Author names and affiliations:

Christine H. Diepenbrock[1]*, Daniel C. Ilut[2], Maria Magallanes-Lundback[3], Catherine B. Kandianis[2,3]†, Alexander E. Lipka[4]†, Peter J. Bradbury[4,5], James B. Holland[6], John P. Hamilton[7], Edmund Wooldridge[7]†, Brieanne Vaillancourt[7], Elsa Góngora-Castillo[7], Jason G. Wallace[8], Jason Cepela[7], Maria Mateos-Hernandez[9]†, Brenda F. Owens[9]†, Tyler Tiede[9]†, Edward S. Buckler[2,4,5], Torbert Rocheford[9], C. Robin Buell[7], Michael A. Gore[2*], and Dean DellaPenna[3*]

[1]University of California, Department of Plant Sciences, Davis, CA 95616.
[2]Cornell University, Plant Breeding and Genetics Section, School of Integrative Plant Science, Ithaca, NY 14853.
[3]Michigan State University, Department of Biochemistry and Molecular Biology, East Lansing, MI 48824.
[4]Cornell University, Institute for Genomic Diversity, Ithaca, NY 14853.
[5]United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853.
[6]United States Department of Agriculture-Agricultural Research Service Plant Science Research Unit, Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695.
[7]Michigan State University, Department of Plant Biology, East Lansing, MI 48824.
[8]University of Georgia, Department of Crop & Soil Sciences, Athens, GA 30602.
[9]Purdue University, Department of Agronomy, West Lafayette, IN 47907.
†Present addresses: Nacre Innovations, Houston, TX 77002 (C.B.K.); University of Illinois at Urbana-Champaign, Department of Crop Sciences, Urbana, IL 61801 (A.E.L.); University of Michigan, Ann Arbor, MI 48109 (E. W.); CONACYT - Unidad de Biotecnologia, Centro de Investigación Científica de Yucatan, Merida, Yucatan, Mexico 97200 (E. G.-C.); University of Minnesota, Bioinformatics and Computational Biology, Minneapolis, MN 55455 (J. C.); Bayer, Stonington, IL 62567 (M. M.-H.); BASF, Dawson, GA 39842 (B.F.O.); Corteva Agriscience, St. Paul, MN 55108 (T.T.). *For correspondence: chdiepenbrock@ucdavis.edu, mag87@cornell.edu, dellapen@msu.edu.

Title: Eleven biosynthetic genes explain the majority of natural variation for carotenoid levels in maize grain

Short title: Genetic basis of maize grain carotenoids

ABSTRACT

Vitamin A deficiency remains prevalent in parts of Asia, Latin America and sub-Saharan Africa where maize is a food staple. Extensive natural variation exists for carotenoids in maize grain; to understand its genetic basis, we conducted a joint linkage and genome-wide association study in the U.S. maize nested association mapping panel. Eleven of the 44 detected quantitative trait loci (QTL) were resolved to individual genes. Six of these were expression QTL (eQTL), showing strong correlations between RNA-seq expression abundances and QTL allelic effect estimates across six stages of grain development. These six eQTL also had the largest percent phenotypic variance explained, and in major part comprised the three to five loci capturing the bulk of genetic variation for each trait. Most of these eQTL had highly correlated QTL allelic effect estimates across multiple traits, suggesting that pleiotropy within this pathway is largely regulated at the expression level. Significant pairwise epistatic interactions were also detected. These findings provide the most comprehensive genome-level understanding of the genetic and molecular control of carotenoids in any plant system, and a roadmap to accelerate breeding for provitamin A and other priority carotenoid traits in maize grain that should be readily extensible to other cereals.

INTRODUCTION

Carotenoids are lipid-soluble isoprenoids (typically C_{40}) synthesized by plants, algae and also some fungi, bacteria and yeast (reviewed in LI *et al.* 2016). Most carotenoids have yellow, orange, or red colors that are a function of the length of their conjugated double bond system and functional groups (KHOO *et al.* 2011). Carotenoids containing oxygen functional groups are termed xanthophylls, and those without such groups are termed carotenes. In plants, carotenoids are biosynthesized and localized in plastids, where they play numerous roles in photosystem structure and light harvesting and in photoprotection through their scavenging of singlet oxygen and dissipation of excess excitation energy via the xanthophyll cycle (JAHNS AND HOLZWARTH 2012). Additionally, 9-*cis* isomers of violaxanthin and neoxanthin are precursors for biosynthesis of abscisic acid (ABA), a plant hormone with critical roles in embryo dormancy and abiotic stress responses (KERMODE 2005; TUTEJA 2007), while 9-*cis* β-carotene is the substrate for synthesis of strigolactones, recently discovered plant hormones involved in branching and in attracting beneficial arbuscular mycorrhizae (reviewed in AL-BABILI AND BOUWMEESTER 2015, Jia et al. 2018).

Provitamin A carotenoids are an essential micronutrient in human and animal diets, as they are converted to vitamin A (retinol) in the body via oxidative cleavage (reviewed in EROGLU AND HARRISON 2013). The most abundant provitamin A carotenoids in the human diet are β-carotene, which yields two molecules of retinol, and β-cryptoxanthin and α-carotene, which yield one (STAHL AND SIES 2005; COMBS AND MCCLUNG 2016). Clinical vitamin A deficiency affects an estimated 127.2 million preschool children and 7.2 million pregnant women in countries determined to be at risk (WEST 2002). Symptoms can include xerophthalmia (‘dry eye’), which often progresses to night blindness, as well as increased morbidity and mortality from infections (reviewed in WEST AND DARNTON-HILL 2008). It is estimated that vitamin A deficiency is responsible for the deaths of 650,000 preschool children per year (RICE *et al.* 2004). Age-related macular degeneration is associated with deficiency in two xanthophylls—lutein and zeaxanthin—in the fovea, or cone-rich center of the retina, affecting 170 million adults in 2014 and projected to increase further with aging global populations (WONG *et al.* 2014; BEATTY *et al.* 1999; KRINSKY *et al.* 2003; ABDEL-AAL EL *et al.* 2013).

Maize is a primary food staple in much of Latin America, sub-Saharan Africa, and Asia,

where vitamin A deficiency remains highly prevalent (WEST 2002). There is extensive natural variation in levels of maize grain carotenoids, which are most highly concentrated in the vitreous (hard) portion of the endosperm (WEBER 1987; BLESSIN *et al.* 1963; HARJES *et al.* 2008). However, as a dietary staple the average provitamin A carotenoid levels of diverse yellow maize lines provide less than 20% of the target level established based on recommended dietary allowances (HARJES *et al.* 2008; BOUIS AND WELCH 2010; OWENS *et al.* 2014). Genetic improvement of maize grain carotenoid (provitamin A) levels through breeding, an example of biofortification, has been proposed as a cost-effective approach for ameliorating vitamin A deficiency in populations who are at risk thereof (GRAHAM *et al.* 2001; WELCH AND GRAHAM 2004; BOUIS AND WELCH 2010; DIEPENBROCK AND GORE 2015).

Carotenoids are derived from the five-carbon central intermediate isopentenyl pyrophosphate (IPP) produced by the plastid-localized methyl-D-erythritol-4-phosphate (MEP) pathway (Figure 1). The committed step toward carotenoid biosynthesis is the head-to-head condensation of two 20-carbon geranylgeranyl diphosphate (GGDP) molecules by the enzyme phytoene synthase to form phytoene. Phytoene is sequentially desaturated and isomerized to form lycopene, which is then cyclized with two β-rings to form β-carotene, or with one β-ring and one ε-ring to form α-carotene. α- and β-carotenes are hydroxylated twice to form lutein and zeaxanthin, respectively, and zeaxanthin further modified to yield violaxanthin and neoxanthin.

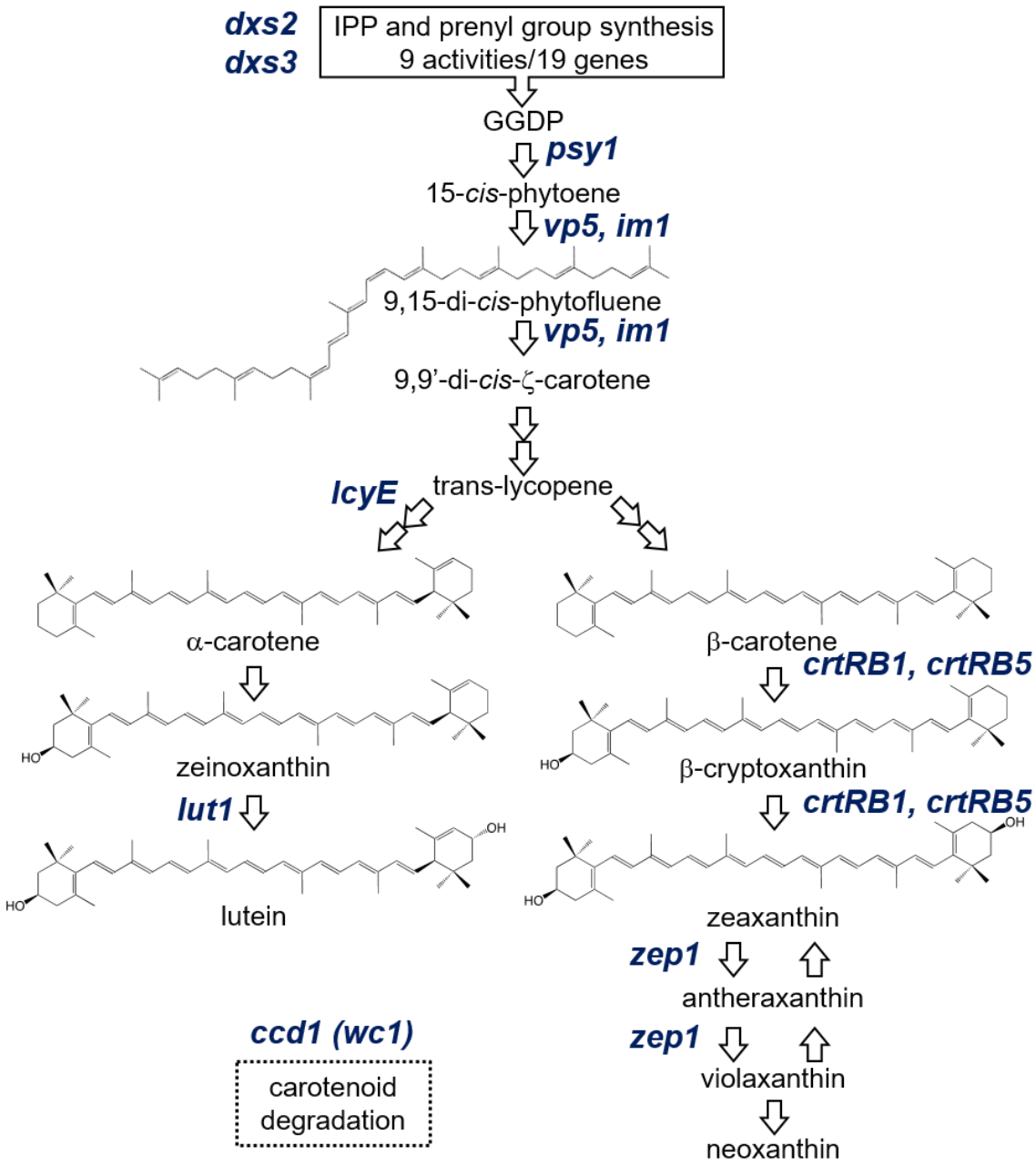


Figure 1. Carotenoid biosynthetic pathway in maize. Precursor pathways are summarized in black boxes. The *a priori* genes identified in this study are denoted in blue italics, and are placed at the pathway step(s) executed by the enzyme that they encode. Compound abbreviation: GGDP, geranylgeranyl diphosphate. Gene abbreviations: 1-deoxy-D-xylulose-5-phosphate synthase (*dxs2* and *3*); phytoene synthase (*psy1)*; phytoene desaturase (*vp5*); plastid terminal oxidase (*im1*); lycopene ε-cyclase *(lcyE);* ε-ring hydroxylase (*lut1*); β-carotene hydroxylase (*crtRB1*); zeaxanthin epoxidase (*zep1*); carotenoid cleavage dioxygenase (*ccd1*), *whitecap1* (*wc1*; a locus containing a varying number of copies of *ccd1*).

The MEP and carotenoid biosynthetic pathways are well-characterized in Arabidopsis, and the encoded genes are highly conserved across species, enabling the straightforward identification of homologs in other species. In maize, the MEP and carotenoid pathways are

encoded by 58 genes, which can be considered *a priori* candidates for influencing natural variation for maize grain carotenoid levels (Supplemental Data Set 1). Four of these *a priori* candidate genes—*lycopene epsilon cyclase* (*lcyE*), *β-carotene hydroxylase 1* (*crtRB1*), *zeaxanthin epoxidase 1* (*zep1*) and *ε-ring hydroxylase* (*lut1*)—have been shown to have genome-wide associations with various carotenoid traits in maize grain (HARJES *et al.* 2008; YAN *et al.* 2010; OWENS *et al.* 2014; SUWARNO *et al.* 2015; AZMACH *et al.* 2018). *zep1* encodes zeaxanthin epoxidase, which converts zeaxanthin to violaxanthin via antheraxanthin, and has been found to be associated with levels of zeaxanthin and total β-xanthophylls, with large explanation of phenotypic variance for these traits (OWENS *et al.* 2014; SUWARNO *et al.* 2015). *lut1* encodes a plastid-localized cytochrome P450 that hydroxylates the ε-ring of α-carotene to form zeinoxanthin, and has been detected in association with zeinoxanthin levels and ratios of α-branch compounds (OWENS *et al.* 2014). Lycopene epsilon cyclase (*lcyE*) catalyzes ε-ring cyclization of lycopene and is a key branch point enzyme: alleles with low expression in grain result in higher flux to β-carotene and β-xanthophylls (HARJES *et al.* 2008). Finally, *crtRB1* encodes a non-heme diiron hydroxylase that sequentially hydroxylates β-carotene to produce β-cryptoxanthin and zeaxanthin; weakly expressed alleles result in higher levels of β-carotene at the expense of β-xanthophylls (YAN *et al.* 2010).

Provitamin A breeding efforts in maize focusing on marker-assisted selection of favorable *lcyE* and *crtRB1* alleles have been successfully ongoing for more than a decade at two CGIAR centers, in coordination with HarvestPlus and numerous public and private partners (SALTZMAN *et al.* 2013; DHLIWAYO *et al.* 2014; SUWARNO *et al.* 2014). Maize hybrid and synthetic varieties that accumulate 40-70% of the target provitamin A level have been released, and others with higher levels are in national performance trials (PIXLEY *et al.* 2013, MENKIR *et al.* 2017). However, introgression of favorable alleles of *lcyE* and/or *crtRB1* can have dramatically different effects depending on the genetic background (BABU *et al.* 2013; MENKIR *et al.* 2017; GEBREMESKEL *et al.* 2018), suggesting that further investigation of these and other involved genes may facilitate and expedite the consistent achievement of target provitamin A levels. Furthermore, to simultaneously enhance and balance several other priority carotenoid traits (e.g., lutein, zeaxanthin, and total carotenoids) requires a more comprehensive understanding of the genetics underlying natural variation in maize grain carotenoid content and composition.

While substantial insight into the carotenoid pathway has been obtained from studies in Arabidopsis, its seed are green and photosynthetic whereas those of most major crops, including maize, are non-photosynthetic. Thus, maize grain is both inherently of interest as a key target crop for biofortification efforts and also potentially provides a more suitable model system for carotenoid accumulation in other major crops. In this light, we used the U.S. maize nested association mapping (NAM) panel to dissect, with high power and resolution, the quantitative trait loci (QTL) and underlying genes responsible for natural variation in grain carotenoid levels.

RESULTS

Genetic dissection of carotenoid accumulation in maize grain

We used the U.S. nested association mapping (NAM) panel—25 families, each comprised of 200 recombinant inbred lines (RILs) with B73 as a common parent—to dissect the genetic basis of carotenoid content and composition in maize grain. Seven grain carotenoid compounds were quantified by high-performance liquid chromatography (HPLC) with photodiode array detection. These traits and one summed trait, total carotenoids, had high estimates of line-mean heritability

(0.70 to 0.94, Table 1). Through a joint-linkage (JL) analysis across all 25 NAM families, we identified 117 individual-trait QTL (10 to 23 for each trait; Tables 1 and 2, Supplemental Data Sets 2 and 3) that each explained 0.7 to 40.7% of the phenotypic variance (Supplemental Data Set 4). To dissect the identified QTL at higher resolution, we performed a genome-wide association study (GWAS) using ~27 million HapMap v1 and v2 markers imputed on the ~3,600 NAM RILs (Table 2, Supplemental Data Set 5). A total of 983 marker-trait associations (101 to 142 per trait) were found to have a resample model inclusion probability (RMIP) value ≥ 0.05 (VALDAR *et al.* 2009) (Table 2, Supplemental Data Set 5). Of these, 422 (42.9%) were within a corresponding trait JL interval (Table 2), with 98 of these being attributed to 42 markers associated with two or more traits.

Trait	No. Lines	BLUEs			Heritabilities	
		Median	SD[a]	Range[b]	Estimate	SE[c]
Phytofluene	3,521	0.39	0.54	-0.43 - 3.02	0.72	0.02
α-carotene	3,556	0.91	0.84	-0.70 - 4.77	0.70	0.03
β-carotene	3,538	0.96	0.97	-0.58 - 5.49	0.78	0.03
Zeinoxanthin	3,465	1.44	1.99	-0.95 - 11.19	0.89	0.01
β-cryptoxanthin	3,538	1.61	1.72	-0.56 - 10.04	0.94	0.01
Lutein	3,581	10.68	6.74	-0.72 - 39.47	0.94	0.01
Zeaxanthin	3,565	7.19	7.19	-0.77 - 42.05	0.94	0.01
Total carotenoids	3,581	27.33	13.69	-0.89 - 85.40	0.94	0.01

Table 1. Sample sizes, ranges, and heritabilities for carotenoid traits. Medians and ranges (in μg g^{-1} dry grain) for untransformed best linear unbiased estimators (BLUEs) of eight carotenoid grain traits evaluated in the U.S. maize nested association mapping (NAM) panel, and estimated heritability on a line-mean basis across two years.
[a]SD, Standard deviation of the BLUEs.
[b]Negative BLUE values are a product of the statistical analysis. Specifically, it is possible for BLUEs to equal any value from -∞ to +∞.
[c]SE, Standard error of the heritability estimate.

Trait	Number of JL-QTL	Median size (SD[a]) of α = 0.01 JL-QTL support interval (Mb)	Number of JL-QTL intervals containing *a priori* genes	Number of GWAS-associated variants in JL-QTL intervals[b]	Maximum RMIP
Phytofluene	14	5.05 (27.96)	6	65	0.67
α-carotene	11	2.64 (16.36)	5	44	0.87
β-carotene	10	3.86 (6.72)	8	39	0.80
Zeinoxanthin	12	2.42 (27.79)	8	40	0.86
β-cryptoxanthin	23	3.52 (26.73)	11	75	0.66

Lutein	14	3.17 (19.59)	7	50	0.88
Zeaxanthin	18	4.22 (29.42)	13	56	0.88
Total carotenoids	15	6.21 (20.77)	7	54	0.73
JL-QTL TOTAL	**117**	**3.80 (23.58)**	**65**	**423**	

Table 2. Genetic association results for carotenoid traits. Summary of joint-linkage quantitative trait loci (JL-QTL) and genome-wide association study (GWAS) variants identified for eight carotenoid grain traits evaluated in the US maize nested association mapping (NAM) panel.

[a] SD, Standard deviation.

[b] GWAS variants residing within JL-QTL support intervals for each trait that exhibited a resample model inclusion probability (RMIP) of 0.05 or greater.

Given that individual carotenoid compounds share a biosynthetic pathway (Figure 1), it was not surprising that 80% of overlapping QTL support intervals were significantly pleiotropic (Supplemental Figure 2, Supplemental Data Set 6). When the 117 individual-trait QTL intervals were merged based on physical overlap, 44 unique QTL were obtained, of which 21 impacted multiple traits (Supplemental Data Set 3). We then applied a triangulation approach (DIEPENBROCK *et al.* 2017) integrating JL-QTL effect estimates (Supplemental Data Set 7), GWAS marker genotypes, and RNA-sequencing (RNA-seq) expression abundances at six developing kernel stages in the NAM parents (Supplemental Data Set 8) to identify genes underlying the 44 unique QTL (Supplemental Figures 3 and 5). The rapid decay of linkage disequilibrium (LD) in proximity of GWAS-detected markers (Supplemental Figure 4), in combination with the high resolving power of the NAM panel (WALLACE *et al.* 2014), supported using a search space spanning ±100 kb of those GWAS signals residing within a unique QTL for gene identification. Based on the confluence of strong triangulation correlations for a single gene within these search spaces, 11 genes were identified as underlying a unique QTL (Figure 2). All 11 genes were contained on the list of 58 *a priori* genes known from prior studies in various plants to play roles in IPP synthesis and carotenoid biosynthesis and degradation (Supplemental Data Set 1). Five of these 11 genes were expression QTL (eQTL), in that their expression levels were significantly associated with the JL allelic effect estimates for the QTL at multiple kernel developmental time points (Figures 2 and 3, Supplemental Figure 3). All QTL with >4% PVE were resolved down to an individual gene, with three exceptions: QTL21 for phytofluene (4.68% PVE), QTL24 for α-carotene (6.99%), and QTL34 for α-carotene (5.21% PVE) (Figure 2, Supplemental Table 2). Of the 33 JL-QTL intervals that could not be resolved to an individual gene, 22 were single-trait intervals (Supplemental Table 2).

eQTL[a]	Common Support Interval	RefGen_v4 ID	Annotated Gene Function	Percent phenotypic variance explained for each trait[b]							
				PHYF	ACAR	BCAR	ZEI	BCRY	LUT	ZEA	TOT CAR
No	1	Zm00001d027936	phytoene desaturase (*vp5*)	**1.8**	**1.4**	**1.4**		**1.7**			
No	2	Zm00001d029822	epsilon ring hydroxylase (*lut1*)				**6.1**	**1.5**	**3.0**	**0.9**	**1.0**
No	5	Zm00001d001909	plastid alternative oxidase (*im1*)	**2.5**	**2.1**		**1.2**	**1.9**			
No	7	Zm00001d003513	zeaxanthin epoxidase 1 (*zep1*)	**1.6**			**0.8**	**2.1**	**0.7**	**17.9**	**5.1**
Yes	25	Zm00001d036345	phytoene synthase 1 (*psy1*)	**11.7**		**5.0**	**9.7**	**5.7**	**8.0**	**8.2**	**24.7**
Yes	27	Zm00001d019060	deoxy xylulose synthase 2 (*dxs2*)	**4.8**	**7.7**	**8.1**	**9.7**	**11.3**	**3.5**	**3.7**	**10.6**
Yes	33	Zm00001d011210	lycopene epsilon cyclase (*lcyE*)		**19.6**	**21.4**	**33.6**	**25.8**	**40.7**	**23.0**	
No	35	Zm00001d045383	deoxy xylulose synthase 3 (*dxs3*)			**1.7**	**1.2**	**1.4**	**1.4**	**1.8**	**4.1**

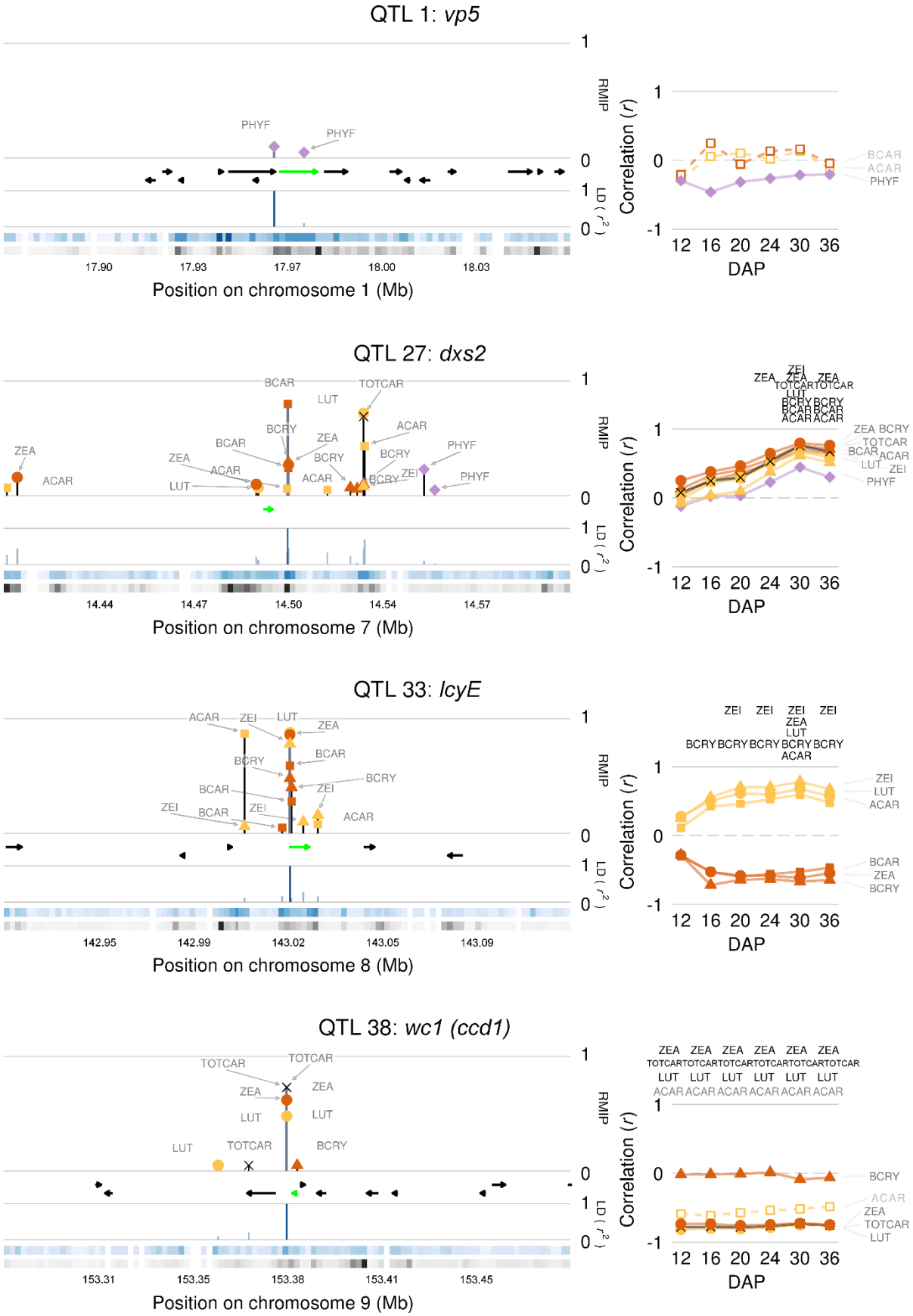
Yes*	38	Zm00001d048373	whitecap 1 (*wc1*) [carotenoid cleavage dioxygenase 1 (*ccd1*)]		**2.3**			**4.1**	**11.1**	**9.7**	**11.1**
No	39	Zm00001d048469	beta carotene hydroxylase 5 (*crtRB5*)							**1.4**	
Yes	43	Zm00001d026056	beta carotene hydroxylase 1 (*crtRB1*)			**8.4**		**4.1**		**3.7**	

Figure 2. Percent phenotypic variance explained (PVE) by joint-linkage quantitative trait loci (JL-QTL). Blue shading corresponds to range of PVEs for JL-QTL, with darker blue indicating higher PVEs.
[a]eQTL (expression QTL) indicates significant correlations between expression values and JL-QTL allelic effect estimates at >2 time points for at least one trait. *Due to the multi-gene array nature of the macrotransposon insertion-derived *whitecap1* locus and its position 1.9 Mb away from the progenitor *ccd1-r* locus (Tan et al. 2017), an FDR-corrected P-value using our ± 100 kb window could not be calculated for the correlation between *ccd1* expression values and effect estimates. While this locus represents a unique case, a strong correlation was still seen between *ccd1* expression (as well as *ccd1* copy number) and QTL38 allelic effect estimates for several traits (Figure 3, Supplemental Table 1).
[b]Trait abbreviations: PHYF, phytofluene; ACAR, α-carotene; BCAR, β-carotene; ZEI, zeinoxanthin; BCRY, β-cryptoxanthin; LUT, lutein; ZEA, zeaxanthin; TOTCAR, total carotenoids.

The role of *a priori* pathway genes

Of the eleven identified *a priori* genes, five encode enzymes that act upstream of the main carotenoid pathway branch point, where the linear carotene lycopene is cyclized to either α- or β-carotene. All five loci were associated with multiple cyclic carotenoids (Figure 2), including provitamin A compounds (β-carotene, β-cryptoxanthin, and α-carotene), and all but phytoene synthase 1 (*psy1*) represent novel associations at the genome-wide level with carotenoid variation in maize grain. Consistent with their sequential positions in the upstream portion of the pathway, all of these loci except for phytoene desaturase (*vp5*) showed positive pleiotropy (i.e., having positively correlated QTL allelic effect estimates between pairs of traits) for the downstream cyclic carotenoids with which they were associated. Two of these five genes, *dxs2* and *dxs3*, encode 1-deoxy-D-xylulose 5-phosphate synthase (DXS), the first enzymatic reaction in the plastid-localized MEP pathway that produces IPP for biosynthesis of carotenoids and other plastidic isoprenoids. *dxs2* and *dxs3* (QTL 27 and 35, respectively) were associated with eight and six traits, with PVEs of 3.5-11.3% and 1.2-4.1%. Of the two genes, *dxs2* was found to be a strong eQTL in the later stages of kernel development (Figure 3). The *dxs2* and *dxs3* homologs were the only MEP pathway genes identified in this study.

Figure 3. Master summaries for selected identified genes. Marker colors indicate pathway branch of the associated trait: yellow for α-branch compounds; orange for β-branch compounds; purple for phytofluene; and black for total carotenoids. Marker shapes correspond to hydroxylation state of the associated trait: squares for α- and β-carotene; triangles for β-cryptoxanthin and zeinoxanthin; circles for lutein and zeaxanthin; diamonds for phytofluene; and an X for total carotenoids.
Left panels: Directional gene models are depicted as black arrows and the identified gene as a green arrow. Lines with trait names above gene models indicate resample model inclusion probability (RMIP) of significant GWAS hits ± 100 kb of the peak RMIP variant. Lines below gene models indicate pairwise linkage disequilibrium (LD; r^2) of each GWAS variant with the peak RMIP variant (dark blue line). The blue ribbon depicts the highest LD, per 200-bp window, to the peak RMIP variant while black ribbons indicate the density of variants tested in GWAS in the 200-bp window ($\log_2$ scale). Darker colors correspond to higher values.
Right panels: Correlations (*r*) between JL-QTL allelic effect estimates and expression of the identified gene across six developing kernel time points. Significant correlations are indicated by trait abbreviations above the respective time point. Traits with both JL and GWAS associations appear in black text to the right of the graph and have solid trend lines and symbols, while those with only JL associations are in gray with dashed trend lines and open symbols.



223

Phytoene synthase (*psy*) catalyzes synthesis of phytoene, the committed biosynthetic intermediate for all carotenoids, from two molecules of GGDP (BUCKNER *et al.* 1996; LI *et al.* 2008). The maize genome contains three *psy* loci, and plants having a functional *psy1* allele accumulate carotenoids in endosperm and embryo whereas those homozygous for a recessive null allele lack carotenoids in endosperm, without affecting the carotenoid content of other tissues (BUCKNER *et al.* 1996, LI *et al.* 2008; reviewed in GILMORE 1997; KOORNNEEF *et al.* 2002). The *psy1* locus serves as a major genetic controller of quantitative variation for carotenoids in maize endosperm, the major site of carotenoid accumulation in grain (ZHU *et al.* 2008; FU *et al.* 2013). *psy1* (QTL 25) had 5.0-11.7% PVE for the seven carotenoids analyzed and also the largest PVE observed in this study for the sum trait of total carotenoids (24.7%). When considering only the 14 families having non-white endosperm parents, indicative of functional *psy1* alleles, this locus still showed PVEs of 1.6-10.8% for the seven measured carotenoids and 22.5% PVE for total carotenoids (Supplemental Data Set 9). *psy1* was an extremely strong eQTL (Supplemental Figure 3), with positive correlations between expression and effect estimates for several traits throughout kernel development, both across all 25 NAM families and across the 14 families having non-white endosperm founders (Supplemental Data Sets 7 and 10).

The next step in the pathway involves the sequential desaturation of phytoene to phytofluene and then ζ-carotene by phytoene desaturase (PDS, encoded by *vp5*; QTL 1), which requires plastoquinone as an electron acceptor (MAYER *et al.* 1990; NORRIS *et al.* 1995; BRAUSEMANN *et al.* 2017). The *vp5* locus was associated with phytofluene (a desaturation intermediate) as well as α-carotene, β-carotene, and β-cryptoxanthin, with PVEs of 1.4-1.8% (Supplemental File 1). Another *a priori* gene (*im1*, QTL 5) encodes a homolog of the Arabidopsis plastid terminal oxidase (PTOX, IMMUTANS) gene (CAROL *et al.* 1999; ALURU *et al.* 2001), which transfers electrons from PQH_2 to molecular oxygen to replenish PQ in the absence of an active photosynthetic electron transport chain, i.e. prior to seedling greening (reviewed in FOUDREE *et al.* 2012). This locus was associated with phytofluene, α-carotene, β-cryptoxanthin, and zeinoxanthin, with PVEs of 1.2-2.5%.

Four biosynthetic genes downstream of the pathway branch point were associated with carotenoid traits, and all had major effects (Figure 2). Lycopene epsilon cyclase (*lcyE*) is the committed step in α-carotene synthesis (CUNNINGHAM *et al.* 1996; BAI *et al.* 2009; CAZZONELLI AND POGSON 2010) and is the major genetic controller of relative flux into the α- or β-carotene pathway branches (HARJES *et al.* 2008). Accordingly, *lcyE* (QTL 33) showed negative pleiotropy (i.e., having negatively correlated QTL allelic effect estimates between pairs of traits) for compounds between the two branches and positive pleiotropy for compounds within the same branch (Supplemental File 1). This locus had the largest PVE observed in this study for each of the six α- and β-branch compounds (PVE = 19.6-40.7%) and—as also reported in HARJES *et al.* (2008)—had no significant impact on total carotenoids (Figure 2).

Subsequent hydroxylations of the ε- and β-rings of α- and β-carotenes are performed by P450s or non-heme dioxygenases, encoded by two and six genes, respectively, in the maize genome. Of these eight genes, three were associated with carotenoid traits in maize grain: *β-carotene hydroxylase 1* (*crtRB1*) and *β-carotene hydroxylase 5 (crtRB5)*, which preferentially convert β-carotene to β-cryptoxanthin and then zeaxanthin, and CYP97C (encoded by *lut1*), a cytochrome P450-type monooxygenase that preferentially hydroxylates the ε-ring of α-carotene to yield zeinoxanthin (TIAN *et al.* 2004; QUINLAN *et al.* 2012). The primary impact of *lut1* (QTL 2) was on the levels of zeinoxanthin (PVE=6.1%) and lutein (PVE=2.9%, Figure 2). *crtrb1* (QTL 43, also known as *hyd3*) had PVEs of up to 8.4% for zeaxanthin, β-carotene, and β-

cryptoxanthin, with negative pleiotropic effects between its substrate, β-carotene, and its products, β-cryptoxanthin and zeaxanthin (Supplemental File 1). *crtRB5* (QTL39, also known as *hyd5*) was identified with PVE of 1.4% for a single trait, zeaxanthin (Figure 2). The subsequent β-branch enzyme, *zeaxanthin epoxidase 1* (*zep1*), had positive pleiotropic effects between β-cryptoxanthin, zeaxanthin, and total carotenoids (Supplemental File 1). *zep1* (QTL 7) primarily had a large effect on its substrate, zeaxanthin, with PVE of 17.9% (Figure 2). Concordant with this large PVE for the highly abundant compound zeaxanthin, this locus also had 5.1% PVE for the sum trait of total carotenoids (Figure 2).

Finally, the maize genome encodes 12 different carotenoid cleavage enzymes involved in ABA and strigolactone synthesis and carotenoid degradation. However, only one, *carotenoid cleavage dioxygenase1* (*ccd1*), was associated with carotenoid levels in this study. CCD1 has been shown to be active toward multiple cyclic and linear carotenoids *in vitro* (VOGEL *et al.* 2008). A single copy of *ccd1* exists in all lines at the progenitor *ccd1-r* locus, and in a limited number of lines containing the dominant *white cap1* (*wc1*) locus a variable number of tandem *ccd1* copies (n = 1 to 11 in the NAM founders) are found within a Tam3L transposon inserted 1.9 Mb proximal to *ccd1-r* (TAN *et al.* 2017). The QTL identified at *wc1* had PVEs of 2.3-11.1% for four compounds, and 11.1% PVE for total carotenoids, making it the second-largest effect QTL for this trait after *psy1* (Figure 2). This locus showed positive pleiotropy for all traits detected (Supplemental File 1) and was a strong eQTL (Figure 3). Correlations of *ccd1* copy numbers in *wc1* alleles from TAN *et al.* (2017) with NAM JL-QTL allelic effect estimates from our study were particularly strong for the most abundant carotenoids, lutein, zeaxanthin, and total carotenoids (r = -0.76 to -0.85), and moderate for α-carotene (-0.56) (Supplemental Table 1). Additionally, there was a strong, positive correlation detected between *ccd1* copy number and *ccd1* expression level ($\log_2$-transformed FPKM) (r = 0.80 to 0.84) across all six developing kernel stages.

Epistasis between identified genes

We tested all pairs of JL-QTL peak markers for epistasis in a joint analysis of all 25 NAM families, including 11 families that each segregated for a recessive allele of *psy1* that conditions extremely low levels of carotenoids in the endosperm. In total, 14 interactions were found to be significant across the 25 families, with PVEs ranging from 0.43 to 6.06%. *psy1* was a partner in only three of these interactions, with *im1*, *lcyE*, and *lut1*, but with PVE <1% in all cases (Figure 4, Supplemental Data Set 11). Only three of the 14 interaction terms had a PVE greater than 1%: *zep1* with *wc1* for zeaxanthin (1.09%), *vp5* with QTL24 for α-carotene (6.06%), and *dxs2* with QTL34 for α-carotene (1.12%).

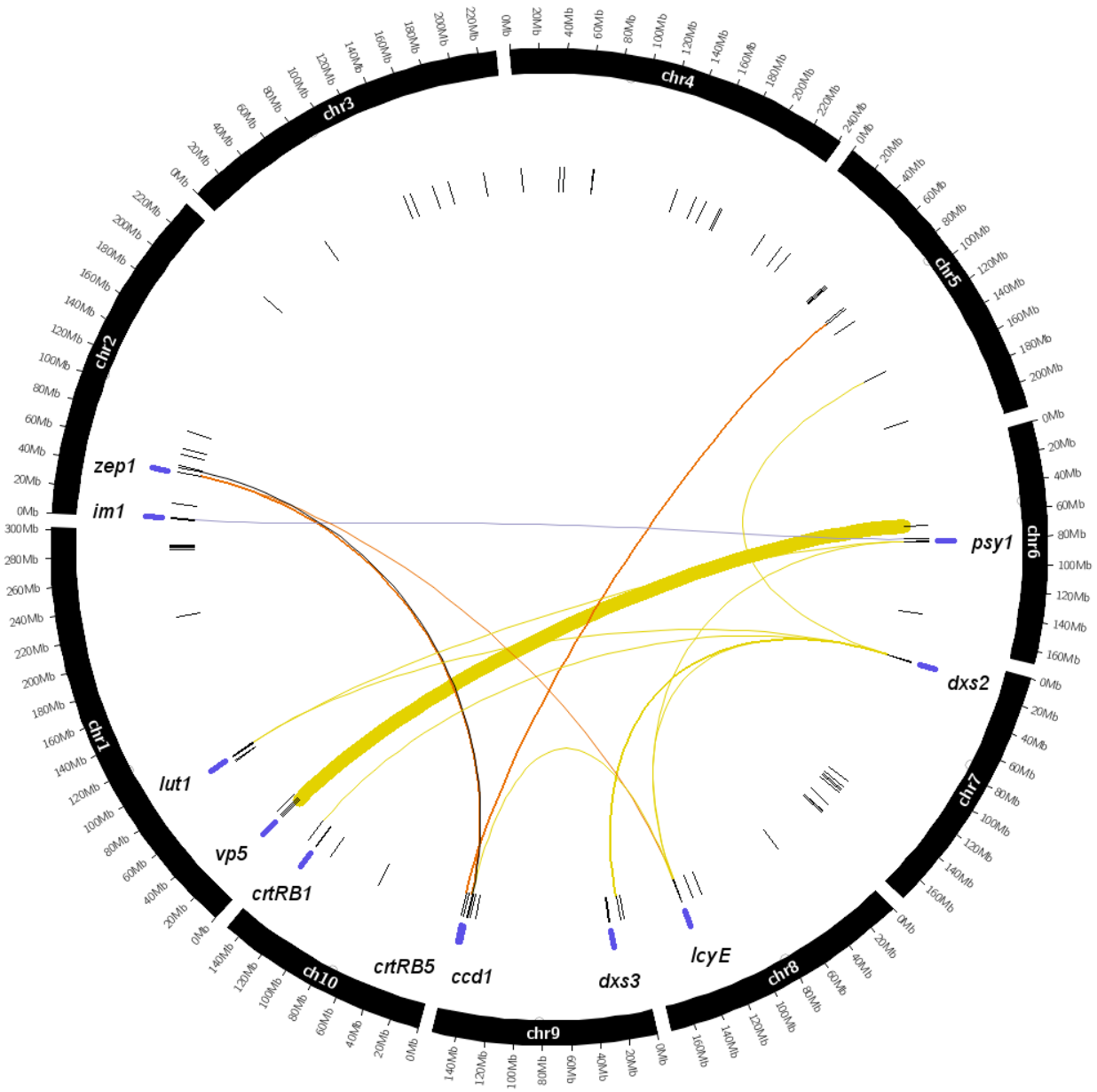


Figure 4. Genome-wide distribution of carotenoid joint-linkage quantitative trait loci (JL-QTL) and their pairwise epistatic interactions in the 25 U.S. NAM families.
From outermost ring to center: Black arcs show chromosomes labeled in 20 Mb increments, with open circles marking centromeres. Identified genes appear as purple capsules, and are labeled with gene names in italics. Radial black lines show positions of peak markers for the 92 individual-trait QTL. Lines linking markers show significant epistatic (additive x additive) interactions, with line thickness proportional to phenotypic variance explained (PVE) by the interaction term; PVEs range from 0.43% to 6.1% in the 25-family models. Links are colored by pathway branch for the interaction: yellow for α-branch compounds; orange for β-branch compounds; purple for phytofluene; and black for total carotenoids.

DISCUSSION

This study provides the most extensive dissection to date of the quantitative genetic basis of maize grain carotenoid levels. The key precursor and core biosynthetic pathway genes

underpinning these traits are now largely elucidated: 11 of 44 carotenoid QTL were resolved to underlying genes (Figure 2, Supplemental Table 2), which together explain 70.33 to 90.87% of all PVE attributed to QTL for each trait (Supplemental Figure 5), with the exception of the acyclic carotenoid phytofluene, with only 48.19% explained. Notably, five of the seven identified genes having PVE ≥4% for any trait were also eQTL (Figure 2), along with *wc1* which exhibited strong correlations between *ccd1* copy number and allelic effect estimates. These six loci also tended to be highly pleiotropic (Supplemental File 1). Taken together, these findings suggest that pleiotropy within the carotenoid pathway is predominantly regulated (directly or indirectly) at the level of gene expression. Epistatic effects were also seen between a small number of the identified genes.

Importantly, this study has identified new and major breeding targets within the precursor pathway for carotenoid biosynthesis. Four of the five identified *a priori* genes residing upstream of lycopene cyclization (*vp5, im1, dxs2,* and *dxs3*) had not previously been associated with natural variation in maize grain carotenoids via GWAS at the genome-wide level. Notably, *dxs2* had the second-largest PVE in this study for two provitamin A carotenoids, β-cryptoxanthin and α-carotene at 11.3% and 7.7%, respectively, second only to that of *psy1* for these traits. These two *dxs* homologs were also associated with natural variation for vitamin-E related traits (tocotrienols and plastochromanol-8; DIEPENBROCK *et al.* 2017), which also utilize IPP in their synthesis. In that vitamin E study, and concordant with the present findings for carotenoids, *dxs2* had larger PVEs and was an eQTL (FDR = 0.05) while *dxs3* was not (Figures 2 and 3, Supplemental Figure 3). Engineering and overexpression studies in Arabidopsis and *E. coli* have shown DXS to be a limiting step within the MEP pathway, and *dxs2* appears to be the major genetic control point in provision of IPP for biosynthesis of carotenoids, tocotrienols, and likely other plastidic isoprenoids in maize grain (HARKER AND BRAMLEY 1999; ESTEVEZ *et al.* 2001). Another gene encoding an MEP pathway enzyme, *seed carotenoid deficient (scd),* was recently characterized to be allelic to *lemon white2-vp12* and encodes a 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HMBPP synthase, or HDS). However, this enzyme is apparently not a rate-limiting step in the MEP pathway as while SCD overexpression complemented a *scd* loss-of-function mutant, the kernel carotenoid levels of the overexpression line were not significantly different from that of wild type (ZHANG et al. 2019).

Genes encoding phytoene desaturase (PDS, *vp5*) and the plastid alternative oxidase (PTOX, *im1*), both of which are necessary for desaturation of phytoene, were identified in this study. PDS sequentially desaturates (oxidizes) phytoene to phytofluene and finally ζ-carotene by introducing a double bond and electrons from each reaction are transferred to plastoquinone (PQ), an essential co-factor for carotenoid biosynthesis (NORRIS *et al.* 1995). In photosynthetic tissues the reduced PQ (i.e., PQH_2) resulting from PDS activity is efficiently re-oxidized by the photosynthetic electron transport chain (ROSSO *et al.* 2006; SHAHBAZI *et al.* 2007; ROSSO *et al.* 2009, reviewed in FOUDREE *et al.* 2012). However, in non-photosynthetic tissues that have a poorly developed electron transport chain—e.g., germinating seedlings and developing maize grain—the plastid alternative oxidase transfers electrons from PQH_2 directly to molecular oxygen to regenerate the PQ needed for additional cycles of desaturation (BEYER *et al.* 1989; MAYER *et al.* 1990; CAROL *et al.* 1999; WU *et al.* 1999, reviewed in RODERMEL 2002; FOUDREE *et al.* 2012). Loss of function of the homologous plastid terminal oxidase in Arabidopsis (IMMUTANS, *IM*) (CAROL *et al.* 1999; WU *et al.* 1999) negatively affects carotenoid synthesis in developing seedlings (WETZEL *et al.* 1994; CAROL *et al.* 1999; ROSSO *et al.* 2009), with photo-oxidized sectors of vegetative tissue in *im* mutants also showing hyper-accumulation of phytoene

(WETZEL *et al.* 1994; WU *et al.* 1999). In our study in maize grain, neither *vp5* (PDS) nor *im1* (PTOX) was an eQTL (Figure 2). This is consistent with previous findings for *IM* in Arabidopsis (ALURU *et al.* 2001), and it has also been noted that *IM* does not directly affect expression of the PDS gene in Arabidopsis (WETZEL AND RODERMEL 1998). While *vp5* and *im1* were largely associated with the same traits (Figure 2), no significant epistatic interaction was observed between them (Figure 4). Of these two, only *im1* showed pleiotropy, with a positive correlation between QTL allelic effects for α-carotene and zeinoxanthin. Further work would be needed to understand the effect of *vp5* and *im1* (PTOX) alleles associated with reduced accumulation of phytofluene (e.g., due to increased conversion to ζ-carotene) on accumulation of downstream, cyclic carotenoids. However, these genes represent another identified target for breeding or metabolic engineering efforts within the precursor pathway for carotenoid biosynthesis. A novel allele of ζ-carotene desaturase (ZDS; *vp9*) in maize was also recently observed to result in increased ζ-carotene and deficiency in downstream carotenoids and ABA (Chen et al. 2017). If one or multiple favorable alleles are identified for this gene, it could be of interest along with *vp5* and *im1*.

Considerable haplotype-level variation is still present at the *psy1* locus in both temperate and tropical maize (SWARTS *et al.* 2017), and an allelic series was seen in the present study (Supplemental Data Set 7). Even within non-white-endosperm maize, previously found to exhibit natural variation for *psy1* (FU *et al.* 2013), the locus was still shown to have large PVEs (Supplemental Data Set 10). This evidence suggests that explicit attention to selecting or fixing a favorable haplotype of *psy1*—along with that of *dxs2*, which also exhibited large PVEs for several traits in this study and positive pleiotropy—may enhance the gains obtained from selection upon other downstream genes (e.g., *lcyE* and *crtRB1*) in provitamin A biofortification of maize, through greater flux of substrates into the carotenoid pathway (FU *et al.* 2013) or other mechanisms (RODRIGUEZ-VILLALON *et al.* 2009).

As mentioned above for *psy1*, these interactions may also have explanations other than that of substrate availability. Namely, in addition to biosynthesis of carotenoids, degradation and the susceptibility to degradation must also be considered. For example, the gene encoding ZEP in Arabidopsis has been identified as a large-effect contributor to natural variation in seed carotenoids (GONZALEZ-JORGE *et al.* 2013), primarily because epoxidation greatly enhances susceptibility to degradation by CCD4, which is plastidic and likely also CCD1, which is cytosolic (AULDRIDGE *et al.* 2006). Combining null alleles of the two genes was found to greatly enhance accumulation of β-carotene and the epoxy-xanthophylls (antheraxanthin, neoxanthin and violaxanthin) two- to three-fold beyond that expected from additive alleles, and also to simultaneously increase zeaxanthin by 40-fold (GONZALEZ-JORGE *et al.* 2016). In our study in maize, *zep1* and *wc1* had an analogous significant interaction for zeaxanthin and total carotenoids, and both were eQTL in the developing kernel stages analyzed. While CCD1 is cytosolic (TAN *et al.* 2003), this enzyme has access to the outer plastid envelope and is thought to have increasing access to plastid-stored carotenoids during desiccation as plastid membranes lose integrity (TAN *et al.* 2017). Monitoring the expression dynamics of these two genes, and also quantifying carotenoid metabolite levels, in the later stages of seed desiccation may provide more information regarding the basis of this interaction. Further investigation of carotenoid retention, due to non-enzymatic factors as well, has been underway for biofortified, provitamin A-dense maize (TALEON *et al.* 2017) and will continue to be important for protection of the accumulated desirable compounds.

Of the 10 to 23 QTL detected per trait, two to five had major effects, from 4 to 41% PVE

depending on the trait (Figure 2). These genes are high-impact targets for genomics-enabled crop improvement strategies. For example, the four major genes to achieve increased provitamin A levels, *lcyE, crtRB1, dxs2,* and *psy1*, explained 75.0% of variation attributed to QTL for β-carotene. Three of the same four genes (*lcyE, dxs2,* and *psy1*) also explained 51.9% of β-cryptoxanthin variation (Figure 2, Supplemental Data Set 4). This finding suggests that simultaneous increases could be achieved for both β-cryptoxanthin and β-carotene through selection at three major genes that each affect several traits. This finding is reinforced by the positive pleiotropy observed at *lcyE, dxs2,* and *psy1* for the two traits. Given that only *crtRB1* and in some cases *lcyE* are currently in use in marker-assisted selection efforts for β-carotene/provitamin A (BABU *et al.* 2013, PRASANNA *et al.* 2020), incorporating this set of additional major-effect genes in selection decisions is likely to provide substantial gains for both β-carotene and β-cryptoxanthin.

While higher heritability was observed for α-carotene in this study than in a smaller inbred diversity panel (OWENS *et al.* 2014), only two large-effect QTL, *lcyE* and *dxs2*, were detected that explained 63.9% of variation attributed to QTL for α-carotene. To increase β-carotene and its derivatives would likely be at the expense of α-carotene, particularly if employing genomics-enabled selection strategies that target *lcyE*. While the monohydroxylated β-cryptoxanthin only has one retinol unit (compared to two for β-carotene), it appears to be more bioavailable than β-carotene (reviewed in PRASANNA *et al.* 2020) and thus should continue to be included in any provitamin A/carotenoid-dense maize breeding program (HARJES *et al.* 2008, DHLIWAYO *et al.* 2014). It was previously recommended that both *lcyE* and *crtRB1* not be selected upon simultaneously, given that a decrease in total carotenoids was seen in lines homozygous for an identified favorable haplotype at both loci (BABU *et al.* 2013). However, neither *lcyE* nor *crtRB1* (separately or in interaction) were associated with total carotenoid levels in this study. Rather, total carotenoid and especially zeaxanthin levels are conditioned upon *zep1* and *ccd1* (Figure 3). *zep1* adds an epoxide group to the β-hydroxy rings of xanthophylls, which greatly decreases their stability in the presence of carotenoid cleavage enzymes. Indeed, disrupting ZEP in Arabidopsis was found to increase zeaxanthin levels by 40-fold, lutein (the most abundant carotenoid in Arabidopsis seed) by 2.2-fold, and total carotenoid levels by 5.7-fold (GONZALEZ-JORGE *et al.* 2016). Thus, decreases in total carotenoids observed upon combining of favorable *lcyE* and *crtRB1* alleles in the same background can likely be circumvented by conditioning on the presence of appropriate *zep1* and *ccd1* alleles. Enabling tandem selection of *lcyE* and *crtRB1* in this manner should afford considerable gains in both total and provitamin A carotenoids, given the large PVEs and allelic effect estimates of *lcyE* observed in this study in families with both temperate and tropical founders (Figure 3, Supplemental Data Set 7).

In addition to provitamin A, multiple other carotenoid traits are of immediate breeding interest (FU *et al.* 2013; KANDIANIS *et al.* 2013). Total carotenoids, a sum trait, is largely controlled by *psy1, wc1, dxs2,* and *zep1*, which together explain 69.0% of phenotypic variation attributed to QTL (Figure 2, Supplemental Data Set 4). Alleles of *dxs2* and *psy1* that increase flux to IPP and phytoene, respectively, could be expected to give rise to higher individual and total carotenoids, while selecting for low *ccd1* copy number (i.e., absence of *wc1*), the only member of the 12-gene cleavage dioxygenase family in maize associated with carotenoid traits in grain, could additionally decrease carotenoid turnover and can act synergistically with biosynthetic enzymes, as previously shown for strong CCD4 and ZEP alleles in Arabidopsis where the introduction of epoxy functional groups by ZEP greatly decreases stability *in vivo*

(GONZALEZ-JORGE *et al.* 2013). Lutein and zeaxanthin, the two most abundant carotenoid compounds in maize grain, are highly correlated with total carotenoids and are of high interest in breeding given their role in eye health. High concentration of zeaxanthin is additionally useful in marketing of biofortified maize due to its dark yellow to orange pigmentation (QUACKENBUSH *et al.* 1961; WEBER 1987), which provides an indicator of a novel, nutritious product (MEENAKSHI *et al.* 2012). The major-effect genes for these compounds are *lcyE*, *wc1* (*ccd1*), and *psy1* (explaining 77.3% of variation attributed to QTL for lutein), and those three genes along with *zep1* for zeaxanthin (explaining 74.1%). Correlations between *ccd1* copy number (VALLABHANENI *et al.* 2010; TAN *et al.* 2017) and QTL effect estimates were highest for lutein and zeaxanthin, with *wc1* explaining 11.1 and 9.7% of phenotypic variation, respectively. These results provide further indication that *zep1* and *wc1 (ccd1*), whose associations with carotenoid trait variation have been recently elucidated (OWENS *et al.* 2014; SUWARNO *et al.* 2015; GONZALEZ-JORGE *et al.* 2016), are key targets for breeding efforts.

Further examination of transcriptional regulation of the genes identified in this study is merited, given that the major genes identified in this study were eQTL (Figure 2). Two transcription factors—a P-box binding factor and R2R3-class MYB—were recently identified, upon transient overexpression, to independently activate the *crtRB1* promoter (resulting in increased *crtRB1* mRNA levels) in maize endosperm as well as the embryo (JIN et al. 2018). The same transcription factors were not found to significantly upregulate *psy1* expression in these tissues, suggesting that other regulators in maize kernels remain unidentified. An R2R3-MYB transcription factor and MADS-box transcription factor were found in monkeyflower and citrus, respectively, to regulate homologs of multiple of the genes identified in this study (STANLEY et al. 2020, LU et al. 2018).

Structural variation in the identified genetic regions also merits further examination. A mutator distance-relative transposable element (TE) in the first intron of maize *dxs2* was recently identified to affect expression of *dxs2,* which was in turn positively correlated with total carotenoid content (FANG et al. 2020). This TE had been present in teosinte; appears to have been subject to selection during maize domestication and improvement; and was found to be nearly fixed in the yellow-grain maize lines examined. While indels and transposable elements have also been identified for *crtRB1* and *lcyE* (YAN et al. 2010, HARJES et al. 2008) and further characterized for use in marker-assisted selection (ZUNJARE et al. 2018, GEBREMESKEL et al. 2017), a more comprehensive examination of structural variation for the 11 genes identified in this study may provide additional insights.

With the genetic resolution of the NAM population, 10 to 23 QTL have captured a large proportion of the phenotypic variation attributed to QTL, even approaching the upper limits of the high heritabilities observed for maize grain carotenoids (Table 1, Supplemental Data Set 4). In summary, genomics-enabled breeding approaches to enhance and balance multiple carotenoid traits in maize grain are now both elucidated and highly feasible. Incorporation of this information into breeding efforts can expedite achievement of the provitamin A target level established by HarvestPlus (BOUIS AND WELCH 2010)—with simultaneous enhancement of other carotenoids beneficial to health—for the consistent delivery of varieties meeting recommended dietary intakes in human populations consuming maize as a staple. The large-effect genes identified in this study are also clear candidates to be assessed as potential controllers of carotenoid variation in seeds of other monocot crop systems.

METHODS

Field Environments and Plant Materials for Genetic Mapping

The design of the maize (*Zea mays*) nested association mapping (NAM) population has been previously described (YU *et al.* 2008; BUCKLER *et al.* 2009; MCMULLEN *et al.* 2009). The experimental field design in 2009 and 2010 for this study—which included the NAM panel, the intermated B73 x Mo17 (IBM) family (LEE *et al.* 2002), and a 281-line inbred diversity panel (FLINT-GARCIA *et al.* 2005)—was conducted as described in CHANDLER *et al.* (2013) and DIEPENBROCK *et al.* (2017). The grain sampling method from at least four ears per plot has additionally been described (DIEPENBROCK et al. 2017).

Carotenoid quantification

Extraction of carotenoids was conducted on ~50 ground kernels per plot, and seven carotenoid compounds—α-carotene, β-carotene, β-cryptoxanthin, lutein, phytofluene, zeaxanthin and zeinoxanthin—as well as the sum trait of total carotenoids were quantified via high-performance liquid chromatography (HPLC) as previously described (OWENS *et al.* 2014), with units of $\mu g\ g^{-1}$ seed. Carotenoids were assessed based on HPLC data passing internal quality control measures that were collected on 9,411 grain samples from 4,871 NAM and 198 IBM RILs, as well as the 850 repeated parental check lines. HPLC-generated measurements of carotenoid compounds and their respective isomers were combined for zeinoxanthin and β-cryptoxanthin, to obtain an overall value for the level of the compound. For technical replicates of the same sample, the mean value of the replicate measurements was recorded for each of the seven carotenoids; i.e., α-carotene, β-carotene, β-cryptoxanthin, lutein, phytofluene, zeaxanthin and zeinoxanthin. Initial total carotenoid values were calculated as the sum of the quantified levels for these seven compounds.

Phenotypic data processing

The HPLC data set was further cleaned to standardize sample genotype names, as associated with experimental field location, and any samples that lacked proper field data were removed. For each NAM family dataset, we retained only samples with genotype assignments belonging to that family or the family's parental genotypes (which were used as checks). Samples from each NAM family and their parental genotypes were categorized as 'yellow- to orange-grain family' (Y) or 'white-grain family' (W) based on the grain endosperm color phenotype of the non-B73 parent of that family. For all samples in the 'W' class for which it was available, the genotype at the psy1 locus, defined as the genomic region spanning chromosome 6 position 82,017,148 to 82,020,879 bp on the AGPv2 maize reference genome, was obtained from GBS SNP marker data downloaded from MaizeGDB (https://www.maizegdb.org/). To eliminate possible sample contamination and select the subset of samples expected to have a functional core carotenoid pathway, samples in the 'W' class were further classified into 'low' and 'high' carotenoid classes using Gaussian decomposition applied to the total carotenoids values for each NAM family × year combination. Gaussian decomposition was performed using the R package 'mclust', specifying two mixture components and one-dimensional variable variance (SCRUCCA *et al.* 2016, R Core Team 2018). Samples with classification uncertainty ≥ 10% were assigned to the 'ambiguous' class. Any samples in the 'W' class that switched carotenoid class assignments between years were removed from further analysis, as well as those that were assigned to the 'low' or 'ambiguous' carotenoid class in both years or that had a homozygous alternate genotype call at the *psy1* locus. Finally, samples in both 'W' and 'Y' classes were subjected to outlier analysis. Specifically, for each NAM family × year combination, a quartile analysis was

performed on the total carotenoid values using the 'boxplot.stats' function in R and any samples with total carotenoids values at least 1.25 * IQR (interquartile range) smaller than the first quartile were marked as low total carotenoids outliers and removed from further analysis.

Following the sample-level filtering, compound measurements were set to missing for any NAM family × year × compound measurement combination that did not have at least 20 samples measured for that compound, and all samples were removed for any NAM family × year combination which did not have at least 40 remaining samples. For the remaining samples, any missing data for a given compound was assigned a value generated by random uniform sampling from the interval (0, min_measured), where min_measured is the lowest (minimum) value measured for that compound in the corresponding NAM family × year group of samples. Following this process, the total carotenoids values were recalculated to include the assigned random uniform values.

As in DIEPENBROCK *et al.* (2017), the IBM RILs were not included in joint-linkage (JL) analysis or the genome-wide association study (GWAS), for they exhibit a differential recombination rate (due to being intermated). However, IBM was still included in the model fitting process to generate best linear unbiased estimators along with the 25 NAM families to provide additional information regarding spatial variation within environments and potential interactions of genotype and environment.

To examine the data for phenotypic outliers, mixed linear model selection was conducted using a custom R script that calls ASRemlR version 4 (BUTLER *et al.* 2017). This model selection process was conducted separately for each of the eight traits. To conduct model selection, a mixed linear model was fit where the grand mean was the only fixed effect, and random effects automatically included in the base model [i.e. without being tested in model selection] were the genotypic effects (family and RIL nested within family) and baseline spatial effects (year and field nested within year). The best random structure was then identified using the Bayesian Information Criterion (BIC; SCHWARZ 1978). The random structures that were tested included all combinations of the following, thus representing one to five terms (or zero, if the base model were to exhibit the most favorable BIC value) to be fit as additional random effects: a laboratory effect (HPLC auto-sampler plate) and certain additional spatial effects (set nested within field within year, block nested within set within field within year, family nested within year, and RIL nested within family within year). The best residual structure was then also identified using BIC, after the best random structure had been identified and included in the mixed linear model. The residual structures that were tested, to account for potential spatial variation across rows and/or columns within each environment, were identity by year; autoregressive for range and identity for row, by field-in-year; identity for range and autoregressive for row, by field-in-year; and autoregressive (first-order, AR1 × AR1) for range and row, by field-in-year. Field-in-year represents a new factor that combines the field name and year, to enable fitting a unique error structure for each of the three fields.

From the final fitted model for each trait, phenotypic outliers with high influence were detected using the DFFITS ('difference in fits') criterion (NETER *et al.* 1996; BELSLEY *et al.* 2005) as previously described in DIEPENBROCK *et al.* (2017), and observations were set to NA if they exceeded a conservative DFFITS threshold previously suggested for this experimental design (HUNG *et al.* 2012). Following outlier removal, the model-fitting process described above was conducted again to estimate best linear unbiased estimators (BLUEs) for the RILs, but with the genotypic effects of family and RIL within family now included as sparse fixed effects rather than random effects. Note that the model was first fit in this step with these genotypic effects as

random, then updated with these effects as sparse fixed. All terms except for the grand mean were then again fitted as random effects to estimate variance components for the calculation of line-mean heritabilities. These heritability estimates were calculated only across the 25 NAM families (HUNG *et al.* 2012), and the delta method was used to obtain standard errors (HOLLAND *et al.* 2003).

The BLUEs generated for each trait were then examined to detect any remaining statistical outliers. Specifically, the Studentized deleted residuals (KUTNER *et al.* 2004) were examined using PROC MIXED in SAS version 9.3 (SAS Institute 2012). These residuals were obtained as in DIEPENBROCK *et al.* (2017), from a parsimonious linear model that contained the grand mean and a single randomly sampled, representative SNP (PZA02014.3) from the original genetic map for the NAM panel (MCMULLEN *et al.* 2009), as fixed effects. The BLUE value of a given RIL for a given trait was set to NA if its corresponding Studentized deleted residual had magnitude greater than the Bonferroni critical value of $t(1 - \alpha/2n; n - p - 1)$. The significance level (α) used in this step was 0.05; n was the sample size of 3,585 RILs; and p was the number of predictors.

The Box-Cox power transformation (BOX AND COX 1964) was then performed separately on BLUEs for each trait as in DIEPENBROCK *et al.* (2017). Briefly, the same parsimonious model used to generate Studentized deleted residuals was also used to identify the most appropriate Box-Cox transformation (with λ tested between -2 and 2, step of 0.05) that corrected for heteroscedasticity and error terms that were not normally distributed. PROC TRANSREG within SAS version 9.3 (SAS Institute 2012) was used to find the optimal λ for each trait (Supplemental Data Set 12) and apply the transformation. Note that the Box-Cox power transformation requires positive input values. Each of the traits had some number of negative BLUE values (ranging from 1 to 283 RILs per trait); these are a reasonable result of the BLUE-fitting process (BURKSCHAT 2009). The lowest possible integer needed to make all values positive for a given trait was added as a constant across that trait vector for all of the RILs before applying the transformation; this constant had a value of 1 for all traits.

Joint linkage analysis

Generation of the 0.1 cM consensus genetic linkage map (14,772 markers) used for joint linkage (JL) analysis, with genotyping by sequencing (GBS) data for ~4,900 NAM RILs as anchors (ELSHIRE *et al.* 2011, GLAUBITZ *et al.* 2014), was previously described (OGUT *et al.* 2015, DIEPENBROCK *et al.* 2017). JL analysis was then conducted as previously described (DIEPENBROCK *et al.* 2017) across the 25 families of the NAM population to map QTL for natural variation in one or more maize grain carotenoid traits. Briefly, joint stepwise regression was implemented using modified source code in TASSEL version 5.2.53 (provided on GitHub), with transformed BLUEs as the response variable and the family main effect forced into the model first as an explanatory variable. The effects of each of the 14,772 markers in the 0.1 cM linkage map, nested within family, were then tested for inclusion in the final model as explanatory variables. The significance threshold for model entry and exit of marker-within-family effects was based on conducting JL analysis on 1000 permutations of transformed BLUEs for each trait and selecting the entry *P*-value thresholds (from a partial F-test) that control the Type I error rate at $\alpha = 0.05$. The permutation-derived entry thresholds are listed in Supplemental Data Set 12. Exit thresholds were set to be twice the value of these empirically derived entry thresholds, so that a marker could not enter and exit the model in the same step.

Upon examining the results of initial JL analysis conducted via the above-described procedure, peak markers in the vicinity of *psy1* for various traits, which is the only locus in this genomic interval expected to control for the presence/absence of endosperm carotenoids in white-grain families, were found to have low minor allele counts exclusively among white-grain families (with fewer than 40 individuals across all 25 families having a genotypic state score greater than zero; zero represents homozygosity for the major allele), resulting in apparent inflation of the phenotypic variance explained by these markers. The RILs that had a genotypic state score greater than zero at the *psy1*-proximal peak marker for a given trait were removed from the data set for all traits (prior to the DFFITs step and BLUE generation), comprising 54 unique RILs removed in total. The analytical pipeline was then re-conducted from the mixed linear model selection step to re-generate JL models, and for use of the results in all downstream analyses. After this additional removal step, it was confirmed that no markers in the vicinity of *psy1* exhibited significant JL signal in any white-grain families. The sample sizes per population in this final data set used in all 25-family analyses are listed in Table 1. The permutation procedure was also applied within the 14 NAM families having parents with non-white endosperm color to enable an additional, separate JL analysis within those families using appropriate thresholds, e.g. due to the reduced sample size.

Some multicollinearity between markers in the consensus map was expected, and indeed for two traits (zeaxanthin and zeinoxanthin), two pairs and one pair of markers present in the final JL model, respectively, had a Pearson's correlation coefficient (r) with magnitude greater than 0.8 between their SNP genotype states. In these cases, the marker with smaller sum of squares within the JL model was removed. A re-scan procedure was then conducted in the vicinity of any of the remaining peak markers for a given trait to test whether the removal of the multicollinear marker(s) meant a shift in the association signal. Specifically, if another marker within the support interval now had a larger sum of squares than the original peak marker, that marker would replace the original peak marker in the model, and this process was repeated (including with re-calculation of the support interval) until a local maximum in the sum of squares was found. The final peak JL markers following re-scan, along with the family term, were then re-fitted to obtain final statistics from the JL analysis of each trait. Allelic effect estimates for each QTL, nested within family, were generated with the 'lm' function (R, *lme4* package) as in DIEPENBROCK *et al.* (2017), with the false discovery rate controlled at 0.05 via the Benjamini-Hochberg procedure (BENJAMINI AND HOCHBERG 1995).

Support intervals ($\alpha = 0.01$) were calculated for the JL-QTL in each final model as previously described (TIAN *et al.* 2011). Logarithm of the odds (LOD) scores were calculated (R, 'logLik' base function). The phenotypic variance explained (PVE) by each joint QTL was calculated using previous methods (LI *et al.* 2011), with some modifications as described in DIEPENBROCK *et al.* (2017) to account for segregation distortion across the families. While transformed data were used throughout the analyses in the present study, including in all steps that required statistical inference, it was also desired to more closely examine the signs and magnitudes of QTL allelic effect estimates on the original trait scale and in directly interpretable units of nutrition. For this single purpose, the final JL model determined using transformed BLUEs was refit with untransformed BLUEs without further model selection or re-scan.

Genome-wide association study

Chromosome-specific residuals for each trait were obtained from the final transformed JL models with the family term and any joint QTL located on the given chromosome removed.

These residuals were used as the response variable in GWAS, whereas the genetic markers tested as explanatory variables consisted of the 26.9 million HapMap v. 1 and 2 variants (SNPs and indels <15 bp), as previously described (WALLACE *et al.* 2014), that were upliftable to RefGen_v4 coordinates. Uplifting of HapMap markers from the B73 RefGen_v2 to RefGen_v4 assembly was conducted by clipping 50 nucleotides from each side of a given marker in its v2 position (101 nucleotides of flanking sequence in total). These were then aligned to the B73 RefGen_v4 assembly using Vmatch (v2.3.0; KURTZ 2019), with the options of -d -p -complete -h1. The resulting alignments were then filtered to keep the highest scoring and unique alignment for each marker. If a marker did not have a high-confidence, unique alignment, it was omitted from the set of upliftable markers.

The upliftable markers were projected onto the NAM RILs using the dense 0.1 cM resolution linkage map, and GWAS was conducted in the NAM-GWAS plugin in TASSEL version 4.1.32 (BRADBURY *et al.* 2007) as previously described (WALLACE *et al.* 2014, DIEPENBROCK *et al.* 2017). Briefly, a forward selection regression procedure was conducted 100 times for each chromosome, with 80% of the RILs from every family sub-sampled each time. For each trait, the model entry threshold was empirically determined by conducting GWAS on 1000 permutations of chromosome-specific residuals for each trait and averaging the results across chromosomes (WALLACE *et al.* 2014) to control the genome-wide Type I error rate at $\alpha = 0.05$ (Supplemental Data Set 12). The significance threshold used for a marker in GWAS was its resample model inclusion probability (RMIP) value, or the proportion of the 100 final GWAS models in which that marker was included (i.e. meeting the model entry threshold). Markers having an RMIP$\geq$ 0.05 were considered in downstream analyses.

RNA sequencing

RNA-seq reads from DIEPENBROCK *et al.* (2017) were downloaded from the National Center for Biotechnology Information Sequence Read Archive (BioProject PRJNA174231) and processed using the same pipeline as described in Hoopes et al. (2018). In brief, read quality was assessed using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (EWELS *et al.* 2016) and then cleaned using Cutadapt (Martin 2011) to remove adaptors and low quality sequences, aligned to AGPv4 of B73 using TopHat2 (Kim *et al.* 2013) with the parameters -i 5 -I 60000 --library-type fr-unstranded, and expression abundances determined using Cufflinks (TRAPNELL *et al.* 2012) in the unstranded mode with a maximum intron length of 60 kb and the AGPv4 annotation. Expression data is available via the MSU Maize Genomics Resource (HOOPES *et al.* 2018; http://maize.plantbiology.msu.edu/index.shtml) via a JBROWSE installation and as a downloadable expression matrix.

FPKM filtering

The gene set was filtered (as in DIEPENBROCK et al. 2017) such that at least one of the kernel developmental samples in at least one sampled founder line had an FPKM greater than 1.0; a total of 30,121 genes remained upon filtering with this criterion. Expression data for genes passing the specified threshold were transformed according to $\log_2(\text{FPKM} + 1)$, where the constant of 1 was added to allow the transformation of '0' values. These $\log_2$-transformed values are herein specified as "gene expression levels" (Supplemental Data Set 8).

Triangulation analysis

The JL support intervals from two or more individual-trait models that were physical

734 overlapping were merged to form common support intervals, as in DIEPENBROCK *et al.* (2017).
735 Physically distinct support intervals detected for a single trait were also retained. Triangulation
736 analyses were conducted as in DIEPENBROCK *et al.* (2017), based on all pairwise Pearson
737 correlations between trait JL-QTL effect estimates; marker genotype state for each significant
738 GWAS marker in the interval for the respective trait(s); and $\log_2$-transformed expression values
739 of genes within ± 100 kb of any of these significant GWAS markers. The search space of 100 kb
740 was selected based on LD decay (Supplemental Figure 4). For those correlations involving one
741 of the five traits with a negative optimal lambda for the Box-Cox transformation (i.e., an inverse
742 power transformation was applied for these traits), the sign of the correlation was reversed in
743 graphical and tabular representations (Figure 3 and Supplemental Figure 3 for master gene
744 summaries, Supplemental Figure 2 and Supplemental Data Set 6 for pleiotropy) to represent the
745 true directionality of the relationship between traits.
746
747 **Epistasis**
748 For the peak markers in the final JL model for each trait, each possible additive x additive
749 pairwise interaction was individually tested for significance in a model containing all marker
750 main effects as in DIEPENBROCK *et al.* (2017). This procedure was conducted both in all 25
751 families and in the 14 families having non-B73 parents with non-white endosperm. The model
752 entry thresholds for these interaction terms was determined (separately for the 25- and 14-family
753 data sets) by modeling 1000 null permutations of transformed trait BLUEs with only additive
754 terms in the model, and selecting the *P*-value approximating a Type I error rate at $\alpha = 0.05$. Final
755 epistatic models were then fit with all marker main effects and any significant interactions. PVE
756 was calculated as described above, except that pairwise genotype scores were collapsed into
757 three classes for interaction terms as previously described (DIEPENBROCK *et al.* 2017).
758 Significant interactions in the 25-family analysis were graphically depicted using the Circos
759 software package (KRZYWINSKI *et al.* 2009) (Figure 4).
760
761 **Pleiotropy**
762 Pleiotropy was assessed as previously described (BUCKLER *et al.* 2009), by applying the JL QTL
763 model for each trait to every other trait. Pearson correlations between the allelic effect estimates
764 for the original trait and the trait to which its model was applied were evaluated for significance
765 at $\alpha = 0.01$ after FDR correction via the Benjamini-Hochberg method. Significant pleiotropic
766 relationships were visualized using the *network* R package (BUTTS 2008; BUTTS 2015)
767 (Supplemental Figure 2). Pleiotropy was also examined within each common support interval
768 (Supplemental File 1) to validate the merging of individual-trait intervals, a step conducted in
769 previous NAM JL analyses (TIAN *et al.* 2011). In this QTL-level analysis, each peak JL marker
770 within the interval was fit for every other trait that had a peak JL marker in the interval.
771
772 **Linkage disequilibrium analysis**
773 The same imputed genotypic data set of 26.9 million segregating markers used in JL-GWAS was
774 used to estimate LD. Specifically, pairwise linkage disequilibrium (LD) of each significant
775 GWAS marker with all other markers within ± 250 kb was estimated through custom Python and
776 R scripts as previously described (WEIR 1996; WALLACE *et al.* 2014; DIEPENBROCK *et al.* 2017).
777 A null distribution was generated by performing the same estimation for 50,000 markers selected
778 at random. LD was examined in both v2 and uplifted v4 coordinates; in the latter case, the small
779 minority of flanking markers that moved outside of the ± 250 kb region upon uplifting were

dropped from the v4 analysis for that given marker.

AUTHOR CONTRIBUTIONS

C.H.D., M.A.G., and D.D.P co-wrote the manuscript; C.H.D., D.C. I., C.B.K., and A.E.L. co-led data analysis; M.M.-L. performed HPLC analyses and metabolite quantifications; B.V., E.G.-C., J.P.H., E.W., and J.C. performed transcriptome analysis; J.P.H. performed uplifting analyses; J.G.W. generated marker data sets; J.G.W. and D.C.I. coded GWAS, GWAS permutation, and figure scripts; J.C. created website/databases; D.C.I. and P.J.B co-led phenotypic data processing; J.B.H. conceptualized and coded spatial model fitting and heritability implementation; P.J.B. modified TASSEL source code and oversaw epistasis calculations; T.R. overall management of NAM population growth; M.M.-H. managed planting, pollination, harvesting, processing of NAM population; B.F.O. and T.T. generated developing kernels for transcriptome analysis; E.S.B. oversaw germplasm, genotyping, and imputation, and advised on mapping analysis; C.R.B. oversaw RNA-seq and transcriptome analysis, managed all informatics; M.A.G. oversaw data analysis, project management, design, coordination; D.D.P. overall project management and coordination, oversaw data and metabolite analyses, biological interpretation.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (DBI-0922493 to D.D.P., C.R.B, E.S.B., and T.R. and DBI-0820619 and IOS-1238014 to E.S.B.), by the USDA-ARS (E.S.B.), by Cornell University startup funds (M.A.G.), and by the University of California, Davis startup funds (C.H.D.). We gratefully acknowledge Arthur Gilmour for expert support in AsREML.

REFERENCES

Abdel-Aal el, S. M., H. Akhtar, K. Zaheer and R. Ali, 2013 Dietary sources of lutein and zeaxanthin carotenoids and their role in eye health. Nutrients 5**:** 1169-1185.

Al-Babili, S., and H. J. Bouwmeester, 2015 Strigolactones, a novel carotenoid-derived plant hormone. Annu Rev Plant Biol 66: 161-186.

Aluru, M. R., H. Bae, D. Wu and S. R. Rodermel, 2001 The Arabidopsis *immutans* mutation affects plastid differentiation and the morphogenesis of white and green sectors in variegated plants. Plant Physiol. 127**:** 67-77.

Azmach, G., A. Menkir, C. Spillane, and M. Gedil 2018 Genetic loci controlling carotenoid biosynthesis in diverse tropical maize lines. G3: Genes|Genomes|Genetics 8**:** 1049-1065.

Babu, R., N. P. Rojas, S. Gao, J. Yan and K. Pixley, 2013 Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. Theor. Appl. Genet. 126**:** 389-399.

Bai, L., E. H. Kim, D. DellaPenna and T. P. Brutnell, 2009 Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. Plant J. 59**:** 588-599.

Barnett, D., E. Garrison, A. Quinlan, M. Strömberg and G. Marth, 2011 BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27**:** 1691-1692.

Beatty, S., M. Boulton, D. Henson, H.-H. Koh and I. J. Murray, 1999 Macular pigment and age related macular degeneration. Br. J. Ophthalmol. 83**:** 867-877.

Berman, J., U. Zorilla-López, V. Medina, G. Farré, G. Sandmann *et al.*, 2017 The Arabidopsis *ORANGE* (*AtOR*) gene promotes carotenoid accumulation in transgenic corn hybrids derived from parental lines with limited carotenoid pools. Plant Cell Rep. 36**:** 933-945.

Beyer, P., M. Mayer and H. Kleinig, 1989 Molecular oxygen and the state of geometric isomerism of intermediates are essential in the carotene desaturation and cyclization reactions in daffodil chromoplasts. Eur. J. Biochem. 184**:** 141-150.

Belsley, D. A., E. Kuh and R. E. Welsch, 2005 *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Hoboken, New Jersey.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. B Met. 57**:** 289-300.

Blessin, C. W., J. D. Brecher, and R. J. Dimler, 1963 Carotenoid of corn and sorghum: V. Distribution of xanthophylls and carotenes in hand-dissected and dry-milled fractions of yellow dent corn. Cereal Chem. 40: 582-586.

Bouis, H. E., and R. M. Welch, 2010 Biofortification--a sustainable agricultural strategy for reducing micronutrient malnutrition in the Global South. Crop Sci. 50**:** S-20-S-32.

Box, G. E. P., and D. R. Cox, 1964 An analysis of transformations. J. Royal Stat. Soc. B Met. 26**:** 211-252.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss *et al.*, 2007 TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23**:** 2633-2635.

Brausemann, A., S. Gemmecker, J. Koschmieder, S. Ghisla, P. Beyer *et al.*, 2017 Structure of phytoene desaturase provides insights into herbicide binding and reaction mechanisms involved in carotene desaturation. Structure 25**:** 1222-1232 e1223.

Brown, P. J., N. Upadyayula, G. S. Mahone, F. Tian, P. J. Bradbury *et al.*, 2011 Distinct genetic architectures for male and female inflorescence traits of maize. PLOS Genet. 7**:** e1002383.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown *et al.*, 2009 The genetic architecture of maize flowering time. Science 325**:** 714-718.

Burdo, B., J. Gray, M. P. Goetting-Minesky, B. Wittler, M. Hunt *et al.*, 2014 The Maize TFome--development of a transcription factor open reading frame collection for functional genomics. Plant J. 80**:** 356-366.

Burkschat, M., 2009 Linear Estimators and Predictors Based on Generalized Order Statistics from Generalized Pareto Distributions. Commun. Stat.-Theor. M. 39**:** 311-326.

Butler, D. G., Cullis, B.R., A. R. Gilmour, Gogel, B.G. and Thompson, R. 2017. ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Butts, C., 2008 Network: a package for managing relational data in R. J. Stat. Softw. 24.

Butts, C., 2015 *Network: Classes for Relational Data*, in *The Statnet Project (http://statnet.org)*.

Carol, P., D. Stevenson, C. Bisanz, J. Breitenbach, G. Sandmann *et al.*, 1999 Mutations in the Arabidopsis gene *IMMUTANS* cause a variegated phenotype by inactivating a chloroplast terminal oxidase associated with phytoene desaturation. Plant Cell 11**:** 57-68.

Cazzonelli, C. I., and B. J. Pogson, 2010 Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci. 15**:** 266-274.

Chandler, K., A. E. Lipka, B. F. Owens, H. Li, E. S. Buckler *et al.*, 2013 Genetic analysis of visually scored orange kernel color in maize. Crop Sci. 53**:** 189-200.

Chen, Y., J. Li, K. Fan, Y. Du, Z. Ren *et al.*, 2017 Mutations in the maize zeta-carotene desaturase gene lead to viviparous kernel. PLOS ONE 12: e0174270.

Chia, J. M., C. Song, P. J. Bradbury, D. Costich, N. de Leon *et al.*, 2012 Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. 44**:** 803-807.

Combs, G. F., 2012 Vitamin A. Vitamins: Fundamental aspects in nutrition and health, 4th Edition, Academic Press, London**:** 93-138.

Cunningham, F. X., Jr., B. J. Pogson, Z. Sun, K. A. McDonald, D. DellaPenna *et al.*, 1996 Functional analysis of the β and ε lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. Plant Cell 8**:** 1613-1626.

Diepenbrock, C. H., C. B. Kandianis, A. E. Lipka, M. Magallanes-Lundback, B. Vaillancourt *et al.*, 2017 Novel loci underlie natural variation in vitamin E levels in maize grain. Plant Cell early online.

Dhliwayo, T., N. Palacios-Rojas, J. Crossa and K. V. Pixley, 2014 Effects of S_1 recurrent selection for provitamin A carotenoid content for three open-pollinated maize cultivars. Crop Sci. 54: 2449-2460.

Diepenbrock, C. H., and M. A. Gore, 2015 Closing the divide between human nutrition and plant breeding. Crop Sci. 55**:** 1437-1448.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6**:** e19379.

Estevez, J. M., A. Cantero, A. Reindl, S. Reichler and P. Leon, 2001 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. J. Biol. Chem. 276**:** 22901-22909.

Ewels, P., M. Magnusson, S. Lundin, and M. Kaller, 2016 MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32**:** 3047-3048.

Fang, H., X. Fu, Y. Wang, J. Xu, H. Feng *et al*., 2020 Genetic basis of kernel nutritional traits during maize domestication and improvement. Plant J. 101**:** 278-292.

Flint-Garcia, S. A., A. C. Thuillet, J. Yu, G. Pressoir, S. M. Romero *et al.*, 2005 Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44**:** 1054-1064.

Foudree, A., A. Putarjunan, S. Kambakam, T. Nolan, J. Fussell *et al.*, 2012 The mechanism of variegation in *immutans* provides insight into chloroplast biogenesis. Front. Plant Sci. 3**:** 260.

Fraser, P. D., and P. M. Bramley, 2004 The biosynthesis and nutritional uses of carotenoids. Prog. Lipid. Res. 43**:** 228-265.

Friso, G., W. Majeran, M. Huang, Q. Sun and K. J. van Wijk, 2010 Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. Plant Physiol. 152**:** 1219-1250.

Fu, Z., Y. Chai, Y. Zhou, X. Yang, M. L. Warburton *et al.*, 2013 Natural variation in the sequence of *PSY1* and frequency of favorable polymorphisms among tropical and temperate maize germplasm. Theor. Appl. Genet. 126**:** 923-935.

Gilmore, A. M., 1997 Mechanistic aspects of xanthophyll cycle-dependent photoprotection in higher plant chloroplasts and leaves. Physiol. Plantarum 99**:** 197-209.

Gilmour, A. R., B. J. Gogel, B. R. Cullis and R. Thompson, 2009 ASReml User Guide: Release 3.0, pp. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9**:** e90346.

Gonzalez-Jorge, S., S. H. Ha, M. Magallanes-Lundback, L. U. Gilliland, A. Zhou *et al.*, 2013 Carotenoid cleavage dioxygenase4 is a negative regulator of beta-carotene content in Arabidopsis seeds. Plant Cell 25**:** 4812-4826.

Gonzalez-Jorge, S., P. Mehrshahi, M. Magallanes-Lundback, A. E. Lipka, R. Angelovici *et al.*, 2016 *ZEAXANTHIN EPOXIDASE* activity potentiates carotenoid degradation in maturing seed. Plant Physiol. 171**:** 1837-1851.

Gore, M. A., J. M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz *et al.*, 2009 A first-generation haplotype map of maize. Science 326**:** 1115-1117.

Graham, R. D., R. M. Welch and H. E. Bouis, 2001 Addressing Micronutrient Malnutrition Through Enhancing the Nutritional Quality of Staple Foods: Principles, Perspectives, and Knowledge Gaps. Adv. Agron. 70**:** 77-142.

Harker, M., and P. M. Bramley, 1999 Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. FEBS Lett. 448**:** 115-119.

Harjes, C. E., T. R. Rocheford, L. Bai, T. P. Brutnell, C. B. Kandianis *et al.*, 2008 Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319**:** 330-333.

Hirsch, C. N., J. M. Foerster, J. M. Johnson, R. S. Sekhon, G. Muttoni *et al.*, 2014 Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26**:** 121-135.

Holland, J. B., W. E. Nyquist and C. T. Cervantes-Martínez, 2003 Estimating and interpreting heritability for plant breeding: an update in *Plant Breeding Reviews*, edited by J. Janick. John Wiley & Sons, Inc.

Hoopes, G. M., J. P. Hamilton, J. C. Wood, E. Esteban, A. Pasha, B. Vaillancourt, N. J. Provart, and C. R. Buell, 2018 An Updated Gene Atlas for Maize Reveals Organ-Specific and Stress-Induced Genes. Plant J. 97**:** 1154-1167.

Hung, H. Y., C. Browne, K. Guill, N. Coles, M. Eller *et al.*, 2012 The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108**:** 490-499.

Jahns, P., and A. R. Holzwarth, 2012 The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. Biochim. Biophys. Acta 1817**:** 182-193.

Jia, K. P., L. Baz, and S. Al-Babili, 2018 From carotenoids to strigolactones. J. Exp. Bot. 69**:** 2189-2204.

Jin, X., C. Bai, L. Bassie, C. Nogareda, I. Romagosa *et al.*, 2019 ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (*Zea mays*) β-carotene hydroxylase 2 gene. New Phytol. 222**:** 793-804.

Kandianis, C. B., R. Stevens, W. Liu, N. Palacios, K. Montgomery *et al.*, 2013 Genetic architecture controlling variation in grain carotenoid composition and concentrations in two maize populations. Theor. Appl. Genet. 126**:** 2879-2895.

Kermode, A. R., 2005 Role of abscisic acid in seed dormancy. J. Plant Growth Regul. 24**:** 319-344.

Khoo, H. E., K. N. Prasad, K. W. Kong, Y. Jiang and A. Ismail, 2011 Carotenoids and their isomers: color pigments in fruits and vegetables. Molecules 16**:** 1710-1738.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14**:** R36.

Koornneef, M., L. Bentsink and H. Hilhorst, 2002 Seed dormancy and germination. Curr. Opin. Plant Biol. 5**:** 33-36.

Krinsky, N. I., J. T. Landrum and R. A. Bone, 2003 Biologic mechanisms of the protective role of lutein and zeaxanthin in the eye. Annu. Rev. Nutr. 23**:** 171-201.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: an information aesthetic for comparative genomics. Genome Res 19**:** 1639-1645.

Kurtz, S.: The Vmatch large scale sequence analysis software. 2019. URL http://www.vmatch.de.

Kutner, M. H., C. J. Nachtsheim, J. Neter and W. Li, 2004 *Applied Linear Statistical Models*. McGraw Hill Irwin, Boston.

Lee, M., N. Sharopova, W. D. Beavis, D. Grant, M. Katt *et al.*, 2002 Expanding the genetic map of maize with the intermated B73 x Mo17 *(IBM) population*. Plant Mol. Biol. 48**:** 453-461.

Li, F., R. Vallabhaneni, J. Yu, T. Rocheford and E. T. Wurtzel, 2008 The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance. Plant Physiol. 147**:** 1334-1346.

Li, H., P. Bradbury, E. Ersoz, E. S. Buckler and J. Wang, 2011 Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS ONE 6**:** e17573.

Li, L., H. Yuan, Y. Zeng and Q. Xu, 2016 Plastids and carotenoid accumulation. Subcell. Biochem. 79**:** 273-293.

Lipka, A. E., M. A. Gore, M. Magallanes-Lundback, A. Mesberg, H. Lin *et al.*, 2013 Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 Genes|Genomes|Genetics 3**:** 1287-1299.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li *et al.*, 2012 GAPIT: genome association and prediction integrated tool. Bioinformatics 28**:** 2397-2399.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger and O. Schabenberger, 2006 SAS® for Mixed Models, Second Edition, SAS Institute Inc., Cary, NC.

Lu, S., Y. Zhang, K. Zhu, W. Yang, J. Ye *et al.*, 2018 The citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. Plant Physiol. 176**:** 2657-2676.

Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17**:** 10-12.

Mayer, M. P., P. Beyer and H. Kleinig, 1990 Quinone compounds are able to replace molecular oxygen as terminal electron acceptor in phytoene desaturation in chromoplasts of *Narcissus pseudonarcissus* L. Eur. J. Biochem. 191**:** 359-363.

McCarty, D. R., 1995 Genetic control and integration of maturation and germination pathways in seed development. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46**:** 71-93.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li *et al.*, 2009 Genetic properties of the maize nested association mapping population. Science 325**:** 737-740.

Meenakshi, J. V., A. Banerji, V. Manyong, K. Tomlins, N. Mittal *et al.*, 2012 Using a discrete choice experiment to elicit the demand for a nutritious food: willingness-to-pay for orange maize in rural Zambia. J. Health Econ. 31**:** 62-71.

Neill, S. J., R. Horgan and A. D. Parry, 1986 The carotenoid and abscisic acid content of viviparous kernels and seedlings of *Zea mays* L. Planta 169**:** 87-96.

Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman, 1996 *Applied linear statistical methods*. Irwin, Chicago.

Norris, S. R., T. R. Barrette and D. DellaPenna, 1995 Genetic dissection of carotenoid synthesis in Arabidopsis defines plastoquinone as an essential component of phytoene desaturation. Plant Cell 7**:** 2139-2149.

Ogut, F., Y. Bian, P. J. Bradbury and J. B. Holland, 2015 Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity 114**:** 552-563.

Owens, B. F., A. E. Lipka, M. Magallanes-Lundback, T. Tiede, C. H. Diepenbrock *et al.*, 2014 A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198**:** 1699–1716.

Palaisa, K. A., 2003 Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15**:** 1795-1806.

Peiffer, J. A., M. C. Romay, M. A. Gore, S. A. Flint-Garcia, Z. Zhang *et al.*, 2014 The genetic architecture of maize height. Genetics 196**:** 1337-1356.

Prasanna, B. M., N. Palacios-Rojas, F. Hossain, V. Muthusamy, A. Menkir *et al.*, 2020 Molecular breeding for nutritionally enriched maize: status and prospects. Front. Genet. 10: 1392.

Quackenbush, F. W., J. G. Firch, W. J. Rabourn, M. McQuistan, E. N. Petzold *et al.*, 1961 Analysis of carotenoids in corn grain. J. Agr. Food Chem. 9**:** 132-135.

Quinlan, R. F., M. Shumskaya, L. M. Bradbury, J. Beltran, C. Ma *et al.*, 2012 Synergistic interactions between carotene ring hydroxylases drive lutein formation in plant carotenoid biosynthesis. Plant Physiol. 160**:** 204-214.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robertson, D. S., 1955 The genetics of vivipary in maize. Genetics 40**:** 745-760.

Rock, C. L., 1997 Carotenoids: biology and treatment. Pharmacol. Ther. 75**:** 185-197.

Rodermel, S., 2002 Arabidopsis variegation mutants. Arabidopsis Book 1**:** e0079.

Rodriguez-Villalon, A., E. Gas and M. Rodriguez-Concepcion, 2009 Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings. Plant J. 60**:** 424-435.

Rosso, D., R. Bode, W. Li, M. Krol, D. Saccon *et al.*, 2009 Photosynthetic redox imbalance governs leaf sectoring in the Arabidopsis thaliana variegation mutants *immutans*, *spotty*, *var1*, and *var2*. Plant Cell 21**:** 3473-3492.

Rosso, D., A. G. Ivanov, A. Fu, J. Geisler-Lee, L. Hendrickson *et al.*, 2006 IMMUTANS does not act as a stress-induced safety valve in the protection of the photosynthetic apparatus of Arabidopsis during steady-state photosynthesis. Plant Physiol. 142**:** 574-585.

Saltzman, A., E. Birol, H. E. Bouis, E. Boy, F. F. De Moura *et al.*, 2013 Biofortification: progress toward a more nourishing future. Glob. Food Sec. 2**:** 9-17.

Schnable, J. C., and M. Freeling, 2011 Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6**:** e17855.

Schwarz, Gideon. Estimating the Dimension of a Model. Ann. Statist. 6 (1978), no. 2, 461--464. doi:10.1214/aos/1176344136.

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, pp. 205-233.

Shahbazi, M., M. Gilbert, A. M. Laboure and M. Kuntz, 2007 Dual role of the plastid terminal oxidase in tomato. Plant Physiol. 145**:** 691-702.

Stahl, W., and H. Sies, 2005 Bioactivity and protective effects of natural carotenoids. Biochim. Biophys. Acta 1740**:** 101-107.

Stanley, L. E., B. Ding, W. Sun, F. Mou, C. Hill *et al.*, 2020 A tetratricopeptide repeat protein regulates carotenoid biosynthesis and chromoplast development in monkeyflowers (*Mimulus*). Plant Cell 32**:** 1536-1555.

Suwarno, W. B., K. V. Pixley, N. Palacios-Rojas, S. M. Kaeppler and R. Babu, 2014 Formation of heterotic groups and understanding genetic effects in a provitamin A biofortified maize breeding program. Crop Sci. 54**:** 14-24.

Suwarno, W. B., K. V. Pixley, N. Palacios-Rojas, S. M. Kaeppler and R. Babu, 2015 Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor. Appl. Genet. 128**:** 851-864.

Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski *et al.*, 2017 Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357**:** 512-515.

Taleon, V., L. Mugode, L. Cabrera-Soto, and N. Palacios-Rojas, 2017 Carotenoid retention in biofortified maize using different post-harvest storage and packaging methods. Food Chem. 232**:** 60-66.

Tan, B.-C., J.-C. Guan, S. Ding, S. Wu, J. W. Saunders *et al.*, 2017 Structure and origin of the *White Cap* locus and its role in evolution of grain color in maize. Genetics 206**:** 135-150.

Tan, B.-C., L. M. Joseph, W.-T. Deng, L. Liu, Q.-B. Li *et al.*, 2003 Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant J. 35**:** 44-56.

Tian, F., P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun *et al.*, 2011 Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 43**:** 159-162.

Tian, L., V. Musetti, J. Kim, M. Magallanes-Lundback and D. DellaPenna, 2004 The *Arabidopsis LUT1* locus encodes a member of the cytochrome P450 family that is required for carotenoid E-ring hydroxylation activity. Proc. Natl. Acad. Sci. 101**:** 402-407.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter, 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7**:** 562-578.

Tuteja, N., 2007 Abscisic acid and abiotic stress signaling. Plant Signal Behav. 2**:** 135-138.

Vallabhaneni, R., L. M. Bradbury and E. T. Wurtzel, 2010 The carotenoid dioxygenase gene family in maize, sorghum, and rice. Arch. Biochem. Biophys. 504**:** 104-111.

Vogel, J. T., B.-C. Tan, D. R. McCarty and H. J. Klee, 2008 The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. J. Biol. Chem. 283**:** 11364-11373.

Wallace, J. G., P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt *et al.*, 2014 Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. 10**:** e1004845.

Wan, C. Y., and T. A. Wilkins, 1994 A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal. Biochem. 223**:** 7-12.

Weber, E. J., 1987 Carotenoids and tocols of corn grain determined by HPLC. J. Am. Oil Chem. Soc. 64**:** 1129-1134.

Weir, B. S., 1996 *Genetic data analysis II*. Sinauer Associates.

Welch, R. M., and R. D. Graham, 2004 Breeding for micronutrients in staple food crops from a human nutrition perspective. J. Exp. Bot. 55**:** 353-364.

West, K. P., Jr., 2002 Extent of vitamin A deficiency among preschool children and women of reproductive age. J. Nutr. 132**:** 28575-28665.

West, K.P., and I. Darnton-Hill. 2008. Vitamin A deficiency. In: R.D. Semba and M. Bloem, editors, Nutrition and health in developing countries, 2nd ed. The Humana Press, Inc., Totowa, NJ. p. 377–433.

Wetzel, C. M., C.-Z. Jiang, L. J. Meehan, D. F. Voytas and S. R. Rodermel, 1994 Nuclear-organelle interactions: the *immutans* variegation mutant of *Arabidopsis* is plastid autonomous and impaired in carotenoid biosynthesis. Plant J. 6**:** 161-175.

Wetzel, C. M., and S. R. Rodermel, 1998 Regulation of phytoene desaturase expression is independent of leaf pigment content in *Arabidopsis thaliana*. Plant Mol. Biol. 37**:** 1045-1053.

Wong, J. C., R. J. Lambert, E. T. Wurtzel and T. R. Rocheford, 2004 QTL and candidate genes phytoene synthase and zeta-carotene desaturase associated with the accumulation of carotenoids in maize. Theor. Appl. Genet. 108**:** 349-359.

Wong, W. L., X. Su, X. Li, C. M. G. Cheung, R. Klein *et al.*, 2014 Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. The Lancet Global Health 2**:** e106-e116.

Wu, D., D. A. Wright, C. Wetzel, D. F. Voytas and S. Rodermel, 1999 The *IMMUTANS* variegation locus of Arabidopsis defines a mitochondrial alternative oxidase homolog that functions during early chloroplast biogenesis. Plant Cell 11**:** 43-55.

Yan, J., C. B. Kandianis, C. E. Harjes, L. Bai, E.-H. Kim *et al.*, 2010 Rare genetic variation at *Zea mays crtRB1* increases beta-carotene in maize grain. Nat. Genet. 42**:** 322-327.

Yilmaz, A., M. Y. Nishiyama, Jr., B. G. Fuentes, G. M. Souza, D. Janies *et al.*, 2009 GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol. 149**:** 171-180.

Young, A. J., 1991 The photoprotective role of carotenoids in higher plants. Physiol. Plantarum 83**:** 702-708.

Yu, J., J. B. Holland, M. D. McMullen and E. S. Buckler, 2008 Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics 178**:** 539-551.

Zhang, L., X. Zhang, X. Wang, J. Xu, M. Wang *et al.*, 2019 SEED CAROTENOID DEFICIENT functions in isoprenoid biosynthesis via the plastid MEP pathway. Plant Physiol. 179: 1723-1738.

Zhu, C., S. Naqvi, J. Breitenbach, G. Sandmann, P. Christou *et al.*, 2008 Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. Proc. Natl. Acad. Sci. 105**:** 18232-18237.

Zunjare, R. U., R. Chhabra, F. Hossain, A. Baveja, V. Muthusamy *et al.*, 2018 Molecular characterization of 5' UTR of the *lycopene epsilon cyclase* (*lcyE*) gene among exotic and indigenous inbreds for its utilization in maize biofortification. 3 Biotech. 8**:** 75.