers tagging the *β-OH*—-snpSB00279, snpSB00280 and snpSB00282—were all significant for lutein concentration. Two of them, snpSB00279 and snpSB00280, were able to differentiate between lutein content in homozygous classes but not between the homozygous and the heterozygous classes (Figure 3.4 E-F). snpSB00282 had only 2 allelic clases and despite being significantly associated with lutein and differentiating between the homozygous and heterozygous means, it had only 5 plants in the 'GC' genotype group (Figure 3.4G and Supplemental Table 4). Marker snpSB00268 was associated with lutein concentration and was able to differentiate between the mean of the two homozygous classes, but not between the heterozygous

and homozygous means (Figure 3.4B). This marker also had a skewed distribution with 243 of the 248 progenies genotypes in the 'AA' homozygous class.

For zeaxanthin, all of the markers were significantly associated with concentration (Table 3.3, Figure 3.5). Four markers— snpSB00264, snpSB00277, snpSB00270 and snpSB00280—had significant differences of zeaxanthin concentration and the means of the three genotype classes (Figure 3.5 A- I). Five markers—snpSB00265, snpSB00266, snpSB00267, snpSB00268 and snpSB00276—were able to group one homozygous class with the heterozygous and differentiate to the second homozygous class (Figure 3.5 B-F). Finally, snpSB00282 was able to differentiate between genotype classes for zeaxanthin content, although only two genotype groups were present.

Table 3.3 Description of KASP markers evaluated and their association with lutein and zeaxanthin content.

SNP	Marker Name	Chromosome	Alleles	Tag[1]	Lutein	Zeaxanthin
S4_62459432	snpSB00266	4	C/T	MDS	8.50E-01	3.50E-05 *
S6_46330663	snpSB00264	6	C/T	ZEP	1.10E-01	1.70E-31 *
S6_46717975	snpSB00265	6	G/A	ZEP	5.60E-01	1.00E-33 *
S6_47643430	snpSB00277	6	C/T	QTL	7.70E-03 *	2.60E-29 *
S6_54257267	snpSB00279	6	A/G	β-OH	4.10E-02 *	1.50E-22 *
S6_54257276	snpSB00280	6	G/A	β-OH	3.80E-02 *	6.00E-22 *
S6_54257380	snpSB00282	6	C/G	β-OH	4.80E-03 *	1.80E-02 *
S8_7569911	snpSB00267	8	A/T	QTL	5.90E-03 *	3.90E-05 *
S8_7570056	snpSB00276	8	G/T	QTL	4.60E-03 *	6.60E-05 *
S10_57162947	snpSB00268	10	C/A	QTL	1.60E-02 *	3.70E-03 *

* P-value< 0.05 for ANOVA; [1] Gene or region in proximity to the SNP.

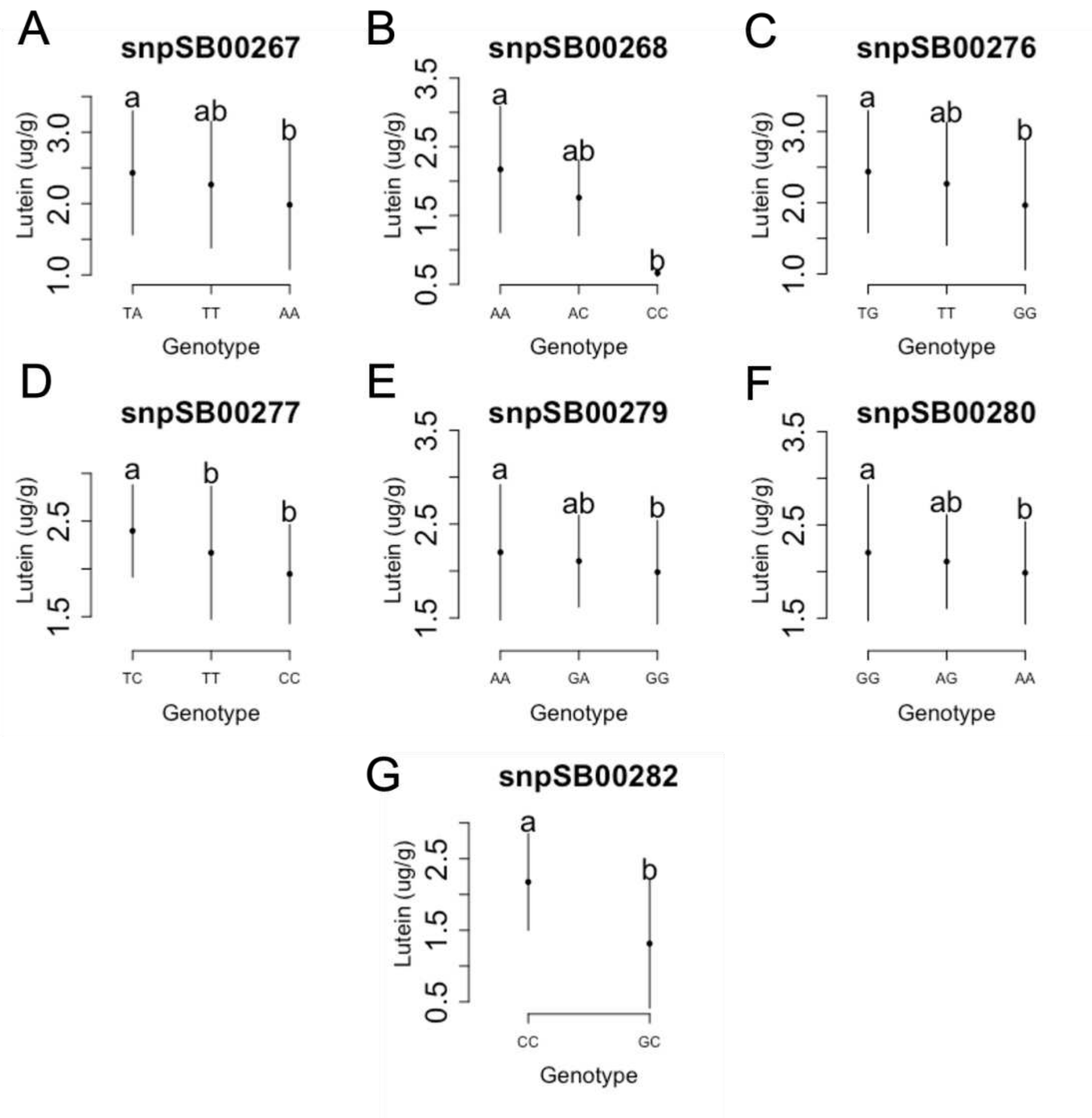


Figure 3.4 Least significant difference (LSD) for markers associated with lutein concentration in the $F_{2:3}$ progenies.

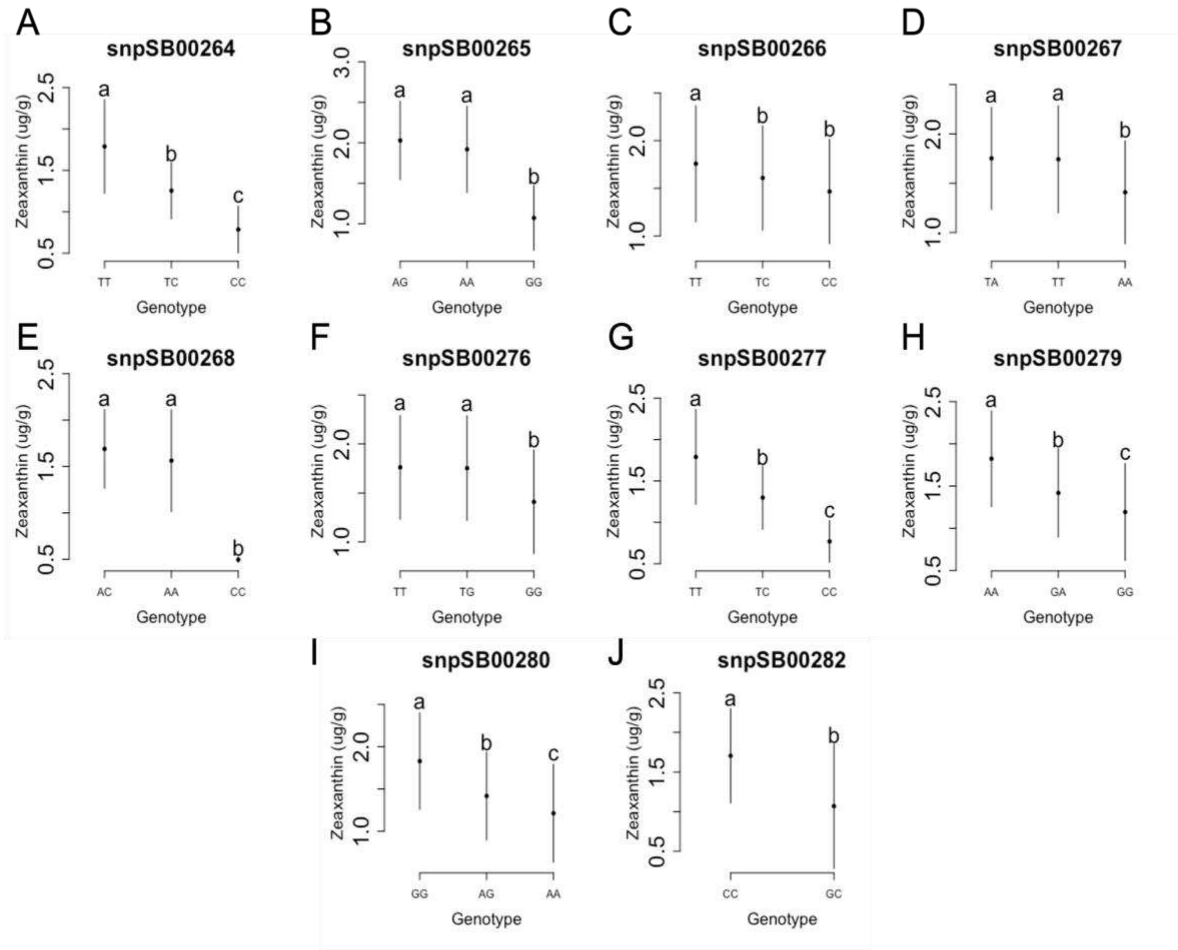


Figure 3.5 Least significant difference (LSD) for markers associated with zeaxanthin concentration in the $F_{2:3}$ progenies.

DISCUSSION

Carotenoid concentration in $F_{2:3}$ progenies

As a major staple crop in regions where VAD is prevalent, sorghum pro-vitamin A biofortification has the potential to alleviate vitamin A deficiencies. In order to be successful, increasing concentrations through breeding is necessary. Despite its potential, current research in sorghum breeding for carotenoids is extremely limited. In this study six biparental families were developed from parents with different carotenoid concentrations. Interestingly, progeny in all the families exhibited transgressive segregation, with carotenoid concentrations both above and below their two parental genotypes. Higher concentrations of β-carotene than parental genotypes was previously reported in F2 progenies when the two highest carotenoid genotypes were crossed (Worzella et al., 1965). However, contrary to results in the current study, when the crosses involved intermediate or low β-carotene parents, the concentration of the progenies was between the two parental genotypes (Worzella et al., 1965). A more recent study found similar patterns of inheritance as in this study, but in a recombinant inbred line (RIL) population (Fernandez et al., 2008). The authors identified progeny that had lutein, zeaxanthin and β-carotene concentrations both above and below parental genotypes (Fernandez et al., 2008). The existence of progenies outside of the parental range is promising for biofortification breeding, because it suggests that there is segregation at the loci controlling variation in carotenoid concentration and that fixation of these alleles can translate into higher concentrations. Crosses must be done between high carotenoid content progenies as well as with other parental lines to further test the potential of biofortification breeding and to increase concentrations of pro-vitamin A carotenoids.

Inferred gene action of candidate genes and breeding implications

Understanding the gene action underlying phenotypic variation can guide breeding programs in how to maximize their gains and to select what type of line development –hybrids or inbreds– is better suited (Fasoula & Fasoula, 2010). In this study, ten breeder friendly biallelic KASP markers were developed based on genomic regions previously associated with carotenoid concentration or genes known to be involved in carotenoid biosynthesis in other species (Table 3.1). Interestingly, among the developed markers different modes of gene actions were observed for both lutein and zeaxanthin concentration.

For lutein, three of the markers —snpSB00268, snpSB00279 and snpSB0080— seem to be in linkage with loci providing incomplete dominance or allelic dose effect with homozygous classes having either higher or lower lutein concentrations and the heterozygous group having an intermediate phenotype (Figure 3.4). Two of these markers—snpSB00279 and snpSB0080—are inside the non-coding sequence of β-OH, a β-carotene 3-hydroxylase (Figure 3.4E and F; Table 3.1) and were also detected for zeaxanthin with an additive gene action as well (Figure 3.5H and I). These markers are perhaps the most promising because in maize, markers tagging the β-carotene 3-hydroxylase homolog gene crtRB1 has also exhibited an allelic dose effect in maize and they have been used for biofortification breeding resulting in significant increases in β-carotene concentration (Babu et al., 2013; Chandran et al., 2019; Yan et al., 2010). Although in this study β-carotene was not able to be measured because it was below the instrument detection limits, previous studies have shown that there is a high correlation among lutein, zeaxanthin and β-carotene (Cruet-Burgos et al., 2020). As these two

markers—snpSB00279 and snpSB0080—are significant for both lutein and zeaxanthin, it can be hypothesized that they can also be used for β-carotene selections.

For zeaxanthin two additional markers—snpSB00264 in proximity to the ZEP (Figure 3.5A) and snpSB00277 in a QTL without any nearby candidate genes (Figure 3.5G)—also exhibited an additive gene action, while snpSB00266 in proximity to the MDS exhibited a recessive gene action (Figure 3.5C). Four markers—snpSB00265, snpSB00267, snpSB00268, snpSB00276—exhibited a complete dominance gene action, with the heterozygous groups and one of the homozygous groups having the same mean, while the second homozygous groups had lower zeaxanthin concentrations (Figure 3.5). Among these markers, snpSB00265 (Figure 3.5B) is the most promising for biofortification efforts as it is inside the coding sequence of the ZEP gene. The ZEP gene has been associated with zeaxanthin variation in diverse germplasm in sorghum as well as in maize (Cruet-Burgos et al., 2020; Owens et al., 2014). Although this marker was not significant for lutein concentration, due to its function in the β-branch it can be hypothesized that it can also predict β-carotene concentration.

CONCLUSIONS

MAS has the potential to accelerate vitamin A biofortification efforts as it can improve the efficiency of a breeding program. When compared to other methods of selection, such as direct selection by HPLC, MAS can reduce the complexity of phenotyping and allows for early selections. For sorghum, given the current low carotenoid concentrations, MAS might be particularly important because concentrations might be too low to be detected by the average HPLC so might require even more

specialized HPLC equipment with increased sensitivity. In this study two markers (snpSB00279 and snpSB0080) were developed inside the non-coding sequence of *β-OH* with the ability to predict lutein and zeaxanthin concentrations. A marker (snpSB00265) inside the *ZEP* gene was also developed that was able to predict zeaxanthin concentrations. Additionally, the hypothesis that carotenoid concentrations in sorghum grain can be increased through breeding was confirmed.

LITERATURE CITED

Abincha, W., Kawuki, R., Ikeogu, U., Obara, J., Hussin, Y., Egesi, C., Kulakow, P., Rabbi, I., Owor, B., & Parkes, E. (2020). Comparison of Near-infrared Spectroscopy with other options for total carotenoids content phenotyping in fresh cassava roots. *Journal of Scientific Agriculture*, 55–60. https://doi.org/10.25081/jsa.2020.v4.6196

Andersson, M., Saltzman, A., & Pfeiffer, W. H. (2017). Progress update: Crop development of biofortified staple food crops under HarvestPlus. *African Journal of Food, Agriculture, Nutrition and Development*, *17*(02), 11905–11935. https://doi.org/10.18697/ajfand.78.HarvestPlus05

Babu, R., Rojas, N. P., Gao, S., Yan, J., & Pixley, K. (2013). Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theoretical and Applied Genetics*, *126*(2), 389–399. https://doi.org/10.1007/s00122-012-1987-3

Belalcazar, J., Dufour, D., Andersson, M. S., Pizarro, M., Luna, J., Londoño, L., Morante, N., Jaramillo, A. M., Pino, L., Becerra López-Lavalle, L. A., Davrieux, F., Talsma, E. F., & Ceballos, H. (2016). High-Throughput Phenotyping and Improvements in Breeding Cassava for Increased Carotenoids in the Roots. *Crop Science*, *56*(6), 2916–2925. https://doi.org/10.2135/cropsci2015.11.0701

Blessin, C. W., Dimler, R. J., & Webster, O. J. (1962). Carotenoids of corn and sorghum II. Carotenoid Loss in Yellow-Endosperm Sorghum Grain during Weathering. *Cereal Chemistry*, *39*, 389–392.

Cardoso, L. de M., Pinheiro, S. S., da Silva, L. L., de Menezes, C. B., de Carvalho, C. W. P., Tardin, F. D., Queiroz, V. A. V., Martino, H. S. D., & Pinheiro-Sant'Ana, H.

M. (2015). Tocochromanols and carotenoids in sorghum (Sorghum bicolor L.): Diversity and stability to the heat treatment. *Food Chemistry*, *172*, 900–908. https://doi.org/10.1016/j.foodchem.2014.09.117

Chandran, S., Pukalenthy, B., Adhimoolam, K., Manickam, D., Sampathrajan, V., Chocklingam, V., Eswaran, K., Arunachalam, K., Joikumar meetei, L., Rajasekaran, R., Muthusamy, V., Hossain, F., & Natesan, S. (2019). Marker-Assisted Selection to Pyramid the Opaque-2 (O2) and β-Carotene (crtRB1) Genes in Maize. *Frontiers in Genetics*, *10*. https://www.frontiersin.org/article/10.3389/fgene.2019.00859

Cruet-Burgos, C., Cox, S., Ioerger, B. P., Perumal, R., Hu, Z., Herald, T. J., Bean, S. R., & Rhodes, D. H. (2020). Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *The Plant Genome*, *13*(1). https://doi.org/10.1002/tpg2.20013

Fasoula, D. A., & Fasoula, V. A. (2010). Gene Action and Plant Breeding. In J. Janick (Ed.), *Plant Breeding Reviews* (pp. 315–374). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470650097.ch9

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., & Kresovich, S. (2008). Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Science*, *48*(5), 1732–1743. https://doi.org/10.2135/cropsci2007.12.0684

Gannon, B., Kaliwile, C., Arscott, S. A., Schmaelzle, S., Chileshe, J., Kalungwana, N., Mosonda, M., Pixley, K., Masi, C., & Tanumihardjo, S. A. (2014). Biofortified orange maize is as efficacious as a vitamin A supplement in Zambian children

even in the presence of high liver reserves of vitamin A: A community-based, randomized placebo-controlled trial123456. *The American Journal of Clinical Nutrition*, *100*(6), 1541–1550. https://doi.org/10.3945/ajcn.114.087379

Hasan, N., Choudhary, S., Naaz, N., Sharma, N., & Laskar, R. A. (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

Hotz, C., Loechl, C., de Brauw, A., Eozenou, P., Gilligan, D., Moursi, M., Munhaua, B., van Jaarsveld, P., Carriquiry, A., & Meenakshi, J. V. (2012). A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *British Journal of Nutrition*, *108*(1), 163–176. https://doi.org/10.1017/S0007114511005174

Hotz, C., Loechl, C., Lubowa, A., Tumwine, J. K., Ndeezi, G., Nandutu Masawi, A., Baingana, R., Carriquiry, A., de Brauw, A., Meenakshi, J. V., & Gilligan, D. O. (2012). Introduction of β-Carotene–Rich Orange Sweet Potato in Rural Uganda Resulted in Increased Vitamin A Intakes among Children and Women and Improved Vitamin A Status among Children. *The Journal of Nutrition*, *142*(10), 1871–1880. https://doi.org/10.3945/jn.111.151829

Irakli, M. N., Samanidou, V. F., & Papadoyannis, I. N. (2011). Development and validation of an HPLC method for the simultaneous determination of tocopherols, tocotrienols and carotenoids in cereals after solid-phase extraction: Liquid Chromotography. *Journal of Separation Science*, *34*(12), 1375–1382. https://doi.org/10.1002/jssc.201100077

Laxman, S. (1997). Plastic bag emasculation technique in sorghum (Sorghum bicolor L.

(Moench)). *Journal of Research ANGRAU*, *25*(2), 38–39.

Mohler, V., & Singrün, C. (2004). General Considerations: Marker-Assisted Selection. In: Lörz, H., Wenzel, G. (eds) Molecular Marker Systems in Plant Breeding and Crop Improvement. In *Biotechnology in Agriculture and Forestry* (Vol. 55, pp. 305–317).

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C. R., Buckler, E. S., DellaPenna, D., Gore, M. A., & Rocheford, T. (2014). A Foundation for Provitamin A Biofortification of Maize: Genome-Wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics*, *198*(4), 1699–1716. https://doi.org/10.1534/genetics.114.169979

Rosales, A., Crossa, J., Cuevas, J., Cabrera-Soto, L., Dhliwayo, T., Ndhlela, T., & Palacios-Rojas, N. (2022). Near-Infrared Spectroscopy to Predict Provitamin A Carotenoids Content in Maize. *Agronomy*, *12*(5), 1027. https://doi.org/10.3390/agronomy12051027

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, *33*(1), 1–14. https://doi.org/10.1007/s11032-013-9917-x

Shen, Y., Su, X., Rhodes, D. H., Herald, T. J., Xu, J., Chen, X., Smith, J. S., & Wang, W. (2017). The pigments of Sorghum Pericarp Are Associated with the Contents of Carotenoids and Pro-vitamin A. *International Journal of Food and Nutritional Science*, *6*(3), 48–56.

Virk, P. S., Andersson, M. S., Arcos, J., Govindaraj, M., & Pfeiffer, W. H. (2021).

Transition From Targeted Breeding to Mainstreaming of Biofortification Traits in Crop Improvement Programs. *Frontiers in Plant Science*, *12*, 703990. https://doi.org/10.3389/fpls.2021.703990

World Health Organization. (2009). *Global prevalence of vitamin A deficiency in populations at risk 1995–2005*. World Health Organization. https://apps.who.int/iris/handle/10665/101008

Worzella, W. W., Khalidy, R., Badawi, Y., & Daghir, S. (1965). Inheritance of Beta-Carotene in Grain Sorghum Hybrids. *Crop Science*, *5*(6), 591–592.

Yan, J., Kandianis, C. B., Harjes, C. E., Bai, L., Kim, E.-H., Yang, X., Skinner, D. J., Fu, Z., Mitchell, S., Li, Q., Fernandez, M. G. S., Zaharieva, M., Babu, R., Fu, Y., Palacios, N., Li, J., DellaPenna, D., Brutnell, T., Buckler, E. S., … Rocheford, T. (2010). Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nature Genetics*, *42*(4), 322–327. https://doi.org/10.1038/ng.551

Zunjare, R. U., Hossain, F., Muthusamy, V., Baveja, A., Chauhan, H. S., Bhat, J. S., Thirunavukkarasu, N., Saha, S., & Gupta, H. S. (2018). Development of Biofortified Maize Hybrids through Marker-Assisted Stacking of β-Carotene Hydroxylase, Lycopene-ε-Cyclase and Opaque2 Genes. *Frontiers in Plant Science*, *9*, 178. https://doi.org/10.3389/fpls.2018.00178

CHAPTER 4. UNRAVELING TRANSCRIPTOMICS OF SORGHUM GRAIN CAROTENOIDS: A STEP FORWARD FOR BIOFORTIFICATION

INTRODUCTION

Sorghum is a staple crop in many countries in South East Asia and Africa where vitamin A deficiency (VAD) is a public health concern (FAOSTAT, 2021; World Health Organization, 2009). Sorghum grain accumulates low concentrations of the pro-vitamin A carotenoids β-carotene, α-carotene, and β-cryptoxanthin, thus increasing their concentrations could improve the nutritional status of sorghum-consuming communities. Studies have demonstrated that sorghum carotenoids are controlled by genetic factors, suggesting that biofortification through breeding is possible (Cruet-Burgos et al., 2020; Fernandez et al., 2008; Worzella et al., 1965). However, understanding of the mechanisms underlying carotenoid biosynthesis and its regulation in sorghum grain is limited.

Carotenoid accumulation in sorghum is dependent on the developmental stage of the grain. A study on eight varieties of sorghum demonstrated that carotenoids accumulate in the grain differentially through development (Kean et al., 2007). Carotenoid accumulation begins at around 10 days after half bloom (DAHB), reaches peak accumulation at 30 DAHB and then starts decreasing (Kean et al., 2007). Similar differential accumulation of carotenoids through development has been observed for maize (Da Silva Messias et al., 2014), tomato (Bramley, 2002), and carrots (Clotault et al., 2008; Jin et al., 2019; Perrin et al., 2017). Even though these studies each focus on

different plant tissues—grain, fruit, and root, respectively—associations in each of the three crops have been identified between carotenoid accumulation through development and transcriptional differences (Bramley, 2002; Clotault et al., 2008; Da Silva Messias et al., 2014; Harjes et al., 2008; Perrin et al., 2017). Given that carotenoid biosynthesis is an essential pathway that is highly conserved in photosynthetic organisms (Armstrong et al., 1990; Armstrong & Hearst, 1996; Bartley et al., 2003; Hirschberg, 2001; Li et al., 2008), sorghum grain carotenoid accumulation through development could also be driven by transcriptional regulation. Alternatively, since the pattern of carotenoid accumulation in sorghum grain corresponds with the milky, soft dough, and hard dough developmental stages (Gerik, n.d.; Roozeboom & Prasad, 2019), the differences in carotenoid accumulation could be a result of differences in nutrient mobilization, sequestration, storage, or plastid biogenesis. Understanding the mechanisms underlying carotenoid accumulation through development can help identify gene targets for breeding efforts.

Variation in carotenoid content in sorghum grain has also been identified among genotypes (Blessin et al., 1958; Cruet-Burgos et al., 2020; Debelo et al., 2020; Kean et al., 2007), with yellow endosperm accessions accumulating higher concentrations. Genetic studies have found associations between carotenoid content and a few regions on the genome, indicating an oligogenic inheritance (Cruet-Burgos et al., 2020; Fernandez et al., 2008; Worzella et al., 1965). Many of the significant associations have been found near a priori candidate genes in the carotenoid biosynthesis (Cruet-Burgos et al., 2020; Fernandez et al., 2008), carotenoid degradation (Fernandez et al., 2008), and the carotenoid precursor methylerythritol phosphate (MEP) pathways (Cruet-Burgos et al., 2020), implying allelic variants in these genes are driving the variation in carotenoid

content among sorghum genotypes. In maize, genomic studies have also suggested that grain carotenoids are oligogenic (Azmach et al., 2018; Chander et al., 2008; Diepenbrock et al., 2021; Kandianis et al., 2013; Owens et al., 2014; Suwarno et al., 2015), and characterization of some of the allelic variants have shown differential gene expression among genotypes (Da Silva Messias et al., 2014; Diepenbrock et al., 2021; Z. Fu et al., 2013; Harjes et al., 2008; Vallabhaneni & Wurtzel, 2009; Yan et al., 2010). However, some of the associated genomic regions could also be involved in other types of regulation, such as post-transcriptional regulation, affecting the enzymatic activities of these candidate genes. Knowledge in expression patterns of sorghum candidate genes in different genotypes could provide a better understanding of the mechanisms underlying carotenoid variation, and help guide molecular breeding efforts by identifying which genes may be more impactful and should be prioritized.

Carotenoid regulation in plants is not fully understood, but it is hypothesized that they are regulated at multiple levels (T. Sun et al., 2022, p. 20; T. Sun & Li, 2020; Zhai et al., 2016, p. 201) because they are essential compounds for many key plant activities such as photosynthesis and photoprotection, as well as precursors to other important molecules such as abscisic acid (ABA). This implies that by direct or indirect regulation, many genes outside of the main carotenoid-related pathways must work in harmony to maintain adequate levels of carotenoids across multiple tissues, organs, and cellular compartments. Although most carotenoid studies in major crops have found evidence that carotenoids are oligogenic traits (Azmach et al., 2018; Cruet-Burgos et al., 2020; Fernandez et al., 2008; Kandianis et al., 2013; Kumar et al., 2018; Owens et al., 2014; Suwarno et al., 2015; Wong et al., 2004), there is evidence that they have a polygenic

architecture as well. The large number of quantitative trait loci (QTL) detected in these studies, as well as the low percentage of phenotypic variance explained by the QTL, suggest a more complex architecture. Furthermore, in maize, an RNA-seq experiment identified over 50 genes with expression associated with carotenoid concentration, among which only 19 of them were involved in carotenoid-related pathways (Z. Fu et al., 2013). Therefore, even though carotenoid biosynthesis and degradation in sorghum grain has been demonstrated to be controlled by only a handful of genomic regions, the total concentration observed in the grain could be dependent on many additive genetic factors.

A deeper understanding of carotenoid genetic architecture and regulation in sorghum grain will help to establish the best breeding schemes and selection methods for biofortification efforts. However, little is known about molecular mechanisms underlying sorghum carotenoid variation. We hypothesize that differences in carotenoid content are driven by differences in transcriptome profiles. To test this hypothesis, in this study we 1) identified which carotenoid *a priori* candidate genes are expressed in sorghum grain; 2) quantified transcriptional differences in carotenoid-related genes throughout sorghum grain development; and 3) characterized differences in transcriptional patterns among genotypes with contrasting carotenoid profiles. The results of this study can guide future breeding efforts for sorghum carotenoid biofortification by identifying genes underlying carotenoid variation at a transcriptional level.

METHODS

Plant Material

Four sorghum accessions—PI329435, PI510924, PI585347, and PI585348—with contrasting carotenoid profiles (Supplemental Table 6) were grown in the Plant Growth

Facilities greenhouse of Colorado State University from March 2021 to July 2021. Accessions were planted using a complete randomized design (CRD) with nine replicates per accession. Grain was collected at three time points 14, 28, and 42 days after pollination (DAP), here defined as the first day pollen emerged on the top part of the panicle. For each accession, 3 biological replicates of each time point were collected. Grain samples were immediately flash frozen with liquid nitrogen and stored at -80C until RNA extractions.

RNA extraction, libraries and sequencing

Total RNA was extracted from grain samples using a SDS-LiCl method (Vennapusa et al., 2020) with some modifications. Briefly, 500 ul of chilled extraction buffer (100mM Tris-HCL (pH=8), 25 mM EDTA 2Na, 2.5% PVP, 2.5 M NaCl, 2.5% β-Mercaptoethanol in DEPC-water), 2 steel grinding balls, and 6 sorghum seeds were added to 2mL tubes. Next, they were immediately transferred to a Bead Ruptor Elite (Omni International, Kennesaw, GA) and ground for 30 seconds at a speed of 4 m/s. Samples were transferred to ice and an additional 500 µl of chilled extraction buffer was added to each sample and mixed with a vortexer. The homogenate was incubated for 5 minutes at room temperature and 100 µl of 20% SDS was added to the suspension. Then, samples were incubated at room temperature for 2 minutes and centrifuged at 16,000xg for 5 minutes at 4℃. The supernatant was collected and 800 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added. Then, the homogenate was mixed with a vortexer and centrifuged at 16,000xg for 5 minutes at 4°C. After centrifugation, the upper aqueous phase was collected and 200 µl of chloroform was added. Samples were then vortexed and centrifuged at 16,000×g for 5 min at 4 °C. The

upper aqueous phase was collected after centrifugation then, 160 µl of 8M LiCl and 50 µl of 3M sodium acetate (pH 4.8) were added and mixed by gentle inversion. To precipitate RNA, samples were then incubated at -20℃ for 24 hours. After precipitation, the samples were centrifuged at 16,000×g for 5 min at 4 ℃. Following centrifugation, the pellet was kept and 500 µl of 2 M LiCl was added and mixed gently. Samples were then centrifuged at 16,000×g for 5 min at 4 ℃ and the supernatant was discarded. To clean the pellet, 500 µl of pre-chilled 80 % ethanol was added and mixed by inversions followed by centrifugation at 16,000×g for 5 min at 4 ℃. The pellet was dried using a speed vac and re-suspended in 30 µl of DEPC-water. RNA was then stored at -80 ℃ for downstream applications.

RNA sequencing, alignment, and transcript counts

Total RNA obtained was sent to the Kansas State University Integrated Genomics Facility (IGF) for library preparation and sequencing. The 36 samples were pooled into three libraries and sequenced independently. Single end-paired reads were obtained from NextSeq 500 with the NextSeq 500/550 High Output v2.5 Kit (75-nucleotide single-end reads) (Illumina, US). The obtained reads were aligned to the sorghum reference genome v3.1.1 using the splice-aware aligner GMAP/GSNAP program version "2021-12-17" (Wu et al., 2016). The aligned transcripts were then counted using HTSeq-counts (Anders et al., 2014). Genes with transcript count numbers of 10 or more were kept for subsequent analysis.

A priori candidate genes

Given the extensive knowledge of carotenoid biosynthesis, analysis was focused on a priori candidate genes. Transcript information for genes involved in the carotenoid biosynthesis pathway, carotenoid degradation, and the carotenoid precursor MEP pathway were obtained from Phytozome (http://www.phytozome.net) and Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg) (Supplemental Table 1). Several candidate genes had more than one transcript expressed, therefore, we use 'genes' when referring to all the transcripts, whereas we use 'transcripts' when referring to only the specified transcript. This is particularly true for the **β**-carotene 3-hydroxylase gene. We chose to use the notation β-OH, but names used in other papers include β-OH, crtRB1, BCH, and hyd3.

Clustering of samples

Transcript counts were transformed with the "varianceStabilizingTransformation" and transformed data was obtained with the "getVarianceStabilizedData" function of the R package DESeq2 (Love et al., 2014), using developmental time points as the design variable. The transformed counts were transposed and a Euclidean distance matrix was calculated using the 'dist' R base function. Average hierarchical clustering was determined with the 'hclust' function in R. Relationships of sample clustering and total carotenoid concentration in mature grains for each sample were visualized with the 'plotDendoAndColors' function of the WGCNA R package (Langfelder & Horvath, 2008) .

Differentially expressed genes

Genes differentially expressed were determined using the DESeq2 package (Love et al., 2014) in R. Genes differentially expressed across DAP were identified using the likelihood ratio test (LTR), with false discovery rate (FDR) < 0.05. Differentially expressed genes for DAP were then separated according to the biochemical pathways, and the patterns of expressions across the DAP were analyzed with the 'degPatterns' function. The 'Wald' test was used to identify genes differentially expressed in the carotenoid concentrations groups ('High vs. Low'). Normalized transcript counts were obtained for differentially expressed genes in the high vs low carotenoid content groups and were aggregated by the carotenoid concentrations group for comparisons.

Correlations between candidate genes and carotenoid concentrations

Correlations between lutein, zeaxanthin, and β-carotene were determined for each of the candidate genes in the carotenoid biosynthesis, carotenoid degradation, and the carotenoid precursor MEP pathway. Average carotenoid concentrations (µg/g) of one biological rep from previous studies in 2015 (Cruet-Burgos et al., 2020) and 2019 (Cruet-Burgos et al., 2022) were used (Supplemental Table 6). Transcript counts for 42 DAP were transformed with the 'vst' transformation and averaged across replicates of the four accessions—PI510924, PI329435, PI585347, and PI585348. Pearson's correlations between each carotenoid concentration and the transformed transcript counts for each gene were calculated in R. Candidate genes with P-value < 0.05 were found to be significantly correlated.

RESULTS

A priori candidate genes expressed in sorghum grain

To better understand the biosynthesis of carotenoids in sorghum grain, a priori candidate genes and their Phytozome-predicted (http://www.phytozome.net) transcripts were analyzed in order to test the hypothesis that known carotenoid-related genes are expressed in sorghum grain. The genes were categorized into three pathways: genes in the carotenoid precursor MEP pathway, genes in the carotenoid biosynthesis pathway, and genes in the carotenoid degradation pathways (Supplemental Table 1, Table 4.1). The carotenoid degradation pathways include genes involved in the biosynthesis of ABA and other apocarotenoids.

In the MEP pathway, all the a priori candidate genes were expressed in the grain (Table 4.1). For the carotenoid biosynthesis pathway, 19 out of the 20 genes identified as a priori candidates were expressed in sorghum grain (Table 4.1). Sobic.010G276400, encoding a phytoene synthase (PSY), was not expressed; however, homologs of this PSY (Sobic.002G292600 and Sobic.008G180800) were expressed in the grain.

For the carotenoid degradation pathways, the a priori candidate gene list consisted of seven genes encoding carotenoid cleavage dioxygenases (CCD's), involved in the conversion of **β**-carotene to apocarotenoids, and ten genes encoding 9-cis-epoxycarotenoid dioxygenases (NCED's) or abscisic aldehyde oxidase (AAO), involved in the degradation of carotenoids to ABA. Five out of the eight CCD's were expressed in the grain. The three CCD genes that were not expressed were Sobic.005G105700, Sobic.007G170300, and Sobic.010G050300. The ten genes encoding NCDE's and AAO, were all expressed in the sorghum grain.

Differential transcript expression within some genes with multiple transcripts was observed. There were 3 out of the 22 MEP pathway genes, 6 out of the 20 biosynthesis genes, and 2 out of the 17 degradation genes that had had some transcript variants that were expressed and some variants that were not expressed. Differential transcript expression could underlie differences in carotenoid concentrations, however it is also possible that the transcript variants that were not expressed are not functionally important.

Table 4.1 A priori candidate genes expressed in sorghum grain

	Biochemical Pathway		
A priori candidate genes	Carotenoid Precursors (MEP Pathway)	Carotenoid Biosynthesis	Carotenoid Degradation
Expressed	22	19	14
Not Expressed	0	1	3
Total	22	20	17

Factors controlling transcriptome variance

Next, to generate hypotheses on patterns of expression, we explored the intergroup and intragroup (developmental time, high vs low carotenoid lines, individual genotypes) transcriptomic variance for the samples. A hierarchical clustering analysis was conducted for the euclidean distance of the variance stabilizing transformation ‘vst’ counts (Figure 4.1). Developmental time, defined as 14, 28, or 42 days after pollination (DAP), was the factor that explained the majority of the transcriptomic variation, with

perfect clustering of samples according to the DAP (Figure 4.1). Within developmental time clusters, samples were grouped first by the carotenoid content group (High vs. Low), and then by individual genotypes (Figure 4.1).

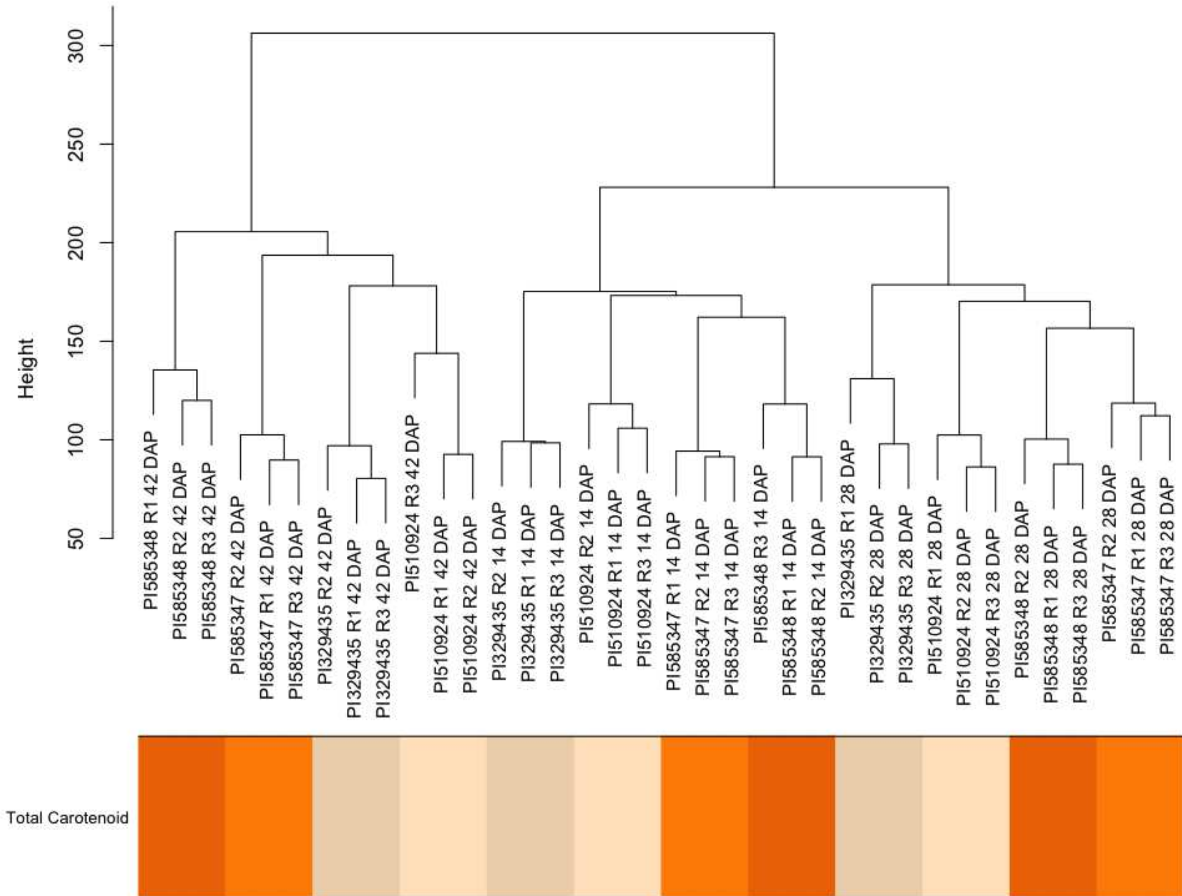


Figure 4.1 Samples hierarchical clustering and heat map of total carotenoid concentration at maturity for each accession. Darker colors in the heat map represent higher carotenoid concentration. R1-R3 denotes each biological replicate.

Differentially expressed genes through grain development

Since developmental time point was the factor that explained the majority of the transcriptome variance, we next sought to test the hypothesis that carotenoid content differences through grain development are driven by differences in gene expression. A likelihood ratio test (LTR) was performed to identify genes differentially expressed across grain development time points. A total of 21,192 transcripts were differentially expressed (False Discovery Rate (FDR) < 0.05) across grain development time points. Among them, 13 genes in the carotenoid precursor MEP pathway, 15 genes in the carotenoid biosynthesis pathway, and 13 genes in the carotenoid degradation pathway were differentially expressed across developmental time points.

Most of the expressed genes in the MEP pathway were differentially expressed between grain development time points. The initial steps in the MEP pathway, up until the synthesis of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) (Figure 4.2A), were more highly expressed during early stages of grain development (Figure 4.2B). Interestingly, genes encoding the enzymes involved in the branching points for the production of isopentyl pyrophosphate (IPP) or dimethyl pyrophosphate (DMAPP) clustered in the opposite pattern with higher expression at the end of grain development (Figure 4.2C). Aside from the genes involved in the branching points, only two other genes, Sobic.003G270500, encoding a farnesyl pyrophosphate synthase (FPPS), and Sobic.010G229400, encoding a geranyl diphosphate synthase (GPPS), were also in this cluster at the end of grain development (Figure 4.2C). However, another *FPPS* (Sobic.009G216800) was highly expressed at initial stages of development (Figure 4.2B), suggesting complementary function.

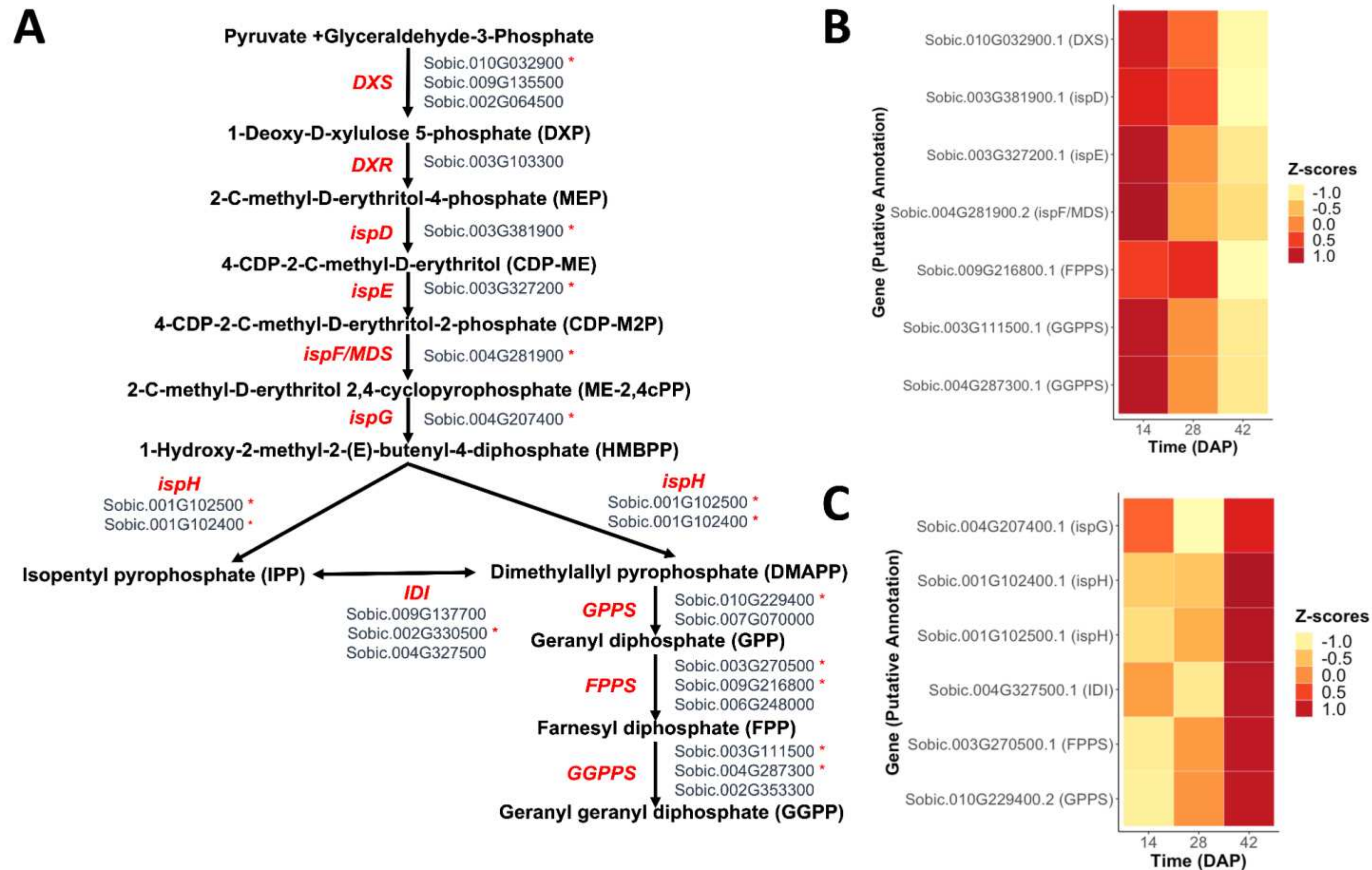


Figure 4.2 Precursor pathway: differentially expressed genes and patterns of expression across grain developmental time points (DAP).
A) MEP pathway with the putative genes catalyzing each reaction. B) Genes expressed in a higher proportion at the beginning of grain development. C) Genes expressed in higher proportion at the end of grain development. Asterisks next to genes, represent the genes that were differentially expressed across times. Positive Z-scores indicate higher expression compared to baseline.

As expected since previous studies show an increase in carotenoid concentration in early development, the expression of most carotenoid biosynthesis genes (Figure 4.3A) was higher during early stages of grain development (Figure 4.3B). Genes involved in the later steps of carotenoid biosynthesis, such as **β**-carotene 3-hydroxylase (β-OH), zeaxanthin epoxidase (ZEP), and violaxanthin de-epoxidase (VDE), had the opposite trend with higher expression during later stages of grain development. Interestingly, there seemed to be complementation of expression for homologous genes encoding a carotenoid isomerase (CRTISO), phytoene desaturase (PDS) and VDE. For CRTISO, the Sobic.005G160500 homolog was more highly expressed during the early stages of grain development (Figure 4.3B), whereas the Sobic.001G01800 homolog was more highly expressed during later stages of development (Figure 4.3C). Similarly, the PDS homolog Sobic.006G177400 (Figure 4.3B) was more highly expressed at 14 DAP, whereas the PDS homolog Sobic.002G383400 (Figure 4.3C) was more highly expressed at 42 DAP. VDE homolog Sobic.006G049200 was more highly expressed during early stages of grain development (Figure 4.3B), while Sobic.003G277400 was more highly expressed during later stages (Figure 4.3C).

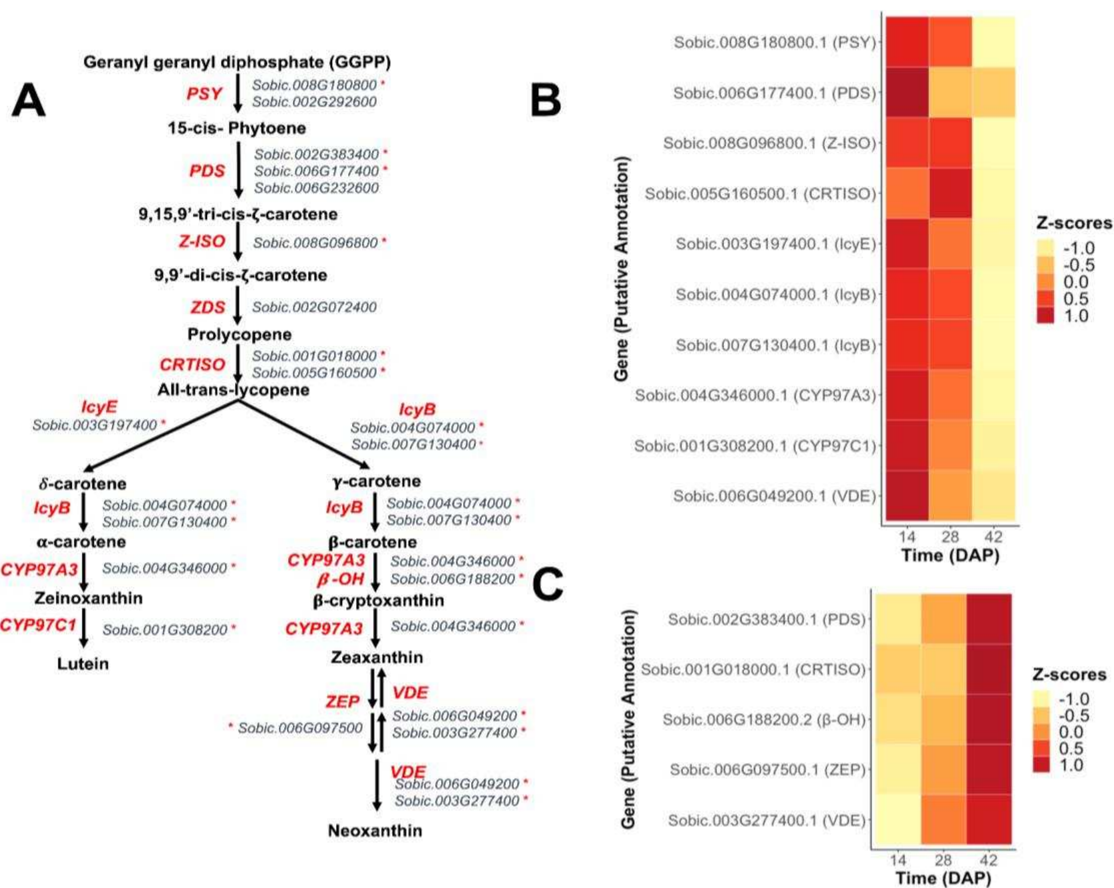


Figure 4.3 Biosynthesis pathway: differentially expressed genes and patterns of expression across grain developmental time points (DAP).

A) Carotenoid biosynthesis pathway with the putative genes catalyzing each reaction. B) Genes expressed in a higher proportion at the beginning of grain development (14 DAP). C) Genes expressed in higher proportion at the end of grain development (42 DAP). Asterisks next to genes represent the genes that were differentially expressed across times. Positive Z-scores indicate higher expression compared to baseline.

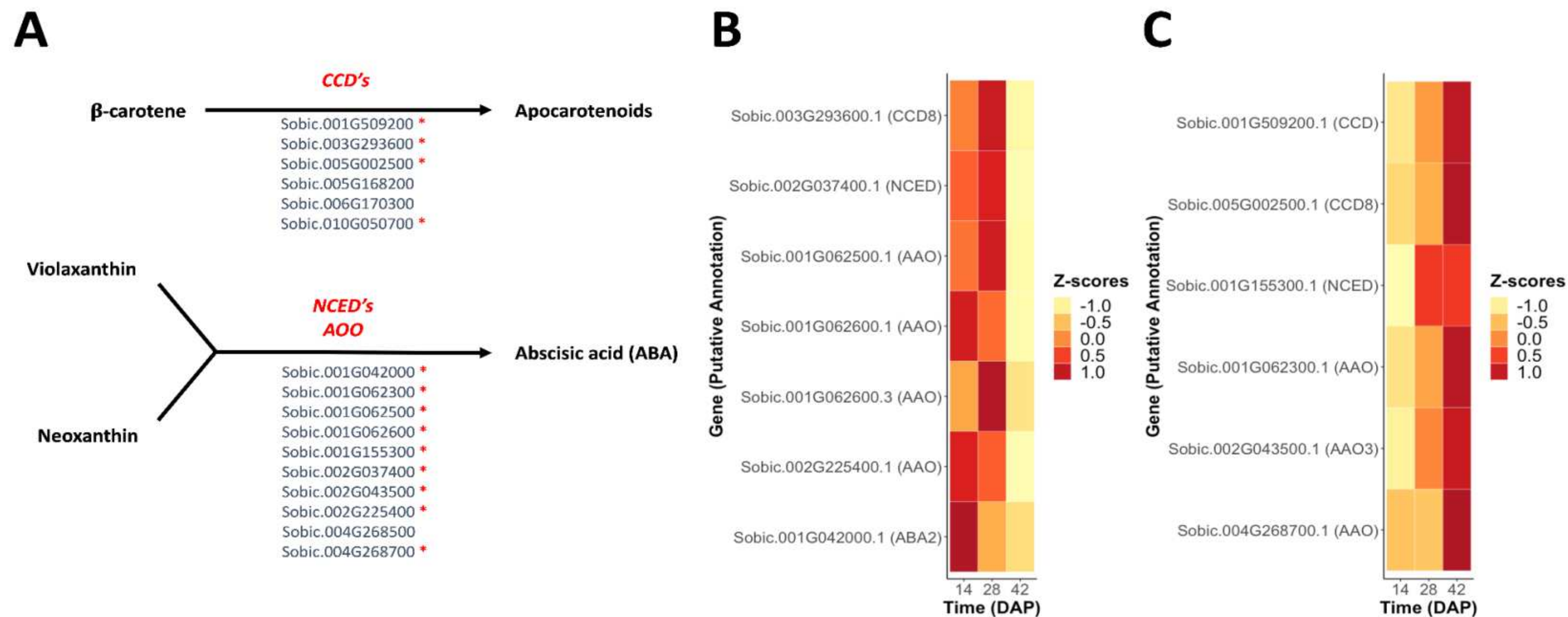


Figure 4.4 Degradation pathways: differentially expressed genes and patterns of expression across grain developmental time points (DAP).

A) Carotenoid degradation pathways with the putative genes catalyzing each reaction. B) Genes expressed in a higher proportion at the beginning of grain development (14 DAP), C) Genes expressed in higher proportion at the end of grain development (42 DAP). Asterisks next to genes represent the genes that were differentially expressed across sorghum grain development time points. Positive Z-scores indicate higher expression compared to baseline.

Carotenoid degradation genes (Figure 4.4A) differentially expressed through grain development were also grouped into two clusters: highly expressed at early stages of grain development (Figure 4.4B) and highly expressed at later stages of grain development (Figure 4.4C). Genes encoding the three main enzyme types involved in carotenoid degradation—CCD's, NCED's, and AAO's—were distributed in both clusters (Fig. 4.4C).

Differentially expressed genes between high vs low carotenoid content groups

Next, since the carotenoid content group (high vs low) was the factor that explained the majority of transcriptome variance after developmental time point, we sought to test the hypothesis that carotenoid content differences between high and low carotenoid lines are driven by differences in gene expression. A 'wald' test was conducted to identify genes differentially expressed between the high and low carotenoid accessions for each of the developmental time points and the 'vst' transformed transcript counts were compared for differentially expressed genes between accessions.

A total of 2,587 genes were differentially expressed between high and low carotenoid content lines at 14 DAP. However, among these genes none were *a priori* candidates in the carotenoid biosynthesis pathway, and only four of them were *a priori* candidate genes involved in either the carotenoid degradation or the carotenoid precursor MEP pathway (Figure 4.5). In the MEP pathway, Sobic.004G287300 (*GGPPS*) and Sobic.009G137700 (*IDI*) were differentially expressed between high and low carotenoid lines. The *GGPPS* was more highly expressed in the high carotenoid content group, whereas the *IDI* was more highly expressed in the low carotenoid group (Figure 4.5A). These results could suggest that there might be a feed-forward regulation in the MEP

pathway resulting in higher carotenoid content. In the carotenoid degradation pathways, Sobic.004G268500 (*NCED4/CCD4*) and Sobic.002G225400 (*AAO*) were differentially expressed between high and low carotenoid lines. However, the expression of these two genes was relatively low. The *NCED* was more highly expressed in the low carotenoid group, whereas the *AAO* was more highly expressed in the high carotenoid group (Figure 4.5B).

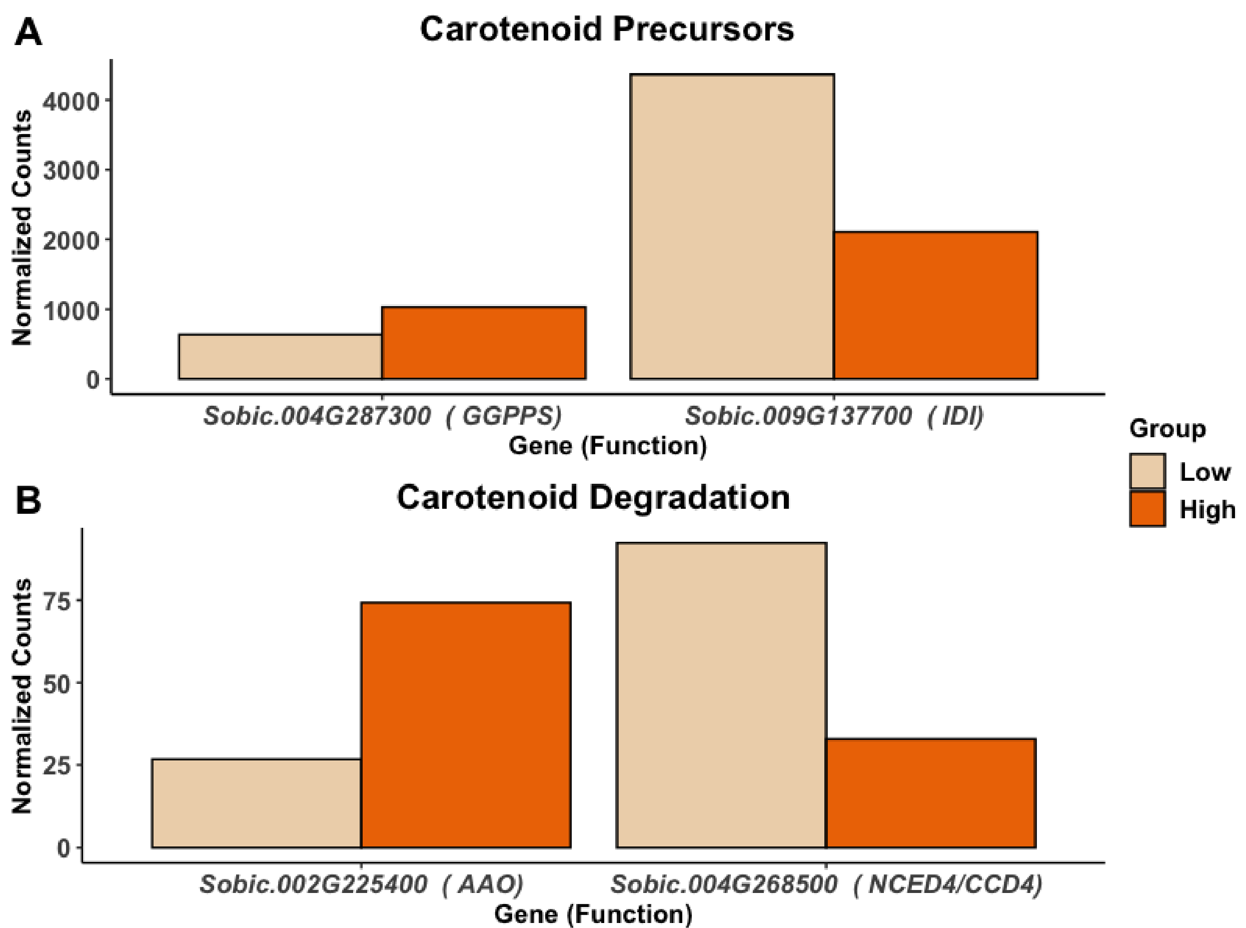


Figure 4.5 Normalized counts for differentially expressed genes between high and low carotenoid accessions at 14 DAP. A) carotenoid precursors (MEP pathway); B) carotenoid degradation pathways

The highest number of total differentially expressed genes for the carotenoid content group (high vs low) was found at 28 DAP with 4,919 genes. Eight genes were *a priori* candidate genes with two of them in the carotenoid degradation pathways and the remaining six in the MEP pathway (Figure 4.6). Genes differentially expressed in the MEP pathway for the high vs low carotenoid content groups had different patterns (Figure 4.6). Of the two GGPPS genes that were differentially expressed, Sobic.003G111500 had higher transcript counts in the low carotenoid group compared to the high carotenoid group, but it had a very low transcript count overall, which could suggest it is not the main GGPPS enzyme functioning in the grain (Figure 4.6A). Two genes, Sobic.010G032900 (*deoxyxylulose 5-phosphate synthase*, *DXS*) and Sobic.009G137700 (*IDI*) had higher expression in the low carotenoid content group. The *IDI* gene was also identified as differentially expressed among the carotenoid content group for 14 days and had the same pattern of expression with higher expression in the lower carotenoid content group. The remaining three genes, Sobic.003G327200 (*ispE*), Sobic.009G216800 (*FPPS*), and Sobic.004G287300 (*GGPPS*), were more highly expressed in the high carotenoid lines at 28 DAP (Figure 4.6A), suggesting they could be a feed-forward mechanism at this development stage. For carotenoid degradation, an *NCED* (Sobic.002G037400) and an *ABA2* (Sobic.001G04200) were differentially expressed (Figure 4.6B). Similar to 14 DAP, these carotenoid degradation genes had opposite patterns of expression with the *NCED* more highly expressed in the low carotenoid group.

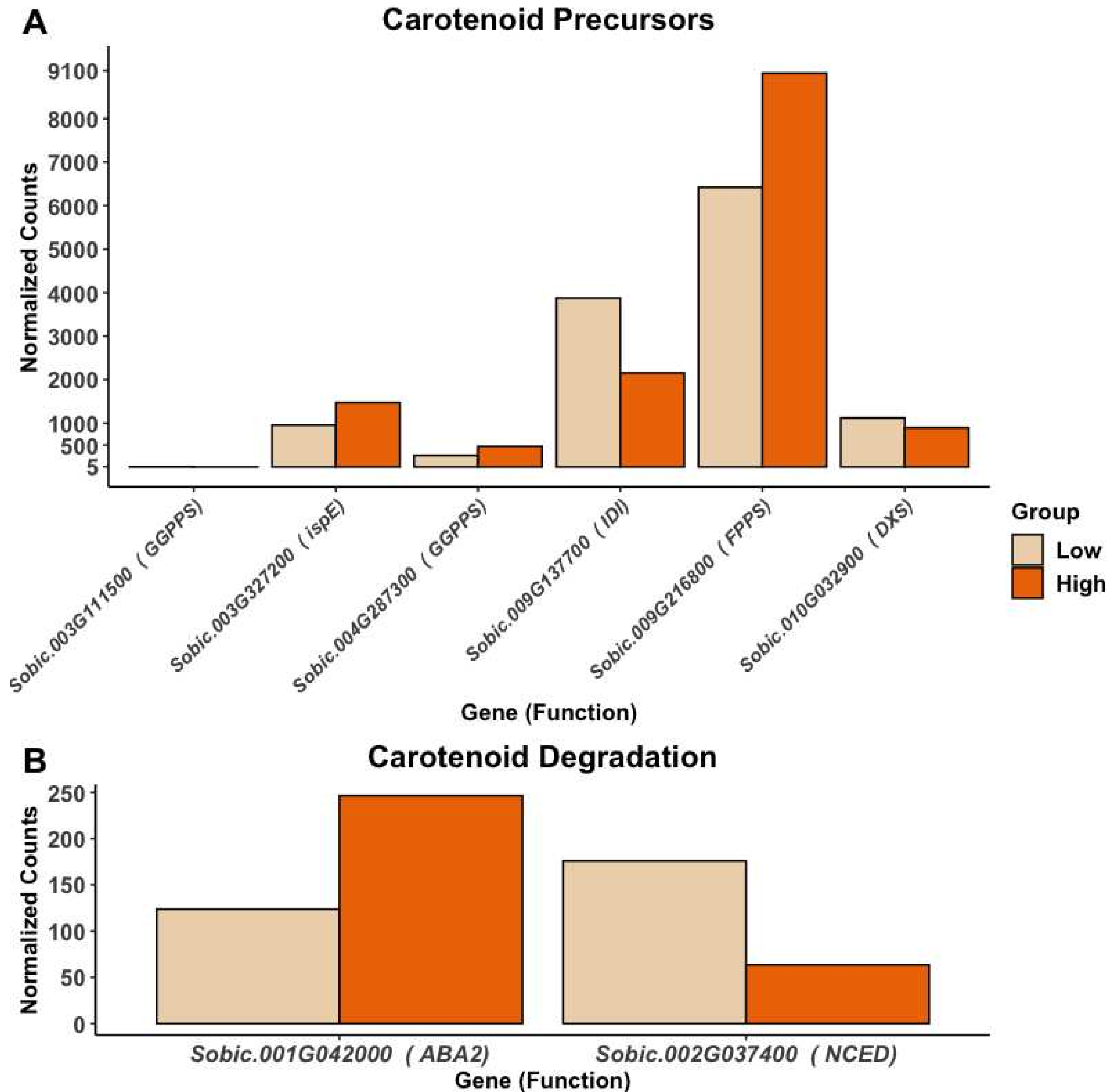


Figure 4.6 Normalized counts for differentially expressed genes between high and low carotenoid accessions at 28 DAP. A) carotenoid precursors (MEP pathway); B) carotenoid degradation pathways

At 42 DAP, there were 4,294 genes differentially expressed among the carotenoid content group, which included the highest number of *a priori* candidate genes (Figure 4.7). Compared to the carotenoid biosynthesis and degradation pathways at this time point, the precursor MEP pathway had the highest number of genes differentially

expressed between high and low carotenoid lines (Figure 4.7A). Among them, three genes were differentially expressed with the same pattern of expression at previous time points: Sobic.004G287300 (*GGPPS*) and Sobic.009G137700 (*IDI*) at both 14 DAP and 28 DAP, and Sobic.003G327200 (*ispE*) at 14 days.

Four additional genes in the MEP pathway were differentially expressed between high and low carotenoid lines at 42 DAP that were not differentially expressed at previous time points. Sobic.004G281900 (2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase, *ispF/MDS*) and Sobic.003G381900 (*ispD*) were more highly expressed in the high carotenoid group, while Sobic.004G207400 (*1-Hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase*, *ispG*) and Sobic.002G064500 (*DXS*) were expressed slightly more in low carotenoid lines (Figure 4.7A). Surprisingly, 42 DAP was the only developmental stage in which carotenoid biosynthesis genes—Sobic.002G292600 (*PSY*) and Sobic.002G383400 (*PDS*)— were differentially expressed between the high versus low carotenoid lines (Figure 4.7B). Both *PSY* and *PDS* were more highly expressed in the high carotenoid lines. The *PSY* expression was more similar between the accessions, with the exception of PI585348 (Figure 4.8A), while *PDS* had bigger differences of expressions (Figure 4.8B). Among the carotenoid degradation genes, Sobic.002G225400 (*AAO*) was more highly expressed in the high carotenoid group (Figure 4.7C), as it was at 14 DAP (Figure 4.5B).

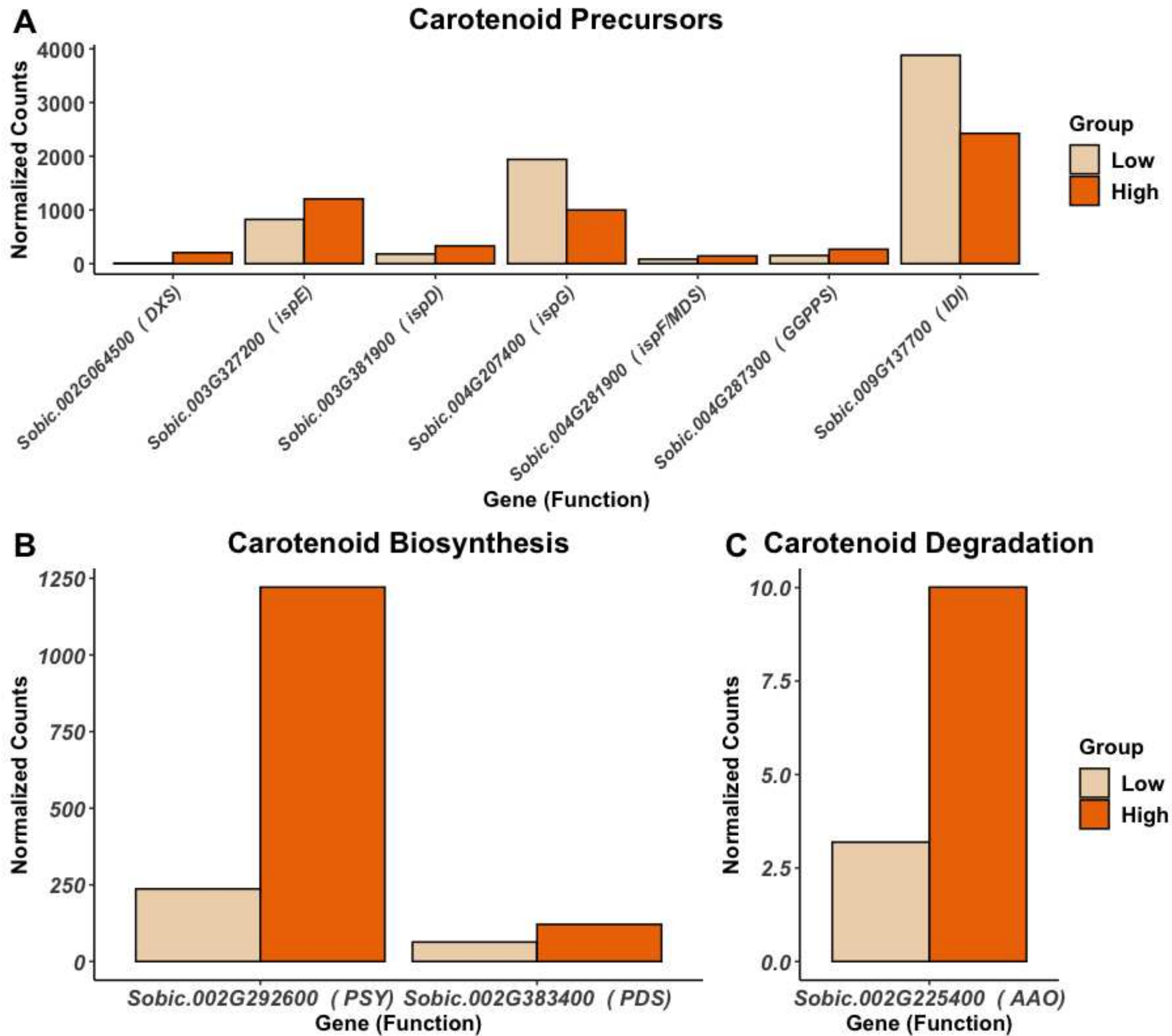


Figure 4.7 Normalized counts for differentially expressed genes between high and low carotenoid accessions at 42 DAP. A) carotenoid precursors (MEP pathway); B) carotenoid degradation pathways

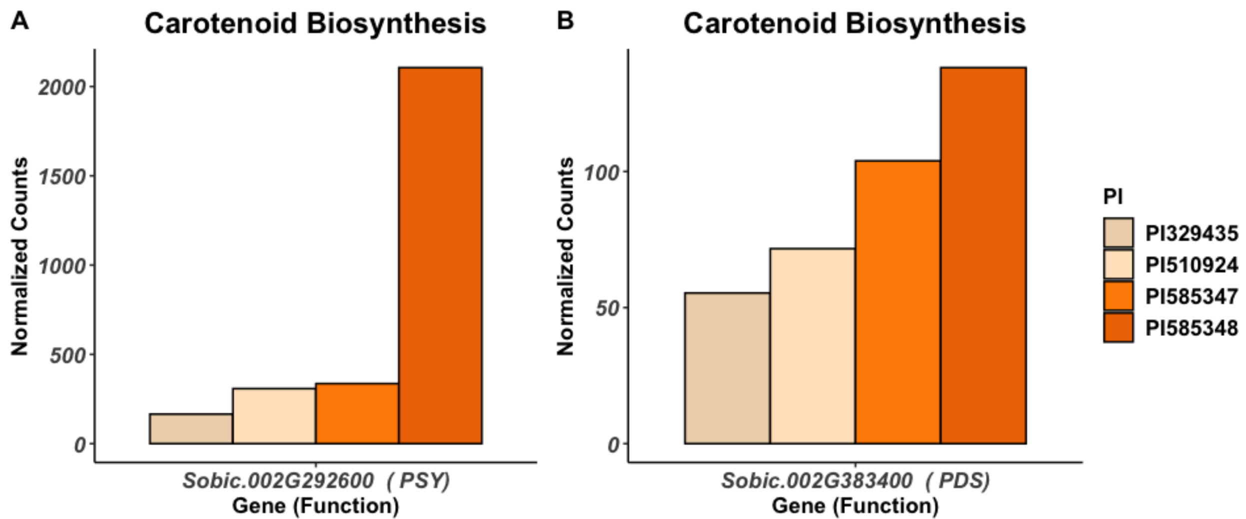


Figure 4.8. Genotypes normalized transcript counts for carotenoid biosynthesis genes differentially expressed at 42 DAP in high versus low carotenoid sorghum accessions. A) Phytoene Synthase (PSY); B) Phytoene desaturase (PDS)

Correlations between candidate gene expression and carotenoid concentration

Since we found differential expression of several *a priori* candidate genes across time points and between high vs low carotenoid content groups, we wanted to further test the hypothesis that gene expression differences underlie carotenoid variation. To look for evidence that expression of individual *a priori* candidate genes underlie differences in individual carotenoid compounds, correlations were calculated between transcript counts and the concentrations of lutein, zeaxanthin, and β-carotene at grain maturity. Among the 55 candidate genes expressed in the grain, only 12 of them were significantly correlated ($P < 0.05$) with concentrations of at least one of the carotenoid compounds (Table 4.2). Expression of three genes—Sobic.003G381900 (*CDP-ME transferase*, *ispD*), Sobic.001G509200 (*CCD*)*,* and Sobic.006G170300 (*CCD*)—were correlated with all three carotenoids (Table 4.2). As expected due to their biosynthesis function in the carotenoid precursor MEP pathways, expression of *ispD* was positively correlated with

concentrations of all three carotenoids. Also as expected, due to the role of CCDs in carotenoid degradation, the expression of one CCD (Sobic.001G509200) was negatively correlated with concentrations of all three carotenoids. Unexpectedly, however, expression of another CCD (Sobic.006G170300) was positively correlated with concentrations of all three carotenoids.

The expression of six genes—Sobic.009G137700 (*IPP isomerase*, *IDI*), Sobic.004G287300 (*geranyl geranyl diphosphate synthase*, *GGPPS*), Sobic.002G383400 (*PDS*), Sobic.006G232600 (*PDS*), Sobic.001G062600 (*AAO*), and Sobic.001G155300 (*NCED*)—were correlated with β-carotene and lutein concentrations, but not with zeaxanthin concentrations (Table 4.2). Of the MEP pathway genes, *IDI* expression was negatively correlated and *GGPPS* expression was positively correlated with β-carotene and lutein concentrations. Of the carotenoid biosynthesis genes, expression of one of the *PDS* genes (Sobic.002G383400) was positively correlated with β-carotene and lutein concentrations. However, expression of the other *PDS* gene (Sobic.006G232600) was negatively correlated with β-carotene and lutein concentrations. Of the carotenoid degradation genes, *AAO* (Sobic.001G062600) expression was positively correlated and *NCED* (Sobic.001G155300) expression was negatively correlated with β-carotene and lutein concentrations (Table 4.2).

Expression of three genes—Sobic.003G327200 (*CDP-ME kinase*, *ispE*), Sobic.006G177400 *(PDS)*, and Sobic.002G043500 *(AAO3)*—were correlated only with zeaxanthin concentrations. In the MEP pathway, expression of *ispE* was positively correlated with zeaxanthin concentrations. In the carotenoid biosynthesis and

degradation pathways, expression of *PDS* and *AAO3* were negatively correlated with zeaxanthin concentrations, respectively (Table 4.2).

Table 4.2 Correlations between *a priori* candidate gene expression and concentrations of lutein, zeaxanthin, and β-carotene at 42 DAP.

Transcript	Enzyme	Compound	r	p-val	Pathway
Sobic.003G381900.1	ispD	Lutein	0.960	0.040	Carotenoid precursor (MEP)
		Zeaxanthin	0.998	0.002	Carotenoid precursor (MEP)
		β-carotene	0.956	0.044	Carotenoid precursor (MEP)
Sobic.001G509200.1	CCD	Lutein	-0.990	0.010	Carotenoid degradation
		Zeaxanthin	-0.953	0.047	Carotenoid degradation
		β-carotene	-0.990	0.010	Carotenoid degradation
Sobic.006G170300.1	CCD	Lutein	0.968	0.032	Carotenoid degradation
		Zeaxanthin	0.996	0.004	Carotenoid degradation
		β-carotene	0.966	0.034	Carotenoid degradation
Sobic.009G137700.1	IDI	Lutein	-0.965	0.035	Carotenoid precursor (MEP)
		β-carotene	-0.966	0.034	Carotenoid precursor (MEP)
Sobic.004G287300.1	GGPPS	Lutein	0.951	0.049	Carotenoid precursor (MEP)
		β-carotene	0.952	0.048	Carotenoid precursor (MEP)
Sobic.002G383400.1	PDS	Lutein	0.971	0.029	Carotenoid biosynthesis
		β-carotene	0.973	0.027	Carotenoid biosynthesis
Sobic.006G232600.1	PDS	Lutein	-0.980	0.020	Carotenoid biosynthesis
		β-carotene	-0.981	0.019	Carotenoid biosynthesis
Sobic.001G062600.3	AAO	Lutein	0.969	0.031	Carotenoid degradation
		β-carotene	0.966	0.034	Carotenoid degradation
Sobic.001G155300.1	NCED	Lutein	-0.976	0.024	Carotenoid degradation
		β-carotene	-0.975	0.025	Carotenoid degradation
Sobic.003G327200.1	ispE	Zeaxanthin	0.991	0.009	Carotenoid precursor (MEP)
Sobic.006G177400.1	PDS	Zeaxanthin	-0.959	0.041	Carotenoid biosynthesis
Sobic.002G043500.1	AAO3	Zeaxanthin	-0.979	0.021	Carotenoid Degradation

DISCUSSION

Sorghum is an important staple crop in regions where vitamin A deficiency is prevalent. Although pro-vitamin A carotenoids do accumulate in sorghum grain, concentrations are low. Breeding has the potential to increase carotenoid concentrations further, but there are current gaps in knowledge of the biosynthesis and regulation of carotenoids in sorghum grain. Identifying genes that control carotenoid accumulation through grain development can help identify potential targets for biofortification breeding. Given the pattern of accumulation of carotenoids in sorghum grain through development and the large proportion of the transcriptional variation for carotenoid-related genes explained by developmental stage (Figure 4.1), there seems to be tight transcriptional controls throughout grain development. This is a promising result for biofortification breeding because it suggests that allelic variants that modify the expression patterns of the genes involved in this transcriptional switch through grain development can be used to further accumulate carotenoids in sorghum grain. For example, in maize the *crtRB1* gene, which encodes a β-carotene hydroxylase (Harjes et al., 2008; Yan et al., 2010), is differentially expressed through grain development and its expression is correlated with carotenoid content (Vallabhaneni et al., 2009). Characterization of this gene showed that there are allelic variants that result in significant differences in gene expression as well as in pro-vitamin A carotenoid accumulation (Yan et al., 2010). Marker-assisted selection targeting the high-carotenoid allele of the *crtRB1* gene has been employed in several studies, resulting in significant increases of β-carotene concentrations (Babu et al., 2013; Chandran et al., 2019; Goswami et al., 2019; Liu et al., 2015; Muthusamy et al., 2014). In this study, we identified genes in the carotenoid precursor MEP pathway, carotenoid

biosynthesis pathway, and carotenoid degradation pathways that have the potential to be impactful for biofortification efforts if favorable allelic variants are found.

Carotenoid precursor pathway targets for biofortification

The MEP pathway synthesizes the GGPP precursor needed for carotenoid biosynthesis, thus the pathway's influence on carotenoid variation has been studied. Several other important plant compounds, such as chlorophylls, tocochromanols, and gibberellins, also use GGPP as a precursor, so there may be competition between the pathways for use of this substrate. Identifying rate limiting steps in the MEP pathway can help to increase carotenoid concentrations in sorghum grain by increasing precursor substrates. In this study, most of the genes that were differentially expressed among the high vs low carotenoid content groups were in the MEP pathway, suggesting that it is a main controller of carotenoids in sorghum grain, and is a potential target for biofortification efforts. A promising gene target is Sobic.003G381900 (*ispD*), because it was differentially expressed across developmental time points (Figure 4.2B), positively associated with all carotenoid concentrations (Table 4.2), and more highly expressed in high carotenoid lines at 42 DAP (Figure 4.7A). These results suggest that high expression of *ispD* increases carotenoid concentrations, perhaps through a feed-forward mechanism.

Sobic.004G281900 (ispF/MDS) is another potential target for biofortification breeding. Although it was not correlated with carotenoid concentrations at 42 DAP, it was more highly expressed in the high carotenoid group at 42 DAP. Notably, we previously identified this candidate gene in a GWAS for sorghum grain zeaxanthin, in which a significant marker-trait association was identified only 62 Mb away from the ispF/MDS gene (Cruet-Burgos et al., 2020). In the same study, we also identified ispF/MDS in a

biosynthesis-pathway targeted GWAS, which we conducted to control for possible false negatives by using only SNPs near a priori candidate genes. It was one of only two gene candidates identified in the pathway targeted GWAS.

Another promising MEP pathway gene target for biofortification breeding is Sobic.009G137700 (IDI). It was negatively associated with carotenoid concentration for lutein and β-carotene (Table 4.2) and had significantly higher expression in low carotenoid lines at each developmental time point (Figure 4.5A, 4.6A and 4.7A). Interestingly, IDI is involved in a branch point in the MEP pathway, so it can be hypothesized that the low carotenoid lines have an allele that preferentially converts DMAPP to IPP, thus reducing the flux of the pathway towards carotenoid biosynthesis.

Lastly, Sobic.004G287300 (GGPPS) is also a promising target, as it was differentially expressed across developmental time points (Figure 4.2), associated with lutein and β-carotene (Table 4.2), and differentially expressed between high and low carotenoid groups for all developmental time points (Figure 4.5A, 4.6A and 4.7A). The expression pattern of this GGPPS gene is perhaps the most exciting in the MEP pathway, because it shows differential expression between high and low carotenoid content groups for all developmental time points. GGPPS is highly expressed in both high and low carotenoid groups at 14 DAP (Figure 4.2B), however, at 28 DAP and 42 DAP, expression is decreased for the low carotenoid group, but remains high in the high carotenoid group (Figures 4.6A, 4.7A). We hypothesize that high carotenoid lines have an allelic variant that remains highly expressed through late stages of grain development, thus increasing the flux to the carotenoid biosynthesis pathway. In maize, a positive correlation between carotenoid concentration and gene expression of GGPPS has been detected

(Vallabhaneni & Wurtzel, 2009), but the sorghum homolog—Sobic.004G287300—has not been associated with variation in carotenoid content through genomic studies. We previously identified a different sorghum GGPPS candidate (Sobic.002G353300) in an association study that identified a SNP associated with β-carotene variation (Cruet-Burgos et al., 2020, p.).

Carotenoid biosynthesis pathway targets for biofortification

Genes directly involved in the carotenoid biosynthesis pathway have been the major targets for biofortification efforts across species due to their potential to control specific carotenoids, such as β-carotene. In this study two carotenoid pathway genes were identified as promising targets for biofortification efforts. The first gene, Sobic.002G383400, encodes a PDS that was positively correlated with lutein and β-carotene concentrations (Table 4.2), was more highly expressed at late stages of grain development, and was more highly expressed in the high carotenoid lines versus the low carotenoid lines at 42 DAP. We hypothesize that the increase in PDS expression at late stages of development in the high carotenoid group increases carotenoid biosynthesis, resulting in higher carotenoid concentrations in the high carotenoid group. However, this gene seems to have a low transcript count number when compared to others (Figures 4.7B and 4.8B), therefore further characterization is required to understand the role of PDS in carotenoid variation.

The second promising gene is Sobic.002G292600 (PSY). This gene is an interesting candidate because, unlike the other candidates, it was not differentially expressed across time points (Figure 4.3), nor was it associated with carotenoid

concentrations (Table 4.2), but it was differentially expressed among high vs low carotenoid content groups at 42 DAP. Interestingly, our highest β-carotene content line, PI585348, —had a notably higher PSY expression compared with the other accessions at 42 DAP (Figure 4.8A). In a QTL mapping study in a biparental sorghum population, SNPs in proximity to this gene were found to be associated with lutein, zeaxanthin, and β-carotene concentrations (Fernandez et al., 2008), suggesting there are PSY allelic variants associated with variation in β-carotene concentration. PSY genes have also been associated with carotenoid concentrations in maize (Wong et al., 2004), as well as with differential expression through development and genotypes (Chander et al., 2008; Da Silva Messias et al., 2014; Z. Fu et al., 2013; Palaisa et al., 2003; Vallabhaneni et al., 2009). As the first committed step in carotenoid biosynthesis, increasing PSY activity could potentially increase carotenoid accumulation in sorghum grain. Sequencing this gene in PI585348 will help to identify potential allelic variants responsible for the high expression and high carotenoid content, which then could be incorporated into molecular breeding efforts.

It is worth noting that ZEP (Sobic.006G097500) was not differentially expressed in carotenoid content groups (High vs. Low) even though it was identified as a major GWAS candidate in a previous study (Cruet-Burgos et al., 2020). These results suggest that ZEP differential expression does not underlie the association between the ZEP region and carotenoid variation. One possibility is that the marker-trait association identified in our GWAS study is in linkage disequilibrium (LD) with ZEP, but that ZEP is not the functional variant. However, this is not likely as there is significant evidence in maize and arabidopsis that ZEP is one of the major controllers of grain carotenoids (Azmach et al.,

2018; Diepenbrock et al., 2021; Owens et al., 2014; Suwarno et al., 2015). We hypothesize that there are allelic variants in ZEP that affect enzymatic activity rather than gene expression. Sequencing of the ZEP gene in accessions with diverse carotenoid profiles is needed to test this hypothesis.

Carotenoid degradation pathway targets for biofortification

Degradation of carotenoids occurs throughout the biosynthesis pathway. Two main routes of degradation are degradation of β-carotene to apocarotenoids such as strigolactones, or degradation of violaxanthin and neoxanthin to abscisic acid (Figure 4.4). Three enzyme groups—CCDs, NCEDs, and AAOs—catalyze the degradation steps (Figure 4.4). Almost all of the *a priori* candidate genes were differentially expressed through grain development (Figures 4.4B and 4.4C). However, when high vs low carotenoid content groups were compared at 14, 28, and 42 DAP, the number of degradation genes differentially expressed at each time point was minimal. Based on these results, it can be hypothesized that degradation genes are not the main drivers of differences among high versus low carotenoid content groups in sorghum grain, and that the differential expression of these genes is to support other biochemical processes for which carotenoids are precursors.

However, there are two degradation genes, Sobic.001G509200 (*CCD*) and Sobic.001G155300 (*NCED*), that have potential for use in sorghum biofortification. Although they were not differentially expressed among the high vs low carotenoid content groups, they were differentially expressed through development, with both of them more highly expressed in later time points (Figure 4.4C). A QTL mapping study in sorghum

identified QTLs in proximity to both of these genes that were associated with lutein variation (Fernandez et al., 2008), suggesting there may be allelic variants of these two genes associated with higher carotenoid concentrations. Allelic variants with reduced activity in either of these genes could be used in molecular breeding to reduce degradation of carotenoids at later stages of grain development. However, despite their potential, carotenoid cleavage enzymes are known to have low substrate specificity (Auldridge et al., 2006; Ohmiya, 2009; Priya et al., 2019), thus they could have limited impact if complementary action of other degradation enzymes occurs.

We were also surprised to see positive correlations between carotenoid concentrations and Sobic.006G170300, annotated as a *CCD* (Table 4.2). CCD's are a family of enzymes known to degrade some carotenoids to apocarotenoids, which are a large group of compounds—including ABA and strigolactones—derived from carotenoids through oxidative cleavage. Therefore we were expecting a negative correlation between Sobic.006G170300 and carotenoid concentrations. In Arabidopsis and maize grains, negative correlations have been observed between *CCD* expression and carotenoid content (Da Silva Messias et al., 2014; Gonzalez-Jorge et al., 2014). However, despite the four CCD's in plants—CCD1, CCD4, CCD7, and CCD8—sharing specificity for the 9,10 double bonds on their substrates, they have shown to be tissue specific. *CCD7* and *CCD8* have tissue specificity for Arabidopsis roots (Gonzalez-Jorge et al., 2014), while *CCD1* and *CCD4* have been found to be expressed in fruits and flowers of several plants (Ohmiya, 2009). Based on sequence similarity with *Arabidopsis thaliana*, Sobic.006G170300 encodes the CCD7 (74.4% similarity to AT2G44990). Even though Sobic.006G170300 overcame our 10 count filtering threshold, it had a low expression

level with only 25 counts across our samples. Therefore it may be possible that the positive correlation here observed (Table 4.2) is a false positive due to low counts and that Sobic.006G170300 encodes a CCD7 with predominant root expression. To test this hypothesis, sorghum CCD's across tissues must be examined. Alternatively, we could hypothesize that the Sobic.006G170300 allele in the high carotenoid lines encodes an enzyme that has reduced carotenoid degradation activity. To test this hypothesis, the gene can be sequenced in both the high and low carotenoid lines for comparison, and/or enzyme activity can be measured and compared.

Duplicate genes and their potential for biofortification

Interestingly, we identified several duplicate genes that appear to be complementary to each other, with one duplicate expressed only in early grain development and the other expressed only in late grain development. For example, PDS genes Sobic.006G177400 and Sobic.002G383400 were more highly expressed at early stages of development and at later stages of development, respectively. We hypothesize that this expression divergence is due to subfunctionalization (the expression patterns of the ancestral gene is divided between the duplicates, so the ancestral expression pattern is not retained) or neofunctionalization (one duplicate gains a new expression pattern and the other duplicate retains the ancestral expression pattern) (Panchy et al., 2016). The type of expression divergence in duplicate genes has implications for biofortification breeding, because the way in which duplicate genes function together will inform the choice in breeding strategies. For example, if increased expression of a duplicate gene is needed to increase carotenoid concentrations, and the duplicate genes have redundant

function, then only one gene may need to be manipulated for biofortification. In contrast, if decreased expression of a gene is needed to increase carotenoid concentrations, then both genes may need to be manipulated for effective biofortification. However, if the duplicate genes do not have redundant function, then manipulating any one of them for biofortification may result in unintended effects due to pleiotropy. Gene complementation tests in recombinant inbred lines (RILs) could be used to test the relationships between these duplicated genes.

Carotenoid regulation in sorghum

Genomic studies suggest that variation in sorghum grain carotenoids are due to an oligogenic genetic architecture, with marker-trait associations detected near a priori candidate genes within the carotenoid precursor MEP pathways, the carotenoid biosynthesis pathway, and the carotenoid degradation pathways (Cruet-Burgos et al., 2020; Fernandez et al., 2008). However, only one of the genes (ispF/MDS) co-localizing with these associations were here identified as differentially expressed in high vs low carotenoid content groups. Comparing these results with those of maize (Babu et al., 2013; Da Silva Messias et al., 2014, 2014; Diepenbrock et al., 2021; Z. Fu et al., 2013; Vallabhaneni & Wurtzel, 2009), it seems that the controls for sorghum carotenoid content are more complex. One hypothesis is that the genetic variation underlying carotenoid variation in sorghum is not causing differences at the transcriptional level. Alternatively, given the interconnectivity of carotenoid biosynthesis with other biological processes, carotenoid content in sorghum grain could have a more complicated inheritance and control; however, more studies are needed. Another hypothesis is that at the gene

expression level, these genes do not independently have large expression differences between high versus low carotenoid content groups, but together an additive expression results in higher carotenoid concentrations, suggesting a polygenic component.

CONCLUSION

To gain insights into the mechanisms underlying sorghum carotenoid biosynthesis and variation, we used a functional genomics approach, analyzing the transcriptomes in high and low carotenoid sorghum varieties throughout grain development. We have developed a clearer understanding of carotenoid regulation in sorghum grain, finding that 1) early MEP and carotenoid biosynthesis pathway genes are more highly expressed during early grain development, whereas later MEP and carotenoid biosynthesis pathway genes are more highly expressed during late grain development; 2) there is differential expression between high and low carotenoid lines of predominantly MEP and carotenoid degradation pathway genes during early grain development, whereas there is differential expression of genes in all three pathways during late grain development; and 3) controls for sorghum grain carotenoids are likely organized across a metabolic network, interacting with pathways involved in synthesizing other compounds during grain fill. Additionally, we identified potential gene targets for biofortification breeding, particularly *GGPPS*, *PDS*, and *PSY*. Moving forward, QTL analysis, sequence analysis, and marker testing are needed to identify causal allelic variants underlying carotenoid variation in sorghum. This transcriptomics study contributes to efforts to develop molecular breeding tools for sorghum carotenoid biofortification.

LITERATURE CITED

Anders, S., Pyl, P. T., & Huber, W. (2014). *HTSeq – A Python framework to work with high-throughput sequencing data* (p. 002824). bioRxiv. https://doi.org/10.1101/002824

Armstrong, G. A., Alberti, M., & Hearst, J. E. (1990). Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. *Proceedings of the National Academy of Sciences*, *87*(24), 9975–9979. https://doi.org/10.1073/pnas.87.24.9975

Armstrong, G. A., & Hearst, J. E. (1996). Genetics and molecular biology of carotenoid pigment biosynthesis. *The FASEB Journal*, *10*(2), 228–237. https://doi.org/10.1096/fasebj.10.2.8641556

Auldridge, M. E., McCarty, D. R., & Klee, H. J. (2006). Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Current Opinion in Plant Biology*, *9*(3), 315–321. https://doi.org/10.1016/j.pbi.2006.03.005

Azmach, G., Menkir, A., Spillane, C., & Gedil, M. (2018). Genetic Loci Controlling Carotenoid Biosynthesis in Diverse Tropical Maize Lines. *G3 Genes|Genomes|Genetics*, *8*(3), 1049–1065. https://doi.org/10.1534/g3.117.300511

Babu, R., Rojas, N. P., Gao, S., Yan, J., & Pixley, K. (2013). Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theoretical and Applied Genetics*, *126*(2), 389–399. https://doi.org/10.1007/s00122-012-1987-3

Bartley, G., Scolnik, P., & Giuliano, G. (2003). Molecular Biology of Carotenoid

Biosynthesis in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *45*, 287–301. https://doi.org/10.1146/annurev.pp.45.060194.001443

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

Bramley, P. M. (2002). Regulation of carotenoid formation during tomato fruit ripening and development. *Journal of Experimental Botany*, *53*(377), 2107–2113. https://doi.org/10.1093/jxb/erf059

Chander, S., Guo, Y. Q., Yang, X. H., Zhang, J., Lu, X. Q., Yan, J. B., Song, T. M., Rocheford, T. R., & Li, J. S. (2008). Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. *Theoretical and Applied Genetics*, *116*(2), 223–233. https://doi.org/10.1007/s00122-007-0661-7

Chandran, S., Pukalenthy, B., Adhimoolam, K., Manickam, D., Sampathrajan, V., Chocklingam, V., Eswaran, K., Arunachalam, K., Joikumar meetei, L., Rajasekaran, R., Muthusamy, V., Hossain, F., & Natesan, S. (2019). Marker-Assisted Selection to Pyramid the Opaque-2 (O2) and β-Carotene (crtRB1) Genes in Maize. *Frontiers in Genetics*, *10*. https://www.frontiersin.org/article/10.3389/fgene.2019.00859

Clotault, J., Peltier, D., Berruyer, R., Thomas, M., Briard, M., & Geoffriau, E. (2008). Expression of carotenoid biosynthesis genes during carrot root development. *Journal of Experimental Botany*, *59*(13), 3563–3573. https://doi.org/10.1093/jxb/ern210

Cruet-Burgos, C., Cox, S., Ioerger, B. P., Perumal, R., Hu, Z., Herald, T. J., Bean, S. R., & Rhodes, D. H. (2020). Advancing provitamin A biofortification in sorghum:

Genome-wide association studies of grain carotenoids in global germplasm. *The Plant Genome*, *13*(1). https://doi.org/10.1002/tpg2.20013

Cruet-Burgos, C., Morris, G. P., & Rhodes, D. H. (2022). *Characterization of Grain Carotenoids in Global Sorghum Germplasm to Guide Genomics-Assisted Breeding Strategies* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1774418/v1

Da Silva Messias, R., Galli, V., Dos Anjos e Silva, S. D., & Rombaldi, C. V. (2014). Carotenoid Biosynthetic and Catabolic Pathways: Gene Expression and Carotenoid Content in Grains of Maize Landraces. *Nutrients*, *6*(2), 546–563. https://doi.org/10.3390/nu6020546

Debelo, H., Albertsen, M., Simon, M., Che, P., & Ferruzzi, M. (2020). Identification and Characterization of Carotenoids, Vitamin E and Minerals of Biofortified Sorghum. *Current Developments in Nutrition*, *4*(Supplement_2), 1792–1792. https://doi.org/10.1093/cdn/nzaa067_019

Diepenbrock, C. H., Ilut, D. C., Magallanes-Lundback, M., Kandianis, C. B., Lipka, A. E., Bradbury, P. J., Holland, J. B., Hamilton, J. P., Wooldridge, E., Vaillancourt, B., Góngora-Castillo, E., Wallace, J. G., Cepela, J., Mateos-Hernandez, M., Owens, B. F., Tiede, T., Buckler, E. S., Rocheford, T., Buell, C. R., … DellaPenna, D. (2021). Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *The Plant Cell*, *33*(4), 882–900. https://doi.org/10.1093/plcell/koab032

FAOSTAT. (2021). *Food and Agriculture Organization of the United Nations*. http://www.fao.org/faostat/en/#data/QC

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., & Kresovich, S. (2008). Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Science*, *48*(5), 1732–1743. https://doi.org/10.2135/cropsci2007.12.0684

Fu, Z., Chai, Y., Zhou, Y., Yang, X., Warburton, M. L., Xu, S., Cai, Y., Zhang, D., Li, J., & Yan, J. (2013). Natural variation in the sequence of PSY1 and frequency of favorable polymorphisms among tropical and temperate maize germplasm. *Theoretical and Applied Genetics*, *126*(4), 923–935. https://doi.org/10.1007/s00122-012-2026-0

Gerik, T. (n.d.). *Sorghum Growth and Development* (p. 8).

Gonzalez-Jorge, S., Ha, S.-H., Magallanes-Lundback, M., Gilliland, L. U., Zhou, A., Lipka, A. E., Nguyen, Y.-N., Angelovici, R., Lin, H., Cepela, J., Little, H., Buell, C. R., Gore, M. A., & DellaPenna, D. (2014). CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds. *The Plant Cell*, *25*(12), 4812–4826. https://doi.org/10.1105/tpc.113.119677

Goswami, R., Zunjare, R. U., Khan, S., Baveja, A., Muthusamy, V., & Hossain, F. (2019). Marker-assisted introgression of rare allele of β-carotene hydroxylase (crtRB1) gene into elite quality protein maize inbred for combining high lysine, tryptophan and provitamin A in maize. *Plant Breeding*, *138*(2), 174–183. https://doi.org/10.1111/pbr.12676

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., Stapleton, A. E., Vallabhaneni, R., Williams, M., Wurtzel, E. T., Yan, J., &

Buckler, E. S. (2008). Natural Genetic Variation in *Lycopene Epsilon Cyclase* Tapped for Maize Biofortification. *Science*, *319*(5861), 330–333. https://doi.org/10.1126/science.1150255

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology*, *4*(3), 210–218. https://doi.org/10.1016/S1369-5266(00)00163-1

Jin, X., Bai, C., Bassie, L., Nogareda, C., Romagosa, I., Twyman, R. M., Christou, P., & Zhu, C. (2019). ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (Zea mays) β-carotene hydroxylase 2 gene. *New Phytologist*, *222*(2), 793–804. https://doi.org/10.1111/nph.15614

Kandianis, C. B., Stevens, R., Liu, W., Palacios, N., Montgomery, K., Pixley, K., White, W. S., & Rocheford, T. (2013). Genetic architecture controlling variation in grain carotenoid composition and concentrations in two maize populations. *Theoretical and Applied Genetics*, *126*(11), 2879–2895. https://doi.org/10.1007/s00122-013-2179-5

Kean, E. G., Ejeta, G., Hamaker, B. R., & Ferruzzi, M. G. (2007). Characterization of Carotenoid Pigments in Mature and Developing Kernels of Selected Yellow-Endosperm Sorghum Varieties. *Journal of Agricultural and Food Chemistry*, *55*(7), 2619–2626. https://doi.org/10.1021/jf062939v

Kumar, J., Saripalli, G., Gahlaut, V., Goel, N., Meher, P. K., Mishra, K. K., Mishra, P. C., Sehgal, D., Vikram, P., Sansaloni, C., Singh, S., Sharma, P. K., & Gupta, P. K. (2018). Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (Triticum aestivum L.) using multi-locus and multi-traits GWAS. *Euphytica*, *214*(11), 219. https://doi.org/10.1007/s10681-018-2284-2

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Li, F., Vallabhaneni, R., & Wurtzel, E. T. (2008). PSY3, a New Member of the Phytoene Synthase Gene Family Conserved in the Poaceae and Regulator of Abiotic Stress-Induced Root Carotenogenesis. *Plant Physiology*, *146*(3), 1333–1345. https://doi.org/10.1104/pp.107.111120

Liu, L., Jeffers, D., Zhang, Y., Ding, M., Chen, W., Kang, M. S., & Fan, X. (2015). Introgression of the crtRB1 gene into quality protein maize inbred lines using molecular markers. *Molecular Breeding: New Strategies in Plant Improvement*, *35*(8), 154. https://doi.org/10.1007/s11032-015-0349-7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Muthusamy, V., Hossain, F., Thirunavukkarasu, N., Choudhary, M., Saha, S., Bhat, J. S., Prasanna, B. M., & Gupta, H. S. (2014). Development of β-Carotene Rich Maize Hybrids through Marker-Assisted Introgression of β-carotene hydroxylase Allele. *PLoS ONE*, *9*(12), e113583. https://doi.org/10.1371/journal.pone.0113583

Ohmiya, A. (2009). Carotenoid cleavage dioxygenases and their apocarotenoid products in plants. *Plant Biotechnology*, *26*(4), 351–358. https://doi.org/10.5511/plantbiotechnology.26.351

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C. R.,

Buckler, E. S., DellaPenna, D., Gore, M. A., & Rocheford, T. (2014). A Foundation for Provitamin A Biofortification of Maize: Genome-Wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics*, *198*(4), 1699–1716. https://doi.org/10.1534/genetics.114.169979

Palaisa, K. A., Morgante, M., Williams, M., & Rafalski, A. (2003). Contrasting Effects of Selection on Sequence Diversity and Linkage Disequilibrium at Two Phytoene Synthase Loci[W]. *The Plant Cell*, *15*(8), 1795–1806. https://doi.org/10.1105/tpc.012526

Panchy, N., Lehti-Shiu, M., & Shiu, S.-H. (2016). Evolution of Gene Duplication in Plants. *Plant Physiology*, *171*(4), 2294–2316. https://doi.org/10.1104/pp.16.00523

Perrin, F., Hartmann, L., Dubois-Laurent, C., Welsch, R., Huet, S., Hamama, L., Briard, M., Peltier, D., Gagné, S., & Geoffriau, E. (2017). Carotenoid gene expression explains the difference of carotenoid accumulation in carrot root tissues. *Planta*, *245*(4), 737–747. https://doi.org/10.1007/s00425-016-2637-9

Priya, R., Sneha, P., Dass, J. F. P., Doss C, G. P., Manickavasagam, M., & Siva, R. (2019). Exploring the codon patterns between CCD and NCED genes among different plant species. *Computers in Biology and Medicine*, *114*, 103449. https://doi.org/10.1016/j.compbiomed.2019.103449

Roozeboom, K. L., & Prasad, P. V. V. (2019). Sorghum Growth and Development. In *Sorghum* (pp. 155–172). John Wiley & Sons, Ltd. https://doi.org/10.2134/agronmonogr58.c8

Sun, T., & Li, L. (2020). Toward the ‘golden’ era: The status in uncovering the regulatory

control of carotenoid accumulation in plants. *Plant Science*, *290*, 110331. https://doi.org/10.1016/j.plantsci.2019.110331

Sun, T., Rao, S., Zhou, X., & Li, L. (2022). Plant carotenoids: Recent advances and future perspectives. *Molecular Horticulture*, *2*(1), 3. https://doi.org/10.1186/s43897-022-00023-2

Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, *128*(5), 851–864. https://doi.org/10.1007/s00122-015-2475-3

Vallabhaneni, R., Gallagher, C. E., Licciardello, N., Cuttriss, A. J., Quinlan, R. F., & Wurtzel, E. T. (2009). Metabolite Sorting of a Germplasm Collection Reveals the Hydroxylase3 Locus as a New Target for Maize Provitamin A Biofortification. *Plant Physiology*, *151*(3), 1635–1645. https://doi.org/10.1104/pp.109.145177

Vallabhaneni, R., & Wurtzel, E. T. (2009). Timing and Biosynthetic Potential for Carotenoid Accumulation in Genetically Diverse Germplasm of Maize. *Plant Physiology*, *150*(2), 562–572. https://doi.org/10.1104/pp.109.137042

Vennapusa, A. R., Somayanda, I. M., Doherty, C. J., & Jagadish, S. V. K. (2020). A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Scientific Reports*, *10*(1), 16887. https://doi.org/10.1038/s41598-020-73958-5

Wong, J. C., Lambert, R. J., Wurtzel, E. T., & Rocheford, T. R. (2004). QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. *Theoretical and Applied Genetics*,

108(2), 349–359. https://doi.org/10.1007/s00122-003-1436-4

World Health Organization. (2009). *Global prevalence of vitamin A deficiency in populations at risk 1995–2005*. World Health Organization. https://apps.who.int/iris/handle/10665/101008

Worzella, W. W., Khalidy, R., Badawi, Y., & Daghir, S. (1965). Inheritance of Beta-Carotene in Grain Sorghum Hybrids. *Crop Science*, *5*(6), 591–592.

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. In E. Mathé & S. Davis (Eds.), *Statistical Genomics: Methods and Protocols* (pp. 283–334). Springer. https://doi.org/10.1007/978-1-4939-3578-9_15

Yan, J., Kandianis, C. B., Harjes, C. E., Bai, L., Kim, E.-H., Yang, X., Skinner, D. J., Fu, Z., Mitchell, S., Li, Q., Fernandez, M. G. S., Zaharieva, M., Babu, R., Fu, Y., Palacios, N., Li, J., DellaPenna, D., Brutnell, T., Buckler, E. S., … Rocheford, T. (2010). Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nature Genetics*, *42*(4), 322–327. https://doi.org/10.1038/ng.551

Zhai, S., Xia, X., & He, Z. (2016). Carotenoids in Staple Cereals: Metabolism, Regulation, and Genetic Manipulation. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.0119

CHAPTER 5. CONCLUSION

Sorghum is an important target for vitamin A biofortification, as it is the fourth most consumed cereal among the countries with severe VAD in South East Asia and Africa. However, among the accessions that were evaluated, the current concentrations of carotenoids in sorghum grain are below target levels and biofortification is necessary to increase them further. In this study high carotenoid lines were identified in global germplasm (CAP/SAP) that could serve as donor parents for breeding efforts . Potentially high carotenoid accessions were also identified in unexplored germplasm that, if validated with HPLC, could also be incorporated into breeding efforts to increase genetic diversity.

This research confirmed that increasing sorghum grain carotenoids through breeding is feasible. Transgressive segregation was observed for the six $F_{2:3}$ biparental families, suggesting that complementary alleles exist in the germplasm. To increase carotenoids further, the next steps should be to generate crosses between high carotenoid progenies in the six $F_{2:3}$ biparental families, as well as with high carotenoid accessions identified both in the global germplasm (SAP/CAP) and in the unexplored germplasm.

Phenotyping carotenoids through HPLC is currently considered the most accurate method of measuring concentrations. However, despite using a relatively high-throughput method for both extraction and HPLC quantification of carotenoids, a limitation of this study is that only one biological replicate and one technical replicate could be quantified for the $F_{2:3}$ progenies due to the high number of samples (n=930) and enormous length of time needed for extraction and quantification. Low replication could potentially skew the phenotype results and significance observed for the KASP markers, as well as as for

the family distributions. Another limitation is that two different HPLC instruments were used in the studies in Chapters 2 and 3, and differences of sensitivity between instruments were observed. The Perkin Elmer LC 300 UHPLC used in Chapter 3 was unable to accurately detect β-carotene concentrations in the progenies and parental lines. It is possible that a longer column and/or an increased run time could resolve this issue, but that would significantly increase the amount of phenotyping time needed. For a sorghum vitamin A biofortification breeding program, this poses an even more challenging scenario as it suggests that an instrument with above average sensitivity would be required during initial stages, further increasing the phenotyping cost.

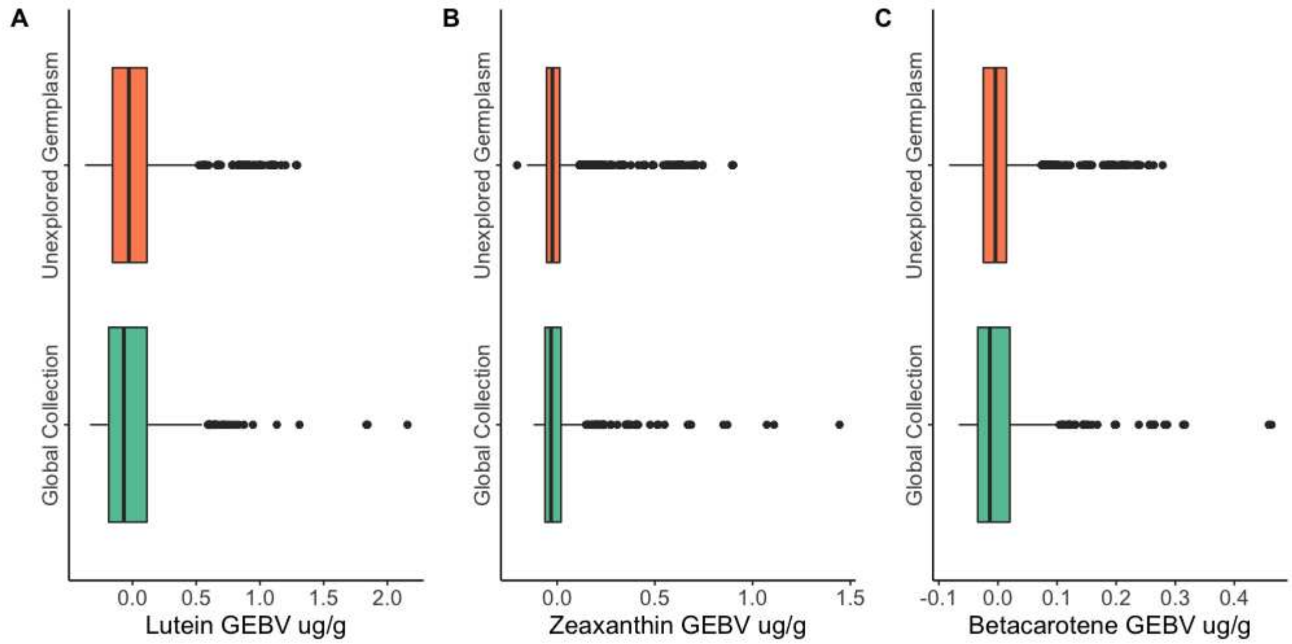
This study also evaluated if genomics-assisted breeding could be a feasible strategy to select sorghum grain carotenoids. Results in this study suggest that carotenoid variation in sorghum has an omnigenic architecture, meaning both oligogenic and polygenic. Under the omnigenic hypothesis, a breeding strategy using both marker-assisted selection and genomic selection at different points in the breeding pipeline could be used to accelerate breeding efforts. This study demonstrated that genomic selection was a good method for selection, obtaining favorable prediction accuracies for the three carotenoids evaluated—lutein, zeaxanthin and β-carotene. However, as most global germplasm seem to lack high carotenoid alleles, I propose that the first steps for biofortification breeding should be to introgress high-carotenoid alleles into germplasm in breeding programs through marker-assisted selection. Ten KASP markers were developed to use in MAS, targeting genomic regions associated with variation in carotenoid concentrations. Among the ten markers, three are very promising, as they are within the open reading frames of carotenoid biosynthesis genes *β-OH* and *ZEP*. One potential shortcoming of this study was the skewness observed for the genetic classes of

most of the markers evaluated, which could have impacted marker significance and predictive ability. For future validations, including parental lines with known low carotenoid content and/or contrasting alleles should be included to obtain a more balanced distribution between genotype classes.

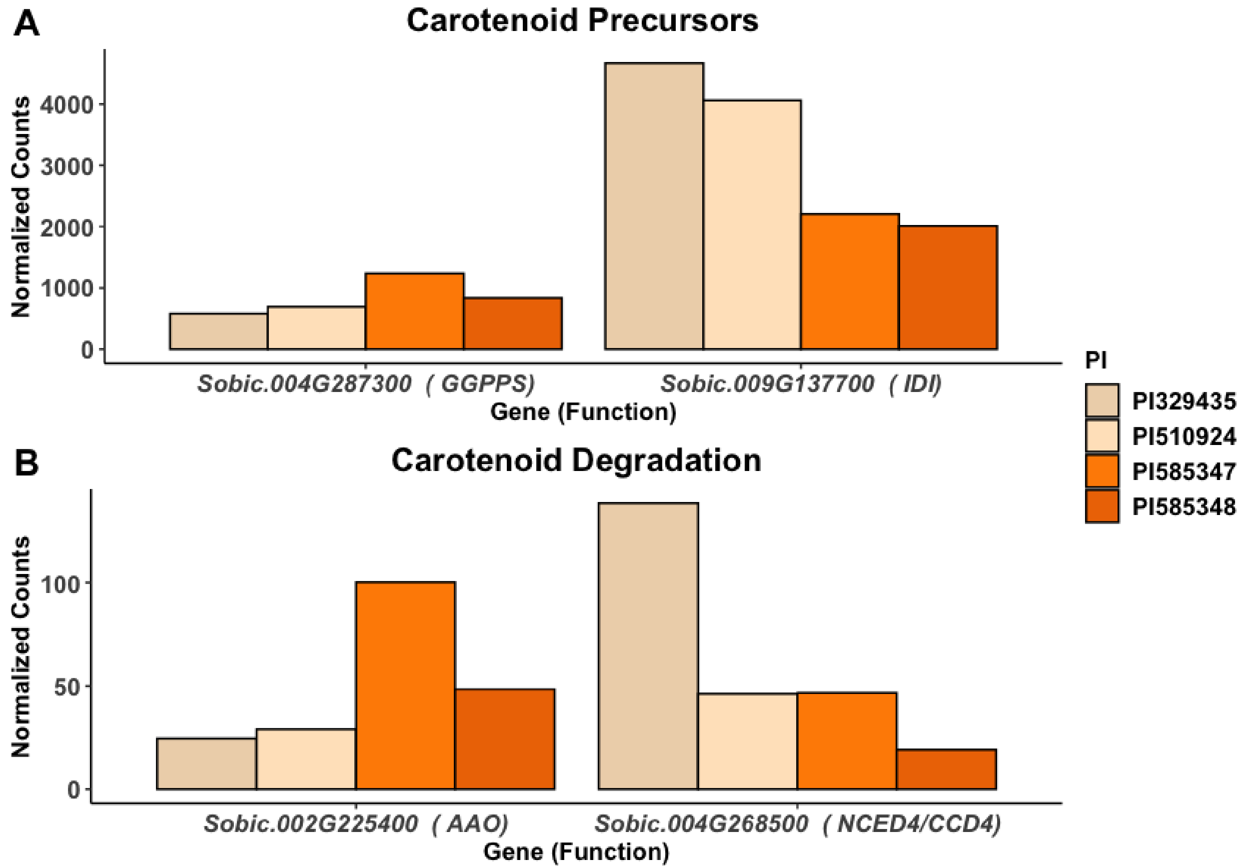
Allelic variants with different patterns of gene expression can potentially be used to increase carotenoid concentrations by increasing biosynthesis or reducing degradation throughout grain development. In our transcriptomic study we observed that many of the genes in the carotenoid precursor MEP pathways, carotenoid biosynthesis and carotenoid degradation pathways are differentially expressed over time through grain development, but not between high vs. low carotenoid content groups. Interestingly, genes that were differentially expressed between high and low carotenoid groups across developmental stages were primarily in the precursor MEP pathway or carotenoid degradation pathways, not the carotenoid biosynthesis pathway. The sorghum *PSY* gene (Sobic.002G292600) should be further investigated, as it was very highly expressed in the PI585348 accession, which is among the highest β-carotene concentration accessions measured. Follow up studies should sequence the four accessions used in the RNA-seq experiment —PI585348, PI585347, PI510924 and PI329435— for the *PSY* gene as well as its flanking regions, to determine if there exists an allelic variant responsible for the differences in expression.

These studies reveal insights into the genetic architecture of sorghum grain carotenoids and provide a profile of differential gene expression to further guide breeding efforts. The molecular tools here developed can be used to initiate efforts in vitamin A biofortification of sorghum.

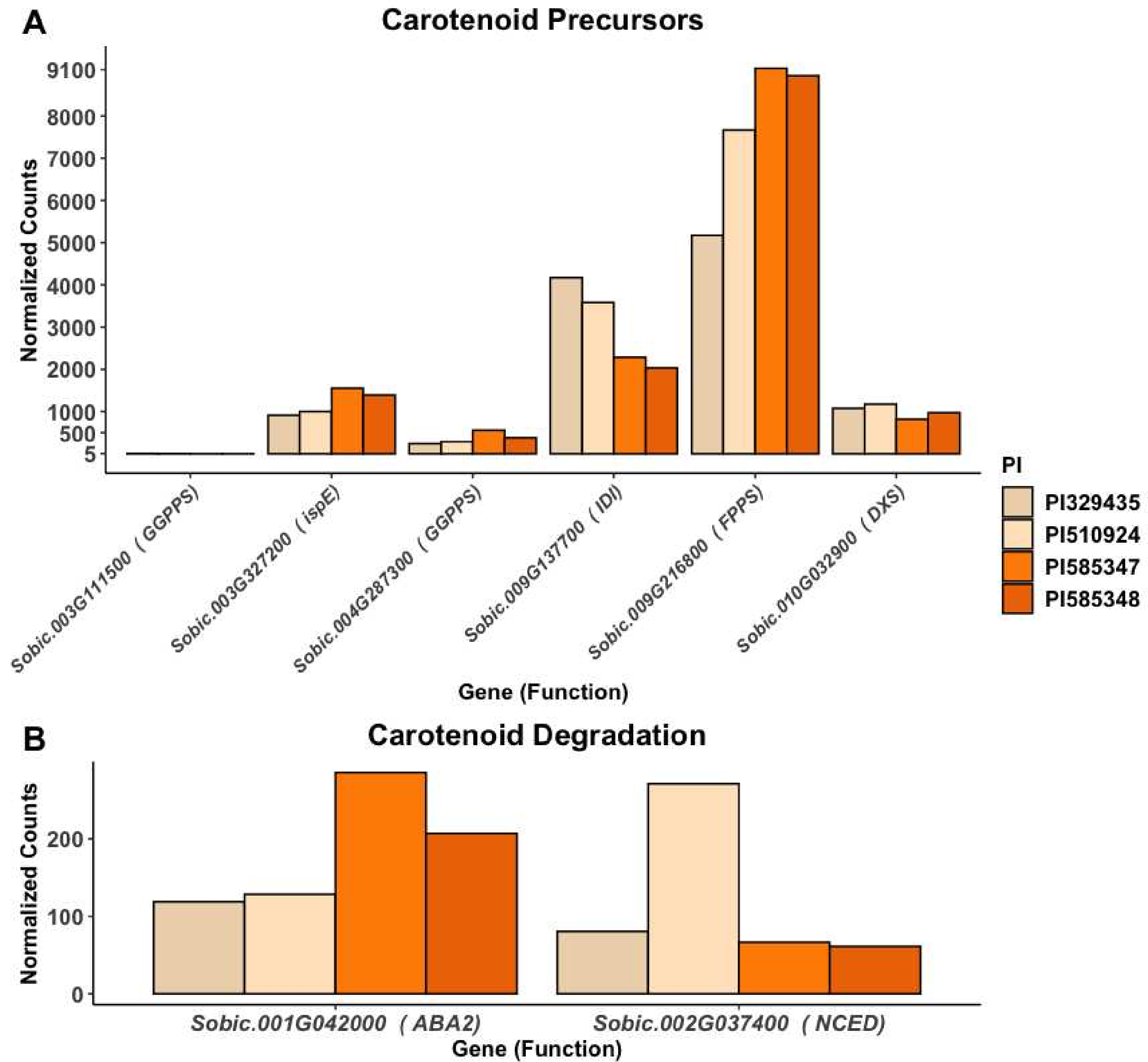
APPENDIX A



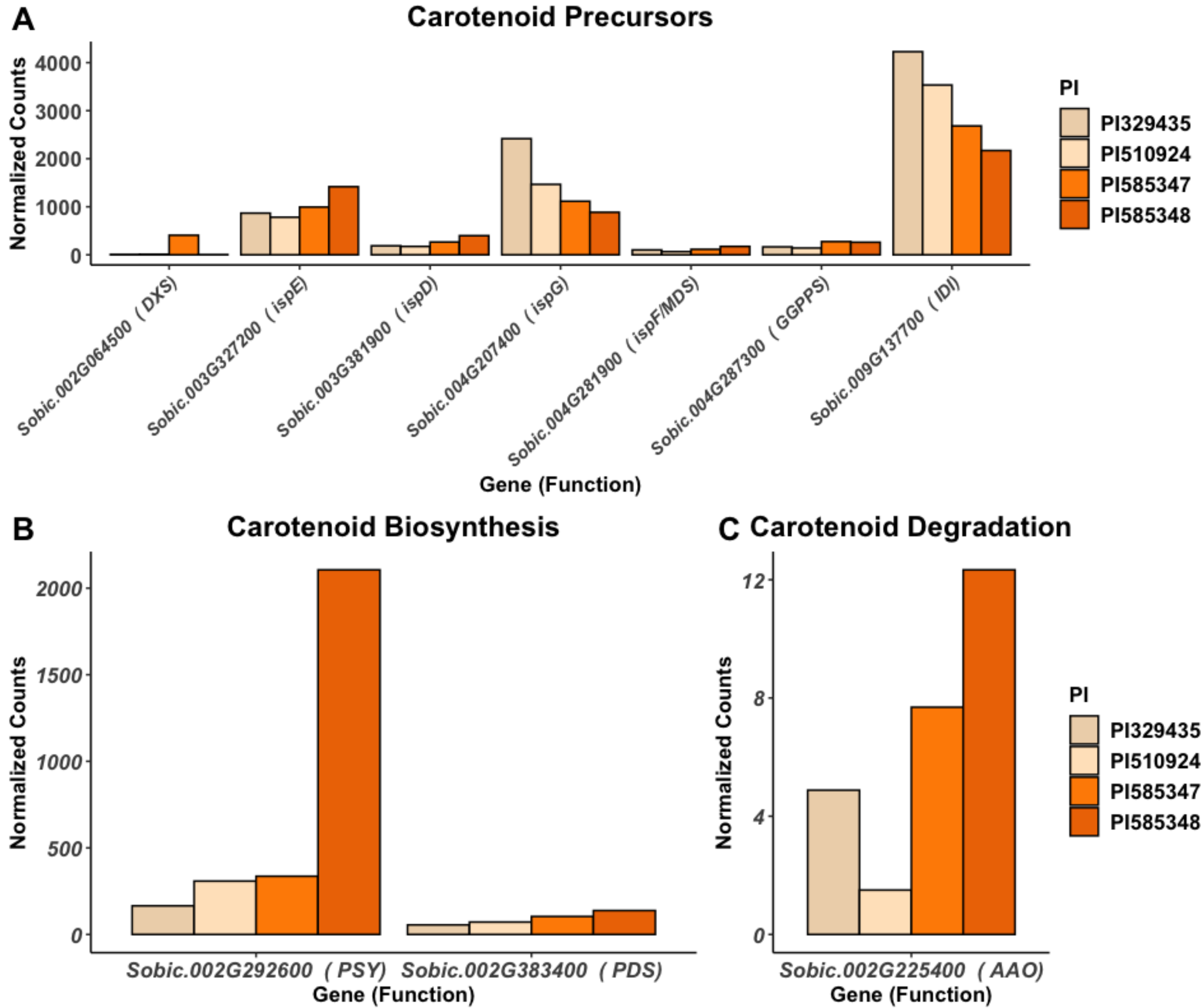
Supplemental Figure 1. Comparison of GEBV for lutein, zeaxanthin and β-carotene for the SAP/CAP global collection and the unexplored germplasm collections. Boxplot of GEBV for A) lutein; B) zeaxanthin; and C) β-carotene



Supplemental Figure 2. Differentially expressed genes at 14 DAP for high versus low carotenoid sorghum accessions. Genotypes normalized transcript counts for A) carotenoid precursors (MEP pathway); B) carotenoid degradation pathways



Supplemental Figure 3. Differentially expressed genes at 28 DAP for high versus low carotenoid sorghum accessions. Genotypes normalized transcript counts for A) carotenoid precursors (MEP pathway); B) carotenoid degradation pathways



Supplemental Figure 4. Differentially expressed genes at 42 DAP for high versus low carotenoid sorghum accessions. Genotypes normalized transcript counts for A) carotenoid precursors (MEP pathway); B) carotenoid biosynthesis; C) carotenoid degradation pathways.

APPENDIX B

Supplemental Table 1. A priori candidate gene list, predicted transcripts and function.

Gene Name	Transcript	K0.ID	KEGG Enzyme ID	KEGG Gene ID	Other IDs	Pathway Name	KEGG Pathway Number
Sobic.001G018000	Sobic.001G018000.1	K09835	carotenoid isomerase	CRTISO		Carotenoid biosynthesis	ko00906
Sobic.001G308200	Sobic.001G308200.1	K09837	carotenoid epsilon hydroxylase	CYP97C1	LUT1	Carotenoid biosynthesis	ko00906
Sobic.002G072400	Sobic.002G072400.1	K00514	zeta-carotene desaturase	ZDS		Carotenoid biosynthesis	ko00906
Sobic.002G292600	Sobic.002G292600.1	K02291	15-cis-phytoene synthase	PSY		Carotenoid biosynthesis	ko00906
Sobic.002G383400	Sobic.002G383400.1	K02293	15-cis-phytoene desaturase	PDS		Carotenoid biosynthesis	ko00906
Sobic.003G197400	Sobic.003G197400.1	K06444	lycopene epsilon-cyclase	lcyE	LUT2/crtL2	Carotenoid biosynthesis	ko00906
Sobic.003G277400	Sobic.003G277400.1	K09839	violaxanthin de-epoxidase	VDE		Carotenoid biosynthesis	ko00906
Sobic.004G074000	Sobic.004G074000.1	K06443	lycopene beta-cyclase	lcyB	crtL1/crtY/LYC	Carotenoid biosynthesis	ko00906
Sobic.004G074000	Sobic.004G074000.2	K06443	lycopene beta-cyclase	lcyB	crtL1/crtY/LYC	Carotenoid biosynthesis	ko00906

Sobic.004G346000	Sobic.004G346000.1	K15747	beta-ring hydroxylase	CYP97A3	LUT5	Carotenoid biosynthesis	ko00906
Sobic.005G160500	Sobic.005G160500.1	K09835	carotenoid isomerase	CRTISO		Carotenoid biosynthesis	ko00906
Sobic.006G049200	Sobic.006G049200.1	K09839	violaxanthin de-epoxidase	VDE		Carotenoid biosynthesis	ko00906
Sobic.006G097500	Sobic.006G097500.1	K09838	zeaxanthin epoxidase	ZEP	aba1	Carotenoid biosynthesis	ko00906
Sobic.006G177400	Sobic.006G177400.1	K02293	15-cis-phytoene desaturase	PDS		Carotenoid biosynthesis	ko00906
Sobic.006G188200	Sobic.006G188200.1	K15746	beta-carotene 3-hydroxylase	β-OH	BCH/crtRB1/hyd3/crtZ	Carotenoid biosynthesis	ko00906
Sobic.006G188200	Sobic.006G188200.2	K15746	beta-carotene 3-hydroxylase	β-OH	BCH/crtRB1/hyd3/crtZ	Carotenoid biosynthesis	ko00906
Sobic.006G232600	Sobic.006G232600.1	K02293	15-cis-phytoene desaturase	PDS		Carotenoid biosynthesis	ko00906
Sobic.006G232600	Sobic.006G232600.2	K02293	15-cis-phytoene desaturase	PDS		Carotenoid biosynthesis	ko00906
Sobic.007G130400	Sobic.007G130400.1	K06443	lycopene beta-cyclase	lcyB	crtL1/crtY/LYC	Carotenoid biosynthesis	ko00906
Sobic.007G130400	Sobic.007G130400.2	K06443	lycopene beta-cyclase	lcyB	crtL1/crtY/LYC	Carotenoid biosynthesis	ko00906
Sobic.007G130400	Sobic.007G130400.3	K06443	lycopene beta-cyclase	lcyB	crtL1/crtY/LYC	Carotenoid biosynthesis	ko00906
Sobic.008G096800	Sobic.008G096800.1	K15744	zeta-carotene isomerase	Z-ISO		Carotenoid biosynthesis	ko00906

Sobic.008G096800	Sobic.008G096800.2	K15744	zeta-carotene isomerase	Z-ISO		Carotenoid biosynthesis	ko00906
Sobic.008G180800	Sobic.008G180800.1	K02291	15-cis-phytoene synthase	PSY		Carotenoid biosynthesis	ko00906
Sobic.010G276400	Sobic.010G276400.2	K02291	15-cis-phytoene synthase	PSY		Carotenoid biosynthesis	ko00906
Sobic.010G276400	Sobic.010G276400.3	K02291	15-cis-phytoene synthase	PSY		Carotenoid biosynthesis	ko00906
Sobic.001G042000	Sobic.001G042000.1	K09841	xanthoxin dehydrogenase	ABA2	SDR1/GIN1/ISI4/CIS4/SRE1	Carotenoid degradation	ko00906
Sobic.001G062300	Sobic.001G062300.1	K11817	indole-3-acetaldehyde oxidase	AAO		Carotenoid degradation	
Sobic.001G062500	Sobic.001G062500.1	K11817	indole-3-acetaldehyde oxidase	AAO		Carotenoid degradation	
Sobic.001G062500	Sobic.001G062500.2	K11817	indole-3-acetaldehyde oxidase	AAO		Carotenoid degradation	
Sobic.001G062600	Sobic.001G062600.1			AAO		Carotenoid degradation	
Sobic.001G062600	Sobic.001G062600.2			AAO		Carotenoid degradation	
Sobic.001G062600	Sobic.001G062600.3			AAO		Carotenoid degradation	

Sobic.001G155300	Sobic.001G155300.1	K09840	9-cis-epoxycarotenoid dioxygenase	NCED		Carotenoid degradation	ko00906
Sobic.001G509200	Sobic.001G509200.1	K11159	carotenoid cleavage oxygenase	CCD	CCD1/NCED1	Carotenoid degradation	
Sobic.002G037400	Sobic.002G037400.1	K09840	9-cis-epoxycarotenoid dioxygenase	NCED		Carotenoid degradation	ko00906
Sobic.002G043500	Sobic.002G043500.1	K09842	abscisic-aldehyde oxidase	AAO3		Carotenoid degradation	ko00906
Sobic.002G225400	Sobic.002G225400.1	K09843	(+)-abscisic acid 8'-hydroxylase	AAO		Carotenoid degradation	ko00906
Sobic.003G293600	Sobic.003G293600.1	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906
Sobic.004G268500	Sobic.004G268500.1	K09840	9-cis-epoxycarotenoid dioxygenase	NCED4/CCD4	NCED4/CCD4	Carotenoid degradation	ko00906
Sobic.004G268700	Sobic.004G268700.1	K09843	(+)-abscisic acid 8'-hydroxylase	AAO		Carotenoid degradation	ko00906
Sobic.005G002500	Sobic.005G002500.1	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906

Sobic.005G105700	Sobic.005G105700.1	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906
Sobic.006G170300	Sobic.006G170300.1	K17912	9-cis-beta-carotene 9',10'-cleaving dioxygenase	CCD7	MAX3/D17	Carotenoid degradation	ko00906
Sobic.007G170300	Sobic.007G170300.1	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906
Sobic.007G170300	Sobic.007G170300.3	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906
Sobic.007G170300	Sobic.007G170300.4	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal 13,14-cleaving dioxygenase	CCD8		Carotenoid degradation	ko00906
Sobic.007G170300	Sobic.007G170300.5	K17913	carlactone synthase / all-trans-10'-apo-beta-carotenal	CCD8		Carotenoid degradation	ko00906

			13,14-cleaving dioxygenase				
Sobic.010G050300	Sobic.010G050300.2	K11159	carotenoid cleavage oxygenase	CCD		Carotenoid degradation	
Sobic.010G050300	Sobic.010G050300.3	K11159	carotenoid cleavage oxygenase	CCD		Carotenoid degradation	
Sobic.010G050300	Sobic.010G050300.4	K11159	carotenoid cleavage oxygenase	CCD		Carotenoid degradation	
Sobic.001G102400	Sobic.001G102400.1	K03527	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	ispH	HDR/CLB6	MEP	M00096
Sobic.001G102500	Sobic.001G102500.1	K03527	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	ispH	HDR/CLB6	MEP	M00096
Sobic.002G064500	Sobic.002G064500.1	K01662	1-deoxy-D-xylulose-5-phosphate synthase	DXS		MEP	M00096
Sobic.002G064700	Sobic.002G064700.1	K01662	1-deoxy-D-xylulose-5-phosphate synthase	DXS		MEP	M00096
Sobic.002G330500	Sobic.002G330500.1	K01823	isopentenyl-diphosphate Delta-isomerase	IDI		MEP	M00366

Sobic.002G353300	Sobic.002G353300.1	K13789	geranylgeranyl diphosphate synthase, type II	GGPPS		MEP	M00366
Sobic.003G103300	Sobic.003G103300.1	K00099	1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR		MEP	M00096
Sobic.003G103300	Sobic.003G103300.2	K00099	1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR		MEP	M00096
Sobic.003G111500	Sobic.003G111500.1	K13789	geranylgeranyl diphosphate synthase, type II	GGPPS		MEP	M00366
Sobic.003G270500	Sobic.003G270500.1	K00787	farnesyl diphosphate synthase	FPPS		MEP	M00366
Sobic.003G327200	Sobic.003G327200.1	K00919	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	ispE	CMK	MEP	M00096
Sobic.003G381900	Sobic.003G381900.1	K00991	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	ispD	CMS	MEP	M00096
Sobic.004G207400	Sobic.004G207400.1	K03526	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	ispG		MEP	M00096

Sobic.004G281900	Sobic.004G281900.2	K01770	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	ispF/MDS	MDS	MEP	M00096
Sobic.004G287300	Sobic.004G287300.1	K13789	geranylgeranyl diphosphate synthase, type II	GGPPS		MEP	M00366
Sobic.004G327500	Sobic.004G327500.1	K01823	isopentenyl-diphosphate Delta-isomerase	IDI		MEP	M00096
Sobic.006G248000	Sobic.006G248000.1	K00787	farnesyl diphosphate synthase	FPPS		MEP	M00366
Sobic.007G070000	Sobic.007G070000.1	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366
Sobic.007G070000	Sobic.007G070000.2	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366
Sobic.007G070000	Sobic.007G070000.3	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366
Sobic.007G070000	Sobic.007G070000.4	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366

Sobic.007G070000	Sobic.007G070000.5	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366
Sobic.009G135500	Sobic.009G135500.1	K01662	1-deoxy-D-xylulose-5-phosphate synthase	DXS		MEP	M00096
Sobic.009G137700	Sobic.009G137700.1	K01823	isopentenyl-diphosphate Delta-isomerase	IDI		MEP	M00096
Sobic.009G216800	Sobic.009G216800.1	K00787	farnesyl diphosphate synthase	FPPS	FPS1	MEP	M00366
Sobic.010G032900	Sobic.010G032900.1	K01662	1-deoxy-D-xylulose-5-phosphate synthase	DXS		MEP	M00096
Sobic.010G229400	Sobic.010G229400.1	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366
Sobic.010G229400	Sobic.010G229400.2	K14066	geranyl diphosphate synthase	GPPS		MEP	M00366

Supplemental Table 2. Best Linear Unbiased Predictors (μg/g) for β-carotene, lutein, and zeaxanthin for SAP/CAP collection.

Taxa	β-carotene BLUP	Lutein BLUP	Zeaxanthin BLUP	Country of Origin	β-carotene rank
PI534092	0.798	2.680	1.057	Nigeria	Top 5%
PI533877	0.792	3.091	1.409	Nigeria	Top 5%
PI563398	0.612	1.457	0.308	UnitedStates	Top 5%
PI563068	0.585	1.314	0.935	UnitedStates	Top 5%
PI563447	0.547	0.756	0.589	UnitedStates	Top 5%
PI585348	0.531	1.059	0.561	Lebanon	Top 5%
PI656010	0.508	1.182	0.731	UnitedStates	Top 5%
PI585369	0.503	1.324	0.629	Lebanon	Top 5%
PI656040	0.496	1.324	0.777	NA	Top 5%
PI585347	0.486	1.569	0.952	Lebanon	Top 5%
PI656096	0.471	1.756	1.372	NA	Top 5%
PI534088	0.442	2.658	1.833	Nigeria	Top 5%
PI656091	0.418	1.444	0.356	NA	Top 5%
PI563455	0.411	1.685	0.724	UnitedStates	Top 5%
PI563392	0.403	1.170	0.622	Uganda	Top 5%
PI655989	0.403	1.221	0.327	UnitedStates	Top 5%
PI563450	0.398	1.420	1.192	UnitedStates	Top 5%
PI563453	0.396	0.893	0.935	UnitedStates	Top 5%
PI533762	0.378	1.411	0.306	NA	Top 5%
PI656004	0.361	1.572	0.563	UnitedStates	Other
PI613536	0.357	2.221	0.388	UnitedStates	Other
PI659695	0.354	0.892	0.196	NA	Other
PI656003	0.347	1.355	0.438	UnitedStates	Other
PI548797	0.323	0.931	0.215	UnitedStates	Other
PI563457	0.320	0.713	0.527	UnitedStates	Other
PI533760	0.314	0.518	0.149	NA	Other
PI533878	0.303	1.286	0.738	Nigeria	Other
PI562712	0.300	0.928	0.159	Mexico	Other
PI576375	0.294	0.661	0.275	Ethiopia	Other
PI561071	0.293	0.610	0.225	UnitedStates	Other
PI576426	0.290	1.335	0.205	Ethiopia	Other
PI534075	0.289	1.635	0.595	Nigeria	Other
PI642791	0.288	1.032	0.158	UnitedStates	Other
PI457747	0.288	1.006	0.176	Ethiopia	Other
PI585363	0.284	0.931	0.329	Lebanon	Other

PI534144	0.284	1.562	0.164	Uganda	Other
PI585374	0.282	0.763	0.575	Lebanon	Other
PI656018	0.282	0.644	0.097	UnitedStates	Other
PI629040	0.280	0.647	0.206	UnitedStates	Other
PI533838	0.276	0.461	0.198	Nigeria	Other
PI534155	0.270	1.294	0.232	Ethiopia	Other
PI533824	0.269	1.308	0.281	Nigeria	Other
PI656002	0.267	0.770	0.362	UnitedStates	Other
PI34911	0.265	0.586	0.271	UnitedStates	Other
PI595720	0.263	0.648	0.220	UnitedStates	Other
PI552861	0.261	0.592	0.134	UnitedStates	Other
PI656024	0.261	0.788	0.275	NA	Other
PI533921	0.258	1.478	0.236	Ethiopia	Other
PI585365	0.257	0.953	0.364	Lebanon	Other
PI595745	0.257	0.657	0.227	UnitedStates	Other
PI656011	0.254	0.633	0.178	UnitedStates	Other
PI656089	0.252	0.615	0.134	NA	Other
PI562777	0.248	0.599	0.304	UnitedStates	Other
PI656052	0.247	0.461	0.127	UnitedStates	Other
PI533821	0.246	0.997	0.175	Tanzania	Other
PI642793	0.245	0.628	0.234	UnitedStates	Other
PI533776	0.244	1.084	0.673	NA	Other
PI655986	0.244	0.698	0.162	UnitedStates	Other
PI656095	0.243	1.131	0.399	NA	Other
PI297247	0.243	0.985	0.184	Uganda	Other
PI565123	0.241	1.481	0.564	Zimbabwe	Other
PI533903	0.240	0.655	0.387	Ethiopia	Other
PI655985	0.236	0.523	0.150	UnitedStates	Other
PI585373	0.234	0.583	0.270	Lebanon	Other
PI656001	0.234	0.526	0.202	UnitedStates	Other
PI534138	0.233	0.892	0.208	Sudan	Other
PI656059	0.233	0.419	0.165	Nicaragua	Other
PI585379	0.233	0.806	0.105	Ethiopia	Other
PI656061	0.232	0.826	0.122	India	Other
PI585291	0.230	0.797	0.188	UnitedStates	Other
PI655977	0.229	0.556	0.202	UnitedStates	Other
PI533785	0.228	0.615	0.085	UnitedStates	Other
PI574455	0.227	0.452	0.216	UnitedStates	Other
PI656114	0.226	0.526	0.157	NA	Other

PI563448	0.222	0.792	0.295	UnitedStates	Other
PI560493	0.222	0.797	0.158	SouthAfrica	Other
PI533831	0.220	0.715	0.190	Sudan	Other
PI563452	0.220	1.227	0.385	UnitedStates	Other
PI563409	0.220	1.041	0.252	UnitedStates	Other
PI563430	0.216	1.138	0.248	UnitedStates	Other
PI656019	0.216	0.629	0.162	UnitedStates	Other
PI533789	0.216	0.754	0.229	NA	Other
PI534133	0.213	0.509	0.142	Ethiopia	Other
PI534135	0.213	0.589	0.140	Ethiopia	Other
PI533910	0.212	0.449	0.101	Sudan	Other
PI585346	0.212	0.684	0.341	Lebanon	Other
PI510950	0.210	0.427	0.092	Botswana	Other
PI656106	0.207	1.382	0.152	NA	Other
PI576130	0.205	0.651	0.162	India	Other
PI534047	0.204	0.752	0.158	NA	Other
PI656092	0.204	0.878	0.185	NA	Other
PI533902	0.204	0.697	0.275	Ethiopia	Other
PI656109	0.204	0.872	0.117	NA	Other
PI561072	0.201	0.726	0.475	UnitedStates	Other
PI656015	0.200	0.618	0.158	Sudan	Other
PI595740	0.200	0.879	0.179	UnitedStates	Other
PI569812	0.200	0.815	0.359	Sudan	Other
PI597982	0.199	0.632	0.218	UnitedStates	Other
PI533901	0.199	0.958	0.227	Japan	Other
PI533938	0.198	0.562	0.175	DemocraticRepublicoftheCongo	Other
PI533752	0.198	0.622	0.138	NA	Other
PI533882	0.198	0.737	0.133	Nigeria	Other
PI533856	0.196	0.730	0.179	India	Other
PI510977	0.196	0.457	0.099	Botswana	Other
PI533761	0.195	0.460	0.175	NA	Other
PI656088	0.194	0.747	0.166	NA	Other
PI533943	0.193	0.476	0.119	India	Other
PI585349	0.192	0.665	0.402	Lebanon	Other
PI510993	0.192	0.470	0.145	Botswana	Other
PI656072	0.192	0.523	0.131	NA	Other
PI655992	0.192	0.406	0.102	UnitedStates	Other

PI508372	0.191	0.843	0.401	Mali	Other
PI510917	0.190	0.557	0.211	Botswana	Other
PI629034	0.189	0.657	0.334	UnitedStates	Other
PI656053	0.188	0.391	0.066	UnitedStates	Other
PI533970	0.187	0.556	0.123	Uganda	Other
PI534123	0.187	0.779	0.182	Ethiopia	Other
PI642998	0.187	1.287	0.130	NA	Other
PI656000	0.186	0.499	0.151	UnitedStates	Other
PI656103	0.185	0.442	0.172	NA	Other
PI585350	0.185	0.539	0.141	Lebanon	Other
PI534132	0.185	0.779	0.158	Ethiopia	Other
PI641836	0.184	0.455	0.139	NA	Other
PI585351	0.184	0.714	0.186	Lebanon	Other
PI563454	0.184	0.708	0.381	UnitedStates	Other
PI35038	0.183	0.600	0.102	UnitedStates	Other
PI655979	0.183	0.571	0.162	UnitedStates	Other
PI533866	0.182	0.786	0.189	NA	Other
PI655982	0.182	0.416	0.119	Australia	Other
PI655991	0.182	0.449	0.152	UnitedStates	Other
PI597964	0.181	0.361	0.162	UnitedStates	Other
PI576437	0.181	0.532	0.294	Brazil	Other
PI607931	0.179	0.485	0.094	UnitedStates	Other
PI655993	0.179	0.581	0.155	UnitedStates	Other
PI656107	0.179	0.509	0.197	NA	Other
PI585355	0.179	0.789	0.442	Lebanon	Other
PI533766	0.178	0.560	0.133	NA	Other
PI656048	0.178	0.440	0.138	Mali	Other
PI656009	0.178	0.395	0.157	UnitedStates	Other
PI24969	0.177	0.848	0.157	China	Other
PI510994	0.177	0.314	0.078	Botswana	Other
PI656049	0.177	0.565	0.157	Botswana	Other
PI510991	0.177	0.473	0.092	Botswana	Other
PI534137	0.175	0.824	0.654	Sudan	Other
PI656022	0.175	0.459	0.138	UnitedStates	Other
PI656047	0.174	0.449	0.282	India	Other
PI655990	0.174	0.615	0.146	UnitedStates	Other
PI659691	0.173	0.839	0.213	UnitedStates	Other
PI576380	0.173	0.314	0.128	Ethiopia	Other
PI533758	0.173	0.532	0.204	NA	Other

PI510925	0.173	0.445	0.111	Botswana	Other
PI534079	0.171	0.544	0.164	Nigeria	Other
PI534115	0.171	0.460	0.152	Ethiopia	Other
PI656071	0.169	0.653	0.171	NA	Other
PI533830	0.169	1.092	0.192	Sudan	Other
PI534124	0.168	1.084	0.284	NA	Other
PI656083	0.168	0.480	0.143	NA	Other
PI597968	0.167	0.572	0.079	UnitedStates	Other
PI655997	0.166	0.817	0.170	UnitedStates	Other
PI533949	0.165	0.386	0.163	Sudan	Other
PI533979	0.165	0.456	0.123	SouthAfrica	Other
PI576425	0.165	0.691	0.188	Ethiopia	Other
PI656085	0.164	0.391	0.094	NA	Other
PI597971	0.164	0.586	0.180	UnitedStates	Other
PI533841	0.164	0.446	0.083	Nigeria	Other
PI656119	0.163	0.688	0.149	NA	Other
PI533976	0.163	0.443	0.119	SouthAfrica	Other
PI656079	0.163	0.321	0.096	NA	Other
PI656060	0.163	0.370	0.113	NA	Other
PI533997	0.162	0.245	0.094	NA	Other
PI597973	0.162	0.799	0.193	UnitedStates	Other
PI533972	0.161	0.578	0.102	Uganda	Other
PI576386	0.161	0.554	0.142	Uganda	Other
PI597972	0.160	0.625	0.159	UnitedStates	Other
PI585353	0.160	0.860	0.184	Lebanon	Other
PI510966	0.159	0.631	0.145	Botswana	Other
PI656111	0.158	0.422	0.125	NA	Other
PI655973	0.158	0.544	0.162	UnitedStates	Other
PI533757	0.157	0.601	0.115	NA	Other
PI656033	0.157	1.072	0.157	UnitedStates	Other
PI585376	0.156	0.863	0.184	Lebanon	Other
PI585359	0.156	0.538	0.157	Lebanon	Other
PI576385	0.156	0.419	0.096	Nigeria	Other
PI656012	0.156	0.746	0.154	UnitedStates	Other
PI576373	0.156	0.525	0.141	Japan	Other
PI510945	0.155	0.629	0.122	Botswana	Other
PI656070	0.155	0.503	0.148	NA	Other
PI655981	0.155	0.491	0.116	Mali	Other
PI576387	0.155	0.547	0.147	Sudan	Other

PI655975	0.155	0.605	0.108	UnitedStates	Other
PI656050	0.155	0.509	0.146	NA	Other
PI533799	0.154	0.401	0.100	NA	Other
PI656110	0.154	0.665	0.154	NA	Other
PI533967	0.154	0.326	0.211	Uganda	Other
PI534053	0.153	0.400	0.122	Uganda	Other
PI576376	0.152	0.648	0.092	Ethiopia	Other
PI659694	0.152	0.453	0.151	India	Other
PI576347	0.151	0.574	0.461	UnitedStates	Other
PI656087	0.151	0.585	0.169	NA	Other
PI561472	0.151	0.370	0.164	Honduras	Other
PI564164	0.151	0.498	0.178	UnitedStates	Other
PI533822	0.150	0.309	0.126	Tanzania	Other
PI534127	0.149	0.741	0.150	NA	Other
PI656055	0.148	0.361	0.121	UnitedStates	Other
PI533965	0.148	0.426	0.125	Uganda	Other
PI576435	0.147	0.546	0.088	Uganda	Other
PI595699	0.147	0.413	0.189	UnitedStates	Other
PI595739	0.146	0.382	0.162	UnitedStates	Other
PI576381	0.146	0.425	0.123	Ethiopia	Other
PI565120	0.145	0.464	0.194	Zimbabwe	Other
PI655988	0.145	0.496	0.111	UnitedStates	Other
PI564163	0.145	0.401	0.110	UnitedStates	Other
PI563449	0.144	0.741	0.375	UnitedStates	Other
PI597945	0.144	0.278	0.172	UnitedStates	Other
PI597965	0.144	0.484	0.136	UnitedStates	Other
PI656082	0.144	0.818	0.140	NA	Other
PI656118	0.143	0.752	0.142	NA	Other
PI656105	0.143	0.478	0.117	NA	Other
PI597976	0.142	0.509	0.075	UnitedStates	Other
PI655970	0.142	0.538	0.134	UnitedStates	Other
PI534112	0.142	0.526	0.111	India	Other
PI533750	0.142	0.549	0.171	NA	Other
PI533754	0.141	0.604	0.128	NA	Other
PI641824	0.141	0.781	0.067	NA	Other
PI576399	0.141	0.620	0.163	NA	Other
PI533912	0.141	0.457	0.127	Sudan	Other
PI655984	0.140	0.635	0.139	UnitedStates	Other
PI659693	0.140	0.864	0.153	NA	Other

PI597950	0.140	0.320	0.065	UnitedStates	Other
PI152651	0.140	0.449	0.086	Sudan	Other
PI656094	0.140	0.258	0.101	NA	Other
PI586046	0.140	0.539	0.305	Nigeria	Other
PI653616	0.139	0.566	0.149	UnitedStates	Other
PI656034	0.139	0.301	0.128	NA	Other
PI533996	0.139	0.505	0.166	NA	Other
PI533842	0.139	0.606	0.171	India	Other
PI533769	0.138	0.426	0.086	NA	Other
PI651492	0.138	0.558	0.180	UnitedStates	Other
PI656044	0.137	0.310	0.105	SouthAfrica	Other
PI576364	0.137	0.613	0.109	India	Other
PI534021	0.137	0.395	0.127	India	Other
PI534157	0.137	0.362	0.118	Ethiopia	Other
PI564165	0.137	0.428	0.114	UnitedStates	Other
PI597960	0.137	0.484	0.097	UnitedStates	Other
PI656104	0.136	0.681	0.139	NA	Other
PI656062	0.135	0.308	0.091	NA	Other
PI510932	0.135	0.260	0.134	Botswana	Other
PI533962	0.135	0.538	0.126	Sudan	Other
PI656029	0.135	0.378	0.106	UnitedStates	Other
PI597952	0.135	0.483	0.151	UnitedStates	Other
PI533957	0.135	0.621	0.133	NA	Other
PI552856	0.135	0.347	0.118	UnitedStates	Other
PI656043	0.134	0.250	0.106	NA	Other
PI533924	0.134	0.963	0.250	Ethiopia	Other
PI597958	0.134	0.343	0.086	UnitedStates	Other
PI656007	0.134	0.299	0.048	NA	Other
PI656036	0.133	0.443	0.136	NA	Other
PI656090	0.133	0.613	0.132	NA	Other
PI561073	0.133	0.323	0.171	UnitedStates	Other
PI510974	0.133	0.503	0.090	Botswana	Other
PI641874	0.132	0.506	0.090	NA	Other
PI510948	0.132	0.309	0.100	Botswana	Other
PI533788	0.132	0.969	0.151	NA	Other
PI656039	0.131	0.275	0.118	NA	Other
PI533807	0.131	0.410	0.141	NA	Other
PI534148	0.130	0.578	0.127	Ethiopia	Other
PI595743	0.130	0.400	0.171	UnitedStates	Other

PI534139	0.130	0.373	0.096	Sudan	Other
PI595714	0.130	0.492	0.117	UnitedStates	Other
PI656046	0.129	0.367	0.044	China	Other
PI576350	0.129	0.368	0.115	UnitedStates	Other
PI656016	0.129	0.324	0.068	UnitedStates	Other
PI533937	0.129	0.372	0.191	Rhodesia	Other
PI533759	0.129	0.596	0.088	NA	Other
PI533845	0.128	0.312	0.081	Nepal	Other
PI656057	0.128	0.474	0.144	UnitedStates	Other
PI656120	0.128	0.360	0.108	NA	Other
PI656086	0.128	0.661	0.136	NA	Other
PI656112	0.128	0.367	0.089	NA	Other
PI511002	0.127	0.364	0.096	Botswana	Other
PI656058	0.127	0.319	0.071	UnitedStates	Other
PI533792	0.127	0.339	0.089	NA	Other
PI659696	0.127	0.904	0.113	NA	Other
PI510921	0.127	0.283	0.060	Botswana	Other
PI656005	0.127	0.425	0.063	UnitedStates	Other
PI656074	0.127	0.260	0.194	NA	Other
PI510918	0.127	0.440	0.088	Botswana	Other
PI656117	0.126	0.309	0.110	NA	Other
PI534108	0.126	0.412	0.095	Uganda	Other
PI533948	0.126	0.297	0.057	UnitedStates	Other
PI576401	0.126	0.357	0.163	India	Other
PI655987	0.125	0.359	0.105	UnitedStates	Other
PI565121	0.125	0.273	0.083	Zimbabwe	Other
PI576434	0.125	0.288	0.093	Nigeria	Other
PI534167	0.124	0.420	0.093	NA	Other
PI656113	0.124	0.449	0.069	NA	Other
PI534099	0.124	0.377	0.106	Japan	Other
PI534063	0.123	0.356	0.096	Nigeria	Other
PI656076	0.123	0.282	0.082	NA	Other
PI576345	0.123	0.614	0.162	SouthAfrica	Other
PI510963	0.122	0.844	0.272	Botswana	Other
PI533794	0.122	0.302	0.076	NA	Other
PI656008	0.122	0.309	0.129	ElSalvador	Other
PI597966	0.121	0.367	0.128	UnitedStates	Other
PI655999	0.121	0.524	0.132	UnitedStates	Other
PI576348	0.121	0.555	0.076	UnitedStates	Other

PI534128	0.121	1.162	0.239	NA	Other
PI534070	0.120	0.272	0.140	Nigeria	Other
PI510981	0.120	0.559	0.092	Botswana	Other
PI655980	0.120	0.683	0.100	UnitedStates	Other
PI566819	0.120	0.468	0.121	UnitedStates	Other
PI585295	0.120	0.441	0.180	UnitedStates	Other
PI655976	0.119	0.384	0.130	UnitedStates	Other
PI533843	0.119	0.241	0.047	India	Other
PI534163	0.119	1.022	0.107	UnitedStates	Other
PI17548	0.119	0.592	0.080	Australia	Other
PI534117	0.119	0.513	0.108	Uganda	Other
PI655998	0.119	0.377	0.095	UnitedStates	Other
PI534097	0.118	0.326	0.074	Japan	Other
PI533985	0.118	0.397	0.110	NA	Other
PI656035	0.118	0.352	0.110	Niger	Other
PI534114	0.117	0.893	0.125	Pakistan	Other
PI533964	0.117	0.390	0.098	Sudan	Other
PI510923	0.117	0.280	0.053	Botswana	Other
PI533913	0.117	0.337	0.093	Sudan	Other
PI642992	0.117	0.411	0.135	NA	Other
PI576339	0.116	0.606	0.161	Zimbabwe	Other
PI656014	0.116	0.563	0.075	NA	Other
PI655995	0.116	0.765	0.082	UnitedStates	Other
PI641849	0.116	0.600	0.062	NA	Other
PI656077	0.116	0.353	0.121	NA	Other
PI576349	0.115	0.550	0.104	UnitedStates	Other
PI533991	0.115	0.482	0.082	NA	Other
PI533961	0.115	0.349	0.114	SouthAfrica	Other
PI510975	0.115	0.465	0.091	Botswana	Other
PI533955	0.115	0.498	0.179	SouthAfrica	Other
PI656075	0.115	0.213	0.094	NA	Other
PI656102	0.115	0.611	0.158	NA	Other
PI533911	0.114	0.403	0.151	Sudan	Other
PI659753	0.114	0.318	0.078	NA	Other
PI510951	0.114	0.357	0.062	Botswana	Other
PI533986	0.114	0.501	0.130	NA	Other
PI542718	0.114	0.402	0.041	China	Other
PI656017	0.114	0.271	0.062	UnitedStates	Other
PI576396	0.113	0.554	0.112	Uganda	Other

PI510931	0.113	0.310	0.082	Botswana	Other
PI576340	0.113	0.371	0.182	SouthAfrica	Other
PI656093	0.113	0.321	0.093	NA	Other
PI534105	0.113	0.378	0.116	Uganda	Other
PI563638	0.112	0.598	0.214	China	Other
PI533936	0.112	0.405	0.111	Tanzania	Other
PI533919	0.111	0.785	0.167	Ethiopia	Other
PI534046	0.111	0.330	0.114	NA	Other
PI655972	0.109	0.384	0.164	UnitedStates	Other
PI576394	0.109	0.352	0.109	Sudan	Other
PI576428	0.108	0.302	0.081	Ethiopia	Other
PI656098	0.108	0.296	0.126	NA	Other
PI533855	0.108	0.247	0.110	India	Other
PI576332	0.108	0.438	0.073	China	Other
PI597957	0.107	0.294	0.076	UnitedStates	Other
PI511257	0.107	0.578	0.173	Somalia	Other
PI656023	0.107	0.241	0.072	NA	Other
PI533998	0.107	0.459	0.121	NA	Other
PI534145	0.106	0.327	0.075	Rhodesia	Other
PI329435	0.106	0.332	0.194	Ethiopia	Other
PI533871	0.105	0.250	0.088	Nigeria	Other
PI656073	0.105	0.312	0.100	NA	Other
PI656042	0.105	0.352	0.100	NA	Other
PI595702	0.105	0.537	0.155	UnitedStates	Other
PI656028	0.105	0.290	0.093	Botswana	Other
PI597980	0.104	0.774	0.146	UnitedStates	Other
PI576333	0.104	0.265	0.052	UnitedStates	Other
PI656068	0.104	0.338	0.104	NA	Other
PI533954	0.104	0.355	0.094	SouthAfrica	Other
PI655983	0.103	0.528	0.105	NA	Other
PI656078	0.103	0.366	0.059	NA	Other
PI276837	0.103	0.466	0.150	Ethiopia	Other
PI533755	0.103	0.305	0.087	NA	Other
PI542406	0.102	0.412	0.172	UnitedStates	Other
PI534028	0.102	0.352	0.084	India	Other
PI534096	0.102	0.311	0.051	Mali	Other
PI510980	0.102	0.361	0.141	Botswana	Other
PI533956	0.102	0.370	0.060	Congo	Other
PI510933	0.101	0.391	0.095	Botswana	Other

PI533852	0.101	0.233	0.100	India	Other
PI563485	0.100	0.458	0.170	Senegal	Other
PI510946	0.100	0.247	0.067	Botswana	Other
PI601816	0.100	0.300	0.102	Mali	Other
PI656097	0.100	0.573	0.136	NA	Other
PI533989	0.100	0.353	0.120	NA	Other
PI510904	0.100	0.314	0.131	Botswana	Other
PI533940	0.100	0.284	0.086	Tanzania	Other
PI534116	0.099	0.557	0.184	Ethiopia	Other
PI329440	0.099	0.271	0.113	Ethiopia	Other
PI656031	0.099	0.224	0.101	NA	Other
PI576393	0.099	0.344	0.102	Ethiopia	Other
PI576418	0.098	0.434	0.103	Nigeria	Other
PI533810	0.098	0.495	0.124	India	Other
PI48770	0.098	0.315	0.094	SouthAfrica	Other
PI656037	0.098	0.212	0.098	SouthAfrica	Other
PI510709	0.097	0.345	0.096	Cameroon	Other
PI534101	0.097	0.447	0.111	Japan	Other
PI656064	0.097	0.284	0.056	NA	Other
PI576422	0.096	0.337	0.093	SouthAfrica	Other
PI656032	0.096	0.244	0.088	Senegal	Other
PI511011	0.096	0.358	0.080	Botswana	Other
PI598069	0.096	0.673	0.087	UnitedStates	Other
PI511020	0.095	0.227	0.073	Botswana	Other
PI534037	0.095	0.465	0.081	Chad	Other
PI510943	0.095	0.588	0.087	Botswana	Other
PI533869	0.095	0.178	0.042	Tanzania	Other
PI511016	0.095	0.334	0.052	Botswana	Other
PI656038	0.095	0.479	0.083	NA	Other
PI533800	0.094	0.261	0.107	NA	Other
PI510922	0.094	0.224	0.078	Botswana	Other
PI656108	0.094	0.268	0.107	NA	Other
PI510989	0.094	0.515	0.078	Botswana	Other
PI656121	0.094	0.314	0.053	NA	Other
PI511018	0.094	0.243	0.064	Botswana	Other
PI511258	0.093	0.490	0.118	Somalia	Other
PI651496	0.092	0.893	0.126	NA	Other
PI656065	0.092	0.399	0.088	NA	Other
PI510924	0.092	0.267	0.068	Botswana	Other

PI510934	0.091	0.305	0.048	Botswana	Other
PI656115	0.091	0.233	0.050	NA	Other
PI533863	0.091	0.326	0.114	Chad	Other
PI533915	0.091	0.167	0.048	Kenya	Other
PI534009	0.000	0.328	0.069	India	Bottom 5%
PI595718	0.000	0.285	0.097	UnitedStates	Bottom 5%
PI534054	0.090	0.252	0.046	Kenya	Bottom 5%
PI533833	0.090	0.468	0.106	Uganda	Bottom 5%
PI576390	0.090	0.299	0.093	India	Bottom 5%
PI291234	0.090	0.460	0.137	Jamaica	Bottom 5%
PI510972	0.089	0.256	0.060	Botswana	Bottom 5%
PI576352	0.089	0.311	0.075	Botswana	Bottom 5%
PI291235	0.088	0.451	0.171	Jamaica	Bottom 5%
PI534104	0.087	0.323	0.103	Uganda	Bottom 5%
PI510944	0.087	0.187	0.085	Botswana	Bottom 5%
PI533814	0.086	0.367	0.116	India	Bottom 5%
PI533927	0.086	0.319	0.110	Ethiopia	Bottom 5%
PI656081	0.081	0.306	0.053	NA	Bottom 5%
PI656056	0.080	0.301	0.120	UnitedStates	Bottom 5%
PI533839	0.077	0.216	0.036	Nigeria	Bottom 5%
PI511015	0.076	0.217	0.058	Botswana	Bottom 5%
PI653617	0.076	0.376	0.117	UnitedStates	Bottom 5%
PI656080	0.069	0.152	0.050	NA	Bottom 5%

Supplemental Table 3. Significant marker-trait associations above the Bonferroni significance threshold for β-carotene, lutein, and zeaxanthin in the SAP/CAP collection.

SNP	Chr[1]	Position	*p*-value	MAF[2]	nobs[3]	Rsquare of Model without SNP	Rsquare of Model with SNP	FDR Adjusted *p*-value	Effect	Trait	Candidate Gene in Proximity
S1_32449786	1	32449786	9.73E-08	0.099	347	0.405	0.457	0.001	-0.136	BLUP_zeaxanthin	NA
S1_4389901	1	4389901	7.37E-08	0.056	347	0.405	0.458	0.001	-0.135	BLUP_zeaxanthin	NA
S1_4389942	1	4389942	7.37E-08	0.056	347	0.405	0.458	0.001	0.135	BLUP_zeaxanthin	NA
S1_79870921	1	79870921	3.44E-09	0.076	347	0.405	0.470	0.000	0.128	BLUP_zeaxanthin	NA
S10_11610905	10	11610905	7.85E-11	0.066	347	0.405	0.485	0.000	0.136	BLUP_zeaxanthin	NA
S10_14377366	10	14377366	1.27E-09	0.051	345	0.326	0.405	0.000	0.076	BLUP_β-carotene	NA
S10_14377366	10	14377366	8.10E-08	0.050	347	0.405	0.458	0.001	0.131	BLUP_zeaxanthin	NA
S10_15044582	10	15044582	6.68E-11	0.062	347	0.405	0.485	0.000	-0.151	BLUP_zeaxanthin	NA
S10_46962872	10	46962872	2.55E-10	0.062	345	0.326	0.412	0.000	0.073	BLUP_β-carotene	NA
S10_47468424	10	47468424	6.12E-09	0.081	345	0.326	0.398	0.000	0.061	BLUP_β-carotene	NA
S10_7481171	10	7481171	5.91E-08	0.077	345	0.326	0.388	0.002	0.064	BLUP_β-carotene	NA
S10_7481175	10	7481175	5.91E-08	0.077	345	0.326	0.388	0.002	0.064	BLUP_β-carotene	NA
S10_7482447	10	7482447	5.91E-08	0.077	345	0.326	0.388	0.002	-0.064	BLUP_β-carotene	NA
S2_59428193	2	59428193	4.94E-09	0.064	345	0.326	0.399	0.000	-0.061	BLUP_β-carotene	NA
S2_61694864	2	61694864	3.09E-08	0.058	347	0.405	0.462	0.001	0.138	BLUP_zeaxanthin	Sobic.002G225400 (CYP707A)
S2_70883355	2	70883355	9.37E-08	0.061	347	0.405	0.458	0.001	-0.168	BLUP_zeaxanthin	NA
S2_72340265	2	72340265	5.70E-09	0.058	345	0.326	0.399	0.000	0.074	BLUP_β-carotene	NA
S3_18675376	3	18675376	1.52E-07	0.061	347	0.357	0.412	0.026	0.248	BLUP_lutein	NA
S4_22084015	4	22084015	1.06E-08	0.058	347	0.405	0.466	0.000	0.137	BLUP_zeaxanthin	NA
S4_23305772	4	23305772	3.04E-10	0.058	347	0.405	0.479	0.000	-0.143	BLUP_zeaxanthin	NA
S4_275231	4	275231	9.91E-08	0.072	347	0.357	0.414	0.026	0.240	BLUP_lutein	NA
S4_275341	4	275341	6.51E-07	0.121	347	0.357	0.407	0.037	-0.176	BLUP_lutein	NA
S4_33122903	4	33122903	8.16E-08	0.065	347	0.405	0.458	0.001	-0.120	BLUP_zeaxanthin	NA

S4_36699206	4	36699206	2.63E-08	0.058	347	0.405	0.462	0.001	0.145	BLUP_zeaxanthin	NA
S4_64220386	4	64220386	6.19E-08	0.056	347	0.405	0.459	0.001	-0.122	BLUP_zeaxanthin	NA
S5_58064120	5	58064120	6.27E-08	0.053	347	0.405	0.459	0.001	0.137	BLUP_zeaxanthin	NA
S5_9829024	5	9829024	3.09E-08	0.050	347	0.405	0.462	0.001	-0.153	BLUP_zeaxanthin	NA
S5_9837818	5	9837818	9.26E-09	0.055	347	0.405	0.466	0.000	0.156	BLUP_zeaxanthin	NA
S5_9840680	5	9840680	7.45E-09	0.050	347	0.405	0.467	0.000	-0.161	BLUP_zeaxanthin	NA
S6_45981388	6	45981388	1.24E-07	0.097	347	0.405	0.457	0.001	-0.113	BLUP_zeaxanthin	NA
S6_46330663	6	46330663	1.92E-08	0.117	347	0.405	0.464	0.000	0.109	BLUP_zeaxanthin	Sobic.006G097500 (ZEP)
S6_46717975	6	46717975	7.83E-11	0.104	347	0.405	0.485	0.000	-0.148	BLUP_zeaxanthin	Sobic.006G097500 (ZEP)
S6_46735094	6	46735094	4.02E-08	0.094	347	0.405	0.461	0.001	0.115	BLUP_zeaxanthin	Sobic.006G097500 (ZEP)
S6_47123508	6	47123508	2.88E-08	0.057	345	0.326	0.392	0.002	-0.065	BLUP_β-carotene	NA
S6_47123508	6	47123508	5.88E-07	0.056	347	0.357	0.407	0.037	-0.218	BLUP_lutein	Sobic.006G097500 (ZEP)
S6_47123508	6	47123508	1.71E-13	0.056	347	0.405	0.509	0.000	-0.173	BLUP_zeaxanthin	Sobic.006G097500 (ZEP)
S6_48562216	6	48562216	8.10E-08	0.065	345	0.326	0.387	0.003	0.066	BLUP_β-carotene	Sobic.006G097500 (ZEP)
S6_48562216	6	48562216	6.83E-11	0.065	347	0.405	0.485	0.000	0.158	BLUP_zeaxanthin	NA
S6_55520350	6	55520350	8.42E-08	0.072	347	0.405	0.458	0.001	-0.113	BLUP_zeaxanthin	NA
S6_55520384	6	55520384	8.42E-08	0.072	347	0.405	0.458	0.001	-0.113	BLUP_zeaxanthin	NA
S7_11375372	7	11375372	5.04E-08	0.059	347	0.405	0.460	0.001	-0.118	BLUP_zeaxanthin	NA
S8_3175664	8	3175664	3.63E-08	0.048	347	0.405	0.461	0.001	-0.153	BLUP_zeaxanthin	NA
S8_3175665	8	3175665	3.63E-08	0.048	347	0.405	0.461	0.001	-0.153	BLUP_zeaxanthin	NA
S8_3175686	8	3175686	3.63E-08	0.048	347	0.405	0.461	0.001	0.153	BLUP_zeaxanthin	NA
S8_55564934	8	55564934	8.55E-08	0.065	347	0.405	0.458	0.001	0.138	BLUP_zeaxanthin	NA
S8_55564944	8	55564944	8.55E-08	0.065	347	0.405	0.458	0.001	0.138	BLUP_zeaxanthin	NA
S8_55668146	8	55668146	3.65E-09	0.053	347	0.405	0.470	0.000	0.152	BLUP_zeaxanthin	NA
S8_7569911	8	7569911	1.69E-12	0.062	347	0.405	0.500	0.000	0.165	BLUP_zeaxanthin	NA

S8_7570056	8	7570056	2.24E-11	0.069	347	0.405	0.490	0.000	0.154	BLUP_zeaxanthin	NA
S9_2167856	9	2167856	7.52E-07	0.073	347	0.357	0.406	0.037	0.192	BLUP_lutein	NA
S9_2167889	9	2167889	3.96E-07	0.069	347	0.357	0.409	0.037	0.205	BLUP_lutein	NA
S9_2168864	9	2168864	6.03E-07	0.072	347	0.357	0.407	0.037	0.202	BLUP_lutein	NA
S9_3720142	9	3720142	4.06E-09	0.072	347	0.405	0.469	0.000	-0.128	BLUP_zeaxanthin	NA
S9_3723048	9	3723048	4.06E-09	0.072	347	0.405	0.469	0.000	-0.128	BLUP_zeaxanthin	NA
S9_5511363	9	5511363	1.80E-08	0.104	347	0.405	0.464	0.000	-0.118	BLUP_zeaxanthin	NA
S9_6565136	9	6565136	7.04E-08	0.055	347	0.405	0.459	0.001	-0.131	BLUP_zeaxanthin	NA

[1] Chromosome ; [2] Minimum Allele Frequency; [3] Number of observations

Supplemental Table 4. Top 5% GEBV for β-carotene in the unexplored germplasm

Taxa	GEBV β-carotene	Panel
PI563027	0.279	Nigeria
IS7812	0.263	Nigeria
PI647027	0.256	Nigeria
IS24737	0.255	Nigeria
PI647029	0.254	Nigeria
PI658758	0.241	Nigeria
IS7509	0.238	Nigeria
PI647024	0.236	Nigeria
PI646209	0.235	Nigeria
PI647025	0.232	Nigeria
PI646215	0.230	Nigeria
PI647018	0.223	Nigeria
PI646213	0.219	Nigeria
PI647020	0.218	Nigeria
IS7486	0.215	Nigeria
IS24713	0.212	Nigeria
PI647026	0.211	Nigeria
IS7806	0.211	Nigeria
PI647019	0.210	Nigeria
PI646210	0.208	Nigeria
PI647028	0.207	Nigeria
PI646214	0.201	Nigeria
PI646208	0.197	Nigeria
PI646217	0.196	Nigeria
PI585414	0.195	Nigeria
PI646212	0.194	Nigeria
IS7408	0.194	Nigeria
PI647021	0.191	Nigeria
PI513550	0.189	Niger
PI646207	0.186	Nigeria
PI646211	0.186	Nigeria
PI513445	0.186	Niger
IS24875	0.184	Nigeria
PI646216	0.178	Nigeria
IS24844	0.177	Nigeria
IS7643	0.159	Nigeria
IS24721	0.156	Nigeria

PI646223	0.156	Nigeria
PI513607	0.153	Niger
IS7386	0.151	Nigeria
PI513492	0.150	Niger
IS24704	0.146	Nigeria
PI513658	0.146	Niger
PI563451	0.145	CAP/SAP
PI585372	0.139	CAP/SAP
PI562978	0.123	Nigeria
PI513683	0.117	Niger
PI569043	0.114	Sudan
PI569234	0.112	Sudan
IS24775	0.107	Nigeria
PI513556	0.103	Niger
PI513581	0.102	Niger
PI569039	0.100	Sudan
IS39797	0.097	Nigeria
PI537083	0.097	Niger
PI513596	0.096	Niger
PI669302	0.095	Niger
PI513741	0.094	Niger
PI537080	0.093	Niger
PI513617	0.093	Niger
PI513620	0.091	Niger
PI513414	0.088	Niger
PI513659	0.087	Niger
PI513413	0.087	Niger
PI513827	0.087	Niger
PI562981	0.086	Nigeria
PI197462	0.085	Ethiopia
PI513452	0.085	Niger
PI513875	0.083	Niger
PI647764	0.082	Niger
PI513816	0.080	Niger
PI513451	0.080	Niger
PI569100	0.080	Sudan
PI540816	0.080	CAP/SAP
PI569150	0.079	Sudan
PI534055	0.079	Niger

PI513435	0.078	Niger
PI513440	0.078	Niger
PI647765	0.078	Niger
PI513660	0.078	Niger
PI513766	0.077	Niger
IS39677	0.077	Nigeria
PI453821	0.077	Ethiopia
IS7389	0.077	Nigeria
PI585406	0.076	Nigeria
PI513600	0.076	Niger
PI513824	0.076	Niger
PI513897	0.075	Niger
PI513795	0.074	Niger
PI568326	0.074	Sudan
PI569343	0.072	Sudan
PI665159	0.072	Senegal
PI562980	0.070	Nigeria
PI513829	0.070	Niger
PI513806	0.069	Niger
PI513797	0.068	Niger
PI585405	0.068	Nigeria
PI569995	0.066	Sudan
PI513808	0.065	Niger
PI570218	0.065	Sudan
PI513792	0.065	Niger
PI569038	0.063	Sudan
PI570410	0.063	Sudan
PI646206	0.062	Nigeria
PI457602	0.062	Ethiopia
PI513867	0.061	Niger
PI513714	0.060	Niger

Supplemental Table 5. KASP marker genotype group distribution in $F_{2:3}$ progenies

Marker Name	Genotype	Number of Progeny
snpSB00264	TT	650
	TC	44
	CC	41
snpSB00265	AA	335
	GG	82
	AG	16
snpSB00266	TT	575
	TC	109
	CC	83
snpSB00267	AA	141
	TA	55
	TT	23
snpSB00268	AA	243
	CC	3
	AC	2
snpSB00276	GG	129
	TG	51
	TT	24
snpSB00277	TT	678
	TC	51
	CC	37
snpSB00279	AA	562
	GA	140
	GG	56

snpSB00280	GG	559
	AG	132
	AA	57
snpSB00281	GG	530
snpSB00282	CC	754
	GC	5

Supplemental Table 6. Grain carotenoid content of sorghum genotypes.

Genotype	Lutein (μg/g)	Zeaxanthin (μg/g)	B-carotene (μg/g)
PI510924	0.42	0.19	0.06
PI329435	0.41	0.39	0.05
PI585347	4.13	2.17	0.59
PI585348	5.36	4.86	0.75

LIST OF ABBREVIATIONS

AAO- Abscisic aldehyde oxidase

ABA- Abscisic acid

ABS- Africa Biofortified Sorghum

ACB- African Centre for Biosafety

ALAD- Arid Land Agricultural Development

BHT-butylated hydroxytoluene

BLUP- best linear unbiased prediction

β-OH- β-carotene 3-hydroxylase

CAP- carotenoid panel

CCD- Carotenoid cleavage dioxygenase

CDP-M2P- 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate

CDP-ME- 4-CDP-2-C-methyl-D-erythritol

COO- Carotenoid cleavage oxygenases

CRD- complete randomized design

CRT- Carotenoid desaturase

CRTISO-Prolycopene isomerase

crtRB- β-Carotene Hydroxylase

CYP97A3- **β** -ring hydroxylase

CYP97C1- Carotenoid epsilon hydroxylase

DAHB- days after half bloom

DAP- days after pollination

DMAPP- Dimethylallyl pyrophosphate

DXP- 1-deoxy-D-xylulose-5-phosphate

DXR- Deoxyxylulose reductoisomerase

EAR- Estimated average requirement

FDR- false discovery rate

FPP- Farnesyl diphosphate

FPPS- Farnesyl diphosphate synthase

GA- Gibberellins

GEBV- genomic estimated breeding value

GGPP- Geranyl geranyl diphosphate

GGPPS- Geranyl geranyl diphosphate synthase

GPP- geranyl diphosphate

GS- genomic selection

GWAS- Genome wide association study

HGGT- Homogentisate geranylgeranyltransferase

HMBPP- 1-Hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate

HPLC- High-performance liquid chromatography

ICARDA- International Center for Agricultural Research in the Dry Areas

ICRISAT- International Crops Research Institute for the Semi-Arid Tropics

IDI- IPP isomerase

IPP- Isopentyl pyrophosphate

ispC- Deoxyxylulose reductoisomerase

ispD- CDP-ME transferase

ispE- CDP-ME kinase

ispF- 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase

ispG- HMBPP synthase

ispH- HMBPP reductase

KASP- Kompetitive allele specific PCR

lcyB- Lycopene beta cyclase

lcyE- Lycopene epsilon cyclase

LSD- least significant difference

MAS- Marker-assisted selection

Mb- megabase

MDS- 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase

ME-2,4cPP- 2-C-methyl-D-erythritol 2,4-cyclodiphosphate

MEP pathway- Methylerythritol phosphate pathway

MEP- 2-C-methyl-D-erythritol 4-phosphate

MLM- mixed linear model

NCED- 9-cis-epoxycarotenoid dioxygenase

NIR- Near infrared reflectance

NPGS- U.S. National Plant Germplasm System

OSP- Orange sweet potato

PCA- principal component analysis

PCR- polymerase chain reaction

PDS- Phytoene desaturase

PSY- Phytoene synthase

QTL- Quantitative trait loci

RCBD- complete randomized design

RIL- Recombinant inbred line

SAP- Sorghum association panel

SNP- single nucleotide polymorphism

UPLC- Ultra performance liquid chromatography

VAD- Vitamin A deficiency

VDE- Violaxanthin de-epoxidase

Z-ISO- ζ –carotene isomerase

ZDS- δ-carotene desaturase

ZEP- Zeaxanthin epoxidase

π- Nucleotide diversity