Received: 26 September 2019 | Accepted: 30 October 2019

DOI: 10.1002/tpg2.20008

The Plant Genome

ORIGINAL RESEARCH

# Natural variation for carotenoids in fresh kernels is controlled by uncommon variants in sweet corn

**Matheus Baseggio[1] | Matthew Murray[2] | Maria Magallanes-Lundback[3] | Nicholas Kaczmar[1] | James Chamness[1] | Edward S. Buckler[1,4,5] | Margaret E. Smith[1] | Dean DellaPenna[3] | William F. Tracy[2] | Michael A. Gore[1]**

[1]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA

[2]Dep. of Agronomy, Univ. of Wisconsin-Madison, Madison, WI 53706, USA

[3]Dep. of Biochemistry and Molecular Biology, Michigan State Univ., East Lansing, MI 48824, USA

[4]Institute for Genomic Diversity, Cornell Univ., Ithaca, NY 14853, USA

[5]US Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

**Correspondence**
Michael A. Gore, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853.
Email: mag87@cornell.edu

**Funding information**
National Science Foundation, Grant/Award Number: IOS-1546657; Cornell University startup funds; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil; USDA-ARS; HATCH, Grant/Award Numbers: 100397, 1010428, 1013637, 1013641, 142 AAC6861

**Abstract**

Sweet corn (*Zea mays* L.) is highly consumed in the United States, but does not make major contributions to the daily intake of carotenoids (provitamin A carotenoids, lutein and zeaxanthin) that would help in the prevention of health complications. A genome-wide association study of seven kernel carotenoids and twelve derivative traits was conducted in a sweet corn inbred line association panel ranging from light to dark yellow in endosperm color to elucidate the genetic basis of carotenoid levels in fresh kernels. In agreement with earlier studies of maize kernels at maturity, we detected an association of β-*carotene hydroxylase* (*crtRB1*) with β-carotene concentration and *lycopene epsilon cyclase* (*lcyE*) with the ratio of flux between the α- and β-carotene branches in the carotenoid biosynthetic pathway. Additionally, we found that 5% or less of the evaluated inbred lines possessing the *shrunken2* (*sh2*) endosperm mutation had the most favorable *lycE* allele or *crtRB1* haplotype for elevating β-branch carotenoids (β-carotene and zeaxanthin) or β-carotene, respectively. Genomic prediction models with genome-wide markers obtained moderately high predictive abilities for the carotenoid traits, especially lutein, and outperformed models with less markers that targeted candidate genes implicated in the synthesis, retention, and/or genetic control of kernel carotenoids. Taken together, our results constitute an important step toward increasing carotenoids in fresh sweet corn kernels.

## 1 | INTRODUCTION

Carotenoids are fat-soluble red, orange, and yellow pigments synthesized by plants that play a critical role in photosynthesis, serving as photoprotectants, antioxidants, and accessory pigments for light harvesting (reviewed in Cuttriss, Cazzonelli, Wurtzel, & Pogson, 2011). When ingested in

**Abbreviations:** AMD, age-related macular degeneration; BLUP, best linear unbiased predictor; FDR, false discovery rate; GBS, genotyping-by-sequencing; GWAS, genome-wide association study; HPLC, high-performance liquid chromatography; IBS, identical-by-state; MLMM, multi-locus mixed-model; QTL, quantitative trait locus; SNP, single-nucleotide polymorphism; WGP, whole-genome prediction.

food by humans, plant-based provitamin A carotenoids, such as α-carotene (one retinyl group), β-carotene (two retinyl groups), and β-cryptoxanthin (one retinyl group), can be converted to retinol, or vitamin A (Combs, 2012). Compromised immune function, blindness, increased risk of maternal mortality and ultimately death from severe infections can result from vitamin A deficiency. In developing countries, more than 127 million preschool-aged children and 7 million pregnant women are vitamin A deficient (West, 2002). Comparatively, less than 1% of the US population (6 yr and older) is considered deficient based on a serum vitamin A concentration of less than <20 μg $dl^{-1}$ (National Center for Environmental Health, 2012). However, even when accounting for consumption of fortified food and dietary supplements, 37% of American adults are below the estimated average requirement (625 μg and 500 μg retinol activity equivalents per day for men and women, respectively) for vitamin A intake (Fulgoni, Keast, Bailey, & Dwyer, 2011).

Lutein and zeaxanthin are the primary pigment compounds found in the retina (macula) of the eye (Bone, Landrum, Hime, Cains, & Zamor, 1993), conferring protection against photooxidative damage to the retina (reviewed in Krinsky, Landrum, & Bone, 2003). Elevated intake of these two non-provitamin A dietary carotenoids has been associated with a reduced risk of progression to late-stage age-related macular degeneration (AMD) (Chew et al., 2014; Wu, Cho, Willett, Sastry, & Schaumberg, 2015), the leading cause of irreversible blindness of elderly adults in the developed world (Congdon et al., 2004; Friedman et al., 2004). The estimated prevalence of AMD in the US population aged 40 and over is 6.5% (Klein et al., 2011), with a forecasted total of almost 18 million early AMD cases in 2050 (Rein et al., 2009). Although no established recommended daily allowance exists for lutein and zeaxanthin, consuming the recommended servings of fruits and vegetables each day would result in a ∼5 mg per day intake of these two carotenoids (reviewed in Mares, 2016). Incidentally, daily intakes of lutein and zeaxanthin in the range of 5–6 mg per day are associated with the lowest AMD rates (reviewed in Mares, 2016), but together lutein and zeaxanthin have an average daily intake of only ∼1.6 mg for an American adult (National Health and Nutrition Examination Survey, 2016).

Given that humans cannot synthesize carotenoids, they must obtain essential and other nutritionally beneficial carotenoids (antioxidants) from their diet to meet minimal nutritional requirements and maintain optimal health (Jerome-Morais, Diamond, & Wright, 2011; Sen & Chakraborty, 2011). The vegetative and seed tissues of fruits and vegetables are important dietary sources of provitamin A carotenoids (α-carotene, β-carotene, and β-cryptoxanthin), lutein, and zeaxanthin (Cazzonelli & Pogson, 2010; Howitt & Pogson, 2006). Extensive variation exists for the content and composition of carotenoids in maize (*Zea mays* L.) grain (Harjes et al., 2008). White endosperm kernels of maize, however, have negligible levels of carotenoids compared to the wide variation of carotenoid levels in kernels with yellow and orange endosperm color (Burt, Grainger, Smid, Shelp, & Lee, 2011; Egesel, Wong, Lambert, & Rocheford, 2003; Harjes et al., 2008; Kurilich & Juvik, 1999). With important implications for human health and nutrition, darker orange endosperm color is a better predictor of high total carotenoids, which is mostly comprised of lutein and zeaxanthin, than provitamin A carotenoid levels (Burt et al., 2011; Harjes et al., 2008). Therefore, endosperm color alone should not be the only consideration when breeding for elevated kernel carotenoid levels (Owens et al., 2014).

**Core Ideas**

- Considerable natural variation found for carotenoids in fresh sweet corn kernels.
- *crtRB1* was associated with concentration of β-carotene—a provitamin A carotenoid.
- *lcyE* controlled flux between the α- and β-carotene branches of carotenoid pathway.
- Favorable variants of *crtRB1* and *lcyE* were uncommon for *sh2* and *su1sh2* lines.
- Whole-genome prediction had moderately high predictive abilities for carotenoids.

In the United States, sweet corn is the third most consumed vegetable (USDA, 2018b); however, consuming 100 g of raw yellow sweet corn (a single medium-sized ear) provides only approximately 1.3 and 1.0% of the recommended daily allowance for vitamin A to adult women (700 retinol activity equivalents per day) and men (900 retinol activity equivalents per day), respectively (Institute of Medicine, 2000; Linus Pauling Institute, 2016; USDA, 2018a). Comparatively, lutein and zeaxanthin are the most abundant carotenoids found in fresh sweet corn kernels (Ibrahim & Juvik, 2009; Kurilich & Juvik, 1999), with 100 g of sweet corn providing on average approximately 11% of the 6 mg per day intake amount of lutein and zeaxanthin associated with a decreased likelihood of AMD (reviewed in Mares, 2016). Extensive fresh kernel carotenoid variation has been observed among a limited number of US sweet corn inbred lines (Ibrahim & Juvik, 2009; Kurilich & Juvik, 1999), but a deeper genetic characterization of the US sweet corn germplasm pool is needed to advance breeding efforts to increase provitamin A carotenoids, lutein, and zeaxanthin levels in fresh kernels.

The carotenoid biosynthetic pathway has been characterized in *Arabidopsis thaliana* and is highly conserved in plants (reviewed in Cuttriss et al., 2011; DellaPenna & Pogson, 2006). Carotenoid synthesis relies on the methylerythritol phosphate pathway to generate two isoprene isomers,

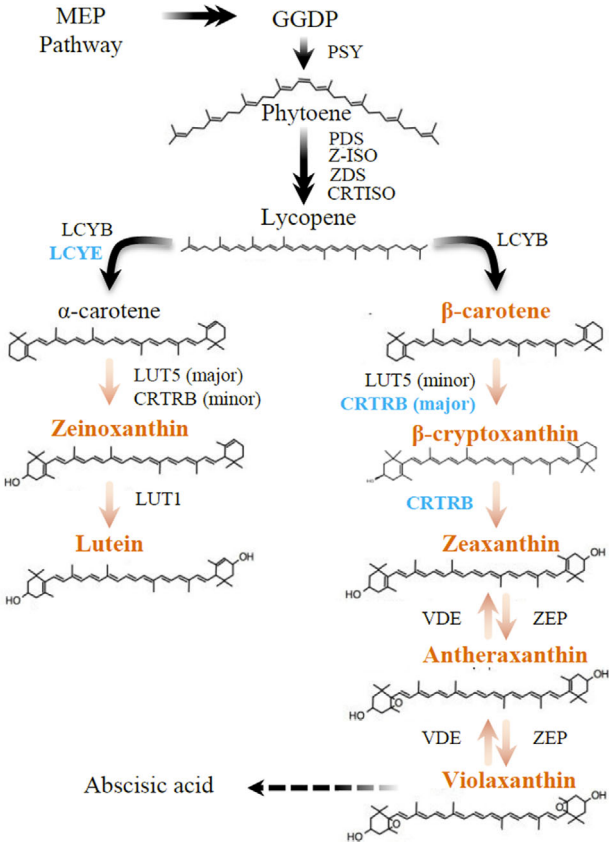


**FIGURE 1** Carotenoid biosynthetic pathway in maize. Multiple maize paralogs are associated with some of the steps shown. When multiple enzymes are indicated for a step in the pathway, the activity level of each enzyme is indicated as major or minor. The seven quantified compounds are shown in bolded orange text. The enzymes in bolded blue text are encoded by candidate genes that are within ±250 kb of the associated single nucleotide polymorphisms (SNPs) identified in our study. Compound abbreviations: GGDP, geranylgeranyl diphosphate; MEP, methylerythritol phosphate. Enzyme abbreviations: CRTRB, β-carotene hydroxylase; CRTISO, carotenoid isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; LUT1, cytochrome P450 ε-ring hydroxylase; LUT5, cytochrome P450 β-ring hydroxylase; PDS, phytoene desaturase; PSY, phytoene synthase; Z-ISO, ε-carotene isomerase; VDE, violaxanthin de-epoxidase; ZEP, zeaxanthin epoxidase; ZDS, ε-carotene desaturase

isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), that are the basis of several reactions resulting in the carotenoid precursor geranylgeranyl diphosphate (GGDP; Figure 1). The synthesis of phytoene by condensing two GGDP molecules via phytoene synthase (PSY) marks the first committed step in carotenoid biosynthesis. Next, two sequential desaturation reactions convert phytoene into lycopene, followed by splitting of the pathway after lycopene into two main branches, α- and β-carotene.

Lycopene β-cyclase (LCYB) introduces a β-ring at both ends of lycopene, resulting in formation of β-carotene. The synthesis of α-carotene necessitates the combined activity of LCYB and lycopene ε-cyclase (LCYE) to add a β-ring to one end of lycopene and an ε-ring to the other. Hydroxylation of one or both β-rings of β-carotene produces β-cryptoxanthin or zeaxanthin, respectively. Zeinoxanthin is formed when the β-ring of α-carotene is hydroxylated, primarily by cytochrome P450 β-ring hydroxylase (LUT5), whereas lutein is produced when the ε-ring of zeinoxanthin is hydroxylated, exclusively by cytochrome P450 ε-ring hydroxylase (LUT1). Xanthophyll cycle carotenoids, zeaxanthin, antheraxanthin, and violaxanthin, play a pivotal role in the dissipation of excess light energy via non-photochemical quenching to minimize photoinhibition (reviewed in Jahns & Holzwarth, 2012). Beyond the xanthophyll cycle, violaxanthin serves as a precursor for synthesis of abscisic acid—a plant hormone that has a central role in the control of seed dormancy and response to abiotic stresses (reviewed in Kermode, 2005; Kundu & Gantait, 2017).

In the last decade, several genes from the carotenoid biosynthetic pathway have been shown to associate with natural variation for carotenoid levels in physiologically mature grain samples from maize (non-sweet corn) association panels at the candidate gene and genome-wide levels. Harjes et al. (2008) showed that four polymorphisms at *lycE* explained 58% of the variation in flux of carotenoids down the α- vs. β-branches of the carotenoid pathway in maize grain. In continuation of this work, Yan et al. (2010) detected an association of β-*carotene hydroxylase 1* (*crtRB1*) with the concentration of β-carotene and its less desirable conversion by hydroxylation to β-cryptoxanthin and zeaxanthin. The authors also showed that the most favorable alleles of *crtRB1*, which had less efficient hydroxylation activity, were rare in frequency in the association panel. Owens et al. (2014) first reported associations of *zeaxanthin epoxidase* (*zep1*) and *cytochrome P450 ε-ring hydroxylase* (*lut1*) with grain carotenoid composition at the genome-wide level. This study also detected a weaker association of carotenoid levels with *1-deoxy-D-xylulose 5-phosphate synthase 2* (*dxs2*) and *cytochrome P450 β-ring hydroxylase* (*lut5*). Lastly, *crtRB3* and *phytoene synthase* (*y1*) have been associated with levels of α-carotene and total carotenoids in maize grain, respectively (Fu et al., 2013; Zhou et al., 2012).

It is unknown whether the carotenoid pathway candidate genes detected in earlier association studies with physiologically mature grain from temperate and tropical maize lines are also critical to the genetic control of carotenoids in developing kernels of sweet corn. This has important implications for the optimization of carotenoid accumulation at fresh eating stage in sweet corn breeding programs, because the content and composition of carotenoids, as well as the expression level of genes from the carotenoid pathway change throughout kernel development (Calvo-Brenes, Fanning, & O'Hare, 2019; Kurilich & Juvik, 1999; Song, Li, He, Chen, & Liu, 2015; Vallabhaneni & Wurtzel, 2009). Not only does the

sweet corn germplasm pool represent a genetically distinct subpopulation of maize that is underrepresented in previously studied maize association populations (Flint-Garcia et al., 2005; Romay et al., 2013), but sweet corn also has one or more mutations in genes from the starch biosynthesis pathway, such as *shrunken2* (*sh2*) and *sugary1* (*su1*), that determine endosperm type (Hannah, Giroux, & Boyer, 1993). The phenotypic consequence of these mutations is a greater accumulation of sugar in the kernel endosperm at the expense of starch relative to wild-type dent corn (Tracy, 1997). Given that carotenoids are primarily synthesized and accumulated in the kernel endosperm (Weber, 1987), it is an open question as to how, if at all, the expression pattern of genes from the carotenoid biosynthetic pathway are altered in a high sucrose endosperm environment. We posit that the genetic architecture of natural variation for carotenoids in fresh sweet corn kernels remains unexplored, thus providing an opportunity for novel elucidation via a genome-wide association study (GWAS) in our newly constructed sweet corn association panel (Baseggio et al., 2019).

Although GWAS is a powerful approach to identify potentially causal genes associated with kernel carotenoid levels, efforts are still needed to translate these findings into innovative breeding strategies to more effectively and efficiently develop nutritionally enhanced sweet corn. Illustrative of such translational genomics work, marker-assisted selection (MAS) for favorable *lcyE* and *crtRB1* alleles with validated effects has been successfully conducted to increase β-carotene content in mature kernels of tropical maize (Babu, Rojas, Gao, Yan, & Pixley, 2013; Zunjare et al., 2018). Suggestive of the potential for increasing carotenoids in sweet corn, Yang, Yan, Wang, Li, and Feng (2018) employed MAS to introgress a favorable allele of *lcyE* identified by Harjes et al. (2008) into four sweet corn lines, resulting in elevated provitamin A carotenoids and total carotenoid levels in fresh kernels. These two genes, however, do not account for all the heritable variation for individual and overall kernel carotenoid levels. In a panel of temperate and tropical maize lines, Owens et al. (2014) showed that a small set of markers, which targeted eight candidate genes underlying QTL associated with carotenoid biosynthesis and retention, was as effective for predicting mature grain carotenoid traits as a genome-wide set of markers. Given the potential for genomic selection to accelerate genetic gain per unit of time (Meuwissen, Hayes, & Goddard, 2001), it is paramount to test whether the QTL-targeted set of eight carotenoid-related candidate genes is sufficient for genomic prediction of fresh kernel carotenoids in the genetically distinct germplasm pool of sweet corn.

In our study, a sweet corn association panel was used to dissect the genetic control of natural variation for carotenoid levels in fresh kernels and construct genomic prediction models to accelerate genetic improvement efforts in fresh sweet corn carotenoid (provitamin A carotenoids, lutein, and zeaxanthin) biofortification breeding programs. To this end, we conducted (i) a GWAS to identify key genes and favorable alleles associated with increased carotenoid levels in fresh sweet corn kernels and (ii) genomic prediction studies to determine the optimal number of genetic markers required to achieve maximal predictive abilities for the potential application of genomic selection for improved fresh kernel carotenoid levels in sweet corn germplasm.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant materials and experimental design

We field-evaluated an association panel of 416 diverse sweet corn inbred lines that samples the allelic diversity of temperate US sweet corn breeding programs (Baseggio et al., 2019) at Cornell University's Musgrave Research Farm in Aurora, NY during the 2014 and 2015 growing seasons. The sweet corn inbred lines included in this panel were homozygous for the following starch-deficient endosperm mutations: *sugary1* (*su1*), *sugary1:sugary enhancer1* (*su1se1*), *shrunken2* (*sh2*), *sugary1:shrunken2* (*su1sh2*), *brittle2* (*bt2*), or *amylose-extender:dull:waxy* (*aeduwx*). Also included in the experiment were 20 non-sweet corn inbred lines and four repeated check sweet corn inbred lines. The association panel was grown in an augmented incomplete block design and two self-pollinated ears were harvested at an immature stage of kernel development from each experimental plot as previously described (Baseggio et al., 2019). Not all plots had harvestable ears because some inbred lines had poor agronomic performance or matured too late. To produce a representative composite kernel sample for each plot, the two hand-harvested ears were immediately frozen in liquid N and shelled, followed by the random sampling and then bulking of frozen kernels. Next, the bulked frozen kernels were quickly transferred to a cryogenic vial and stored at –80°C until grinding. For each bulked sample, we randomly selected 20 to 30 frozen kernels and finely ground them in liquid N. The resultant fine kernel powder of each sample was transferred to a 1.5 ml tube partially submerged in liquid N and stored at –80°C until shipping. All samples were shipped on dry ice to Michigan State University (East Lansing, MI) for extraction and quantification of carotenoids.

### 2.2 | Phenotypic data collection and analysis

Carotenoids were extracted from each ground sample and quantified by high-performance liquid chromatography (HPLC) and fluorometry, with 1 mg $ml^{-1}$ of

β-apo-8′-carotenal as an internal recovery control as previously described (Owens et al., 2014). The seven carotenoid compounds measured in 859 kernel samples from 401 sweet corn, 19 dent, and 4 sweet corn check inbred lines were antheraxanthin, β-carotene, β-cryptoxanthin, lutein, violaxanthin, zeaxanthin, and zeinoxanthin in μg $g^{-1}$ fresh kernel. Given the difficulties associated with identifying and measuring low-abundance carotenoids, lycopene, α-carotene, δ-carotene, and other unidentified carotenes, these compounds were summed to comprise the 'other carotenes' phenotype (less than 8.4% of total carotenoids). Additionally, a series of 11 sums and ratios from Owens et al. (2014) were calculated with minor modifications as follows: zeinoxanthin/lutein, β-cryptoxanthin/zeaxanthin, β-carotene/β-cryptoxanthin, β-carotene/(β-cryptoxanthin+zeaxanthin), α-xanthophylls (sum of lutein and zeinoxanthin), β-xanthophylls (sum of antheraxanthin, β-cryptoxanthin, violaxanthin, and zeaxanthin), β-xanthophylls/α-xanthophylls, total carotenes (sum of β-carotene and other carotenes), total xanthophylls (sum of α- and β-xanthophylls), total carotenes/total xanthophylls, and total carotenoids (sum of the seven carotenoid compounds and other carotenes).

Sweet corn inbred lines homozygous for the recessive null allele of the *y1* gene that encodes *phytoene synthase 1* have carotenoids in the embryo but essentially none in the endosperm with a genetic background-dependent white to pale yellow color (Buckner, Kelson, & Robertson, 1990). Given that, we identified and removed white/pale yellow endosperm lines that had a sample in at least 1 yr with very low levels of total carotenoids quantified by HPLC (total carotenoids: <5.57 μg $g^{-1}$, 2014; <5.98 μg $g^{-1}$, 2015) and simultaneously confirmed this by visual scoring of endosperm color based on images of two immature ears per plot on a green background collected with a hand-held digital camera (Sony DSC-W730, Sony Corporation, Tokyo, Japan) in 2015. Removal of these lines was done to control for the very strong genetic signal at the *y1* locus associated with the Mendelian inherited presence (yellow/orange kernel color) vs. absence (white kernel color) of endosperm carotenoids (Owens et al., 2014) and to allow for the exclusive study of quantitative variation for carotenoid levels. As a result, 345 sweet corn ($n = 322$), dent ($n = 19$), and sweet corn check ($n = 4$) inbred lines with a range from light to dark yellow endosperm color remained.

The levels of zeinoxanthin and compounds comprising 'other carotenes' were below the lower limit of detection for HPLC in two and 27 samples, respectively. Consequently, the values for these samples were approximated within each year by uniform random variables ranging from zero to the minimum detected value for a given carotenoid phenotype as described by Owens et al. (2014). The imputation of missing data with this approach allowed for a maximal sample size to be used in the quantitative analysis of these two phenotypes.

The raw HPLC data of the 19 carotenoid phenotypes from the 345 inbred lines were assessed for normality and screened for significant outliers following the method of Baseggio et al. (2019). Briefly, the Box-Cox power transformation (Box & Cox, 1964) implemented with the 'boxcox' function from the MASS package in R version 3.2.3 (R Core Team, 2015) was used with a simple linear model including genotype, year, set within year, block within set within year, and HPLC autosampler plate within year as fixed effects to select an optimal convenient lambda for each phenotype (Supplemental Table S1). Next, the full mixed linear model 1 of Baseggio et al. (2019) that estimated genetic effects separately from field design effects was fitted for each transformed phenotype in ASReml-R version 3.0 (Gilmour, Gogel, Cullis, & Thompson, 2009). The fitted full model included the following terms: grand mean, check, year, set within year, block within set within year, genotype (non-check line), interaction between genotype and year, HPLC autosampler plate within year, plot grid row within year, plot grid column within year, and residual error following a normal distribution with mean zero and variance $\sigma^2$. All terms were modeled as random effects except for the grand mean and check term. For each phenotype, detected outliers were excluded based on the Studentized deleted residuals (Neter, Kutner, Nachtsheim, & Wasserman, 1996) generated from the fitted mixed linear model.

An iterative mixed linear model fitting procedure was performed with the full model described above in ASReml-R version 3.0 (Gilmour et al., 2009) on each transformed, outlier-screened phenotype as described previously (Baseggio et al., 2019). Briefly, terms fitted as random effects were tested with likelihood ratio tests (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006), and those not significant at $\alpha = .05$ were removed from the model. This resulted in the selection of a final, best fitted model for each phenotype that was then used to generate a best linear unbiased predictor (BLUP) for each genotype. In total, 322 sweet corn inbred lines had BLUPs for at least one of the 19 carotenoid phenotypes, but six inbred lines known to possess *aeduwx* or *bt2* were removed, resulting in a dataset of 316 inbred lines having endosperm mutations (*su1*, *su1se1*, *sh2*, and *su1sh2*) occurring at a higher frequency in the association panel.

Heritability ($\hat{h}_l^2$) on a line-mean basis (Holland, Nyquist, & Cervantes-Martínez, 2003; Hung et al., 2012) was estimated using the variance components from the best fitted model, and standard errors of the estimates were calculated using the delta method (Holland et al., 2003; Lynch & Walsh, 1998). For each pairwise comparison of phenotypes, Pearson's correlation coefficient ($r$) was used to assess the strength of association ($\alpha = .05$) between back-transformed BLUP values (Supplemental Table S2) using the 'cor.test' function in R.

## 2.3 | DNA extraction, sequencing, and genotyping

Of the remaining 316 sweet corn inbred lines with BLUPs, 293 had available raw genotyping-by-sequencing (GBS) (Elshire et al., 2011) data that were generated as described previously (Baseggio et al., 2019). Construction of the SNP marker dataset for the quantitative genetic analysis of the carotenoid phenotypes followed that of Baseggio et al. (2019) with minor modifications. In short, the genotypes of single-nucleotide polymorphisms (SNPs) at 955,690 high confidence loci were called based on the raw GBS data using the production pipeline in TASSEL 5 GBSv1 with the ZeaGBSv2.7 Production TagsOnPhysicalMap file (available at panzea.org, accessed 24 Oct. 2019) in B73 RefGen_v2 coordinates (Glaubitz et al., 2014). To increase the number of lines with both SNP marker and carotenoid data, we merged raw unimputed SNP genotype calls for an additional 16 sweet corn inbred lines from Romay et al. (2013) (ZeaGBSv27_publicSamples_rawGenos_AGPv2–150114.h5, available at panzea.org, accessed 24 Oct. 2019) with those from this study prior to any SNP filtering steps. Initial filtering on the raw unimputed SNP data available for only 309 of the 316 inbred lines consisted of removing SNPs having a minor allele observed in only one line (singletons and doubletons) and retaining only biallelic SNPs with a call rate greater than 10%. Additionally, heterozygous genotype calls with an allele balance score (lowest allele read depth/total read depth) less than 0.3 were set to missing. When two or more samples per line were available, the SNP genotype calls from replicated samples were merged if the identical-by-state (IBS) values from all sample pairwise comparisons exceeded 0.99 as in Romay et al. (2013), and SNP genotypes were set to missing if discordant between replicated samples. If replicated samples had IBS values below this conservative threshold, the sample with the highest call rate was selected to represent the inbred line.

The near complete imputation of missing SNP genotypes was performed using FILLIN (Swarts et al., 2014) with an available set of maize haplotype donors having a window size of 4 kb (available at panzea.org, accessed 24 October 2019). Given that the imputation method is unable to impute all missing genotypes (Swarts et al., 2014), additional filtering was needed for the remaining missing data. Even after imputation, one of the 309 inbred lines still had a SNP call rate less than 40%, thus it was excluded from further analysis. In TASSEL 5 version 20180802, we used a set of filters to further enhance the quality of the imputed dataset by removing SNPs with a call rate less than 70%, a minor allele frequency (MAF) lower than 5%, heterozygosity greater than 10%, an inbreeding coefficient lower than 80%, or a mean read depth greater than 15. The final, complete SNP marker dataset consisted of 172,486 high-quality SNP markers scored on 308 sweet corn inbred lines having a BLUP value for one or more carotenoid phenotypes.

## 2.4 | Genome-wide association study

To conduct a GWAS for each carotenoid phenotype, a univariate mixed linear model was used to test each of the 172,486 SNP markers for association with transformed BLUP values from the 308 inbred lines (Supplemental Table S3) in the GEMMA software version 0.97 (Zhou & Stephens, 2014). The mixed linear model accounted for population stratification and familial relatedness by including principal components (PCs) (Price et al., 2006) and a genomic relationship (kinship) matrix based on VanRaden's method 1 (VanRaden, 2008) calculated in the R package GAPIT version 2017.08.18 (Lipka et al., 2012). The PCs and kinship were calculated based on 12,559 unimputed SNPs—a genome-wide subset of the complete marker dataset—that had a call rate higher than 90%, MAF greater than 5%, heterozygosity less than 10%, inbreeding coefficient greater than 80%, and mean read depth lower than 15. Missing genotypes remaining in both SNP marker datasets were conservatively imputed as heterozygous in GAPIT. The Bayesian information criterion (BIC) (Schwarz, 1978) based on the maximum likelihood estimates of model parameters from GEMMA was used to determine the optimal number of PCs to include as covariates in the mixed linear model. Similarly, the BIC was used to determine whether to also include endosperm mutation type (*su1*, *sh2*, or *su1sh2*), which had been previously scored on the 308 inbred lines by Baseggio et al. (2019), as a covariate in the mixed linear model. This is because *su1* and *sh2* could be strongly associated with endosperm carotenoids (Weber, 1987) as was shown for levels of tocotrienols—a class of tocochromanols mostly found in the endosperm (Grams, Blessin, & Inglett, 1970; Weber, 1987)—in fresh sweet corn kernels from the same association panel (Baseggio et al., 2019).

The likelihood-ratio-based $R^2$ statistic ($R^2_{\mathrm{LR}}$) of Sun et al. (2010) was used to approximate the amount of phenotypic variation explained by a mixed linear model with or without a significant SNP detected in GWAS. The $R^2_{\mathrm{LR}}$ value of each model was calculated with the maximum log-likelihood of the model of interest fitted in GEMMA compared to the maximum log-likelihood of an intercept-only model fitted with the 'lm' function in R. For each phenotype, *P*-values (Wald test) of SNPs tested in GEMMA were adjusted to control the false-discovery rate (FDR) at a level of 5% with the Benjamini–Hochberg multiple test correction (Benjamini & Hochberg, 1995) available in the 'p.adjust' function of R version 3.2.3 (R Core Team, 2015). To identify candidate genes, the search interval was limited to ±250 kb of the physical position of

SNP markers significantly associated with a carotenoid phenotype. This interval size considered the distance at which genome-wide linkage disequilibrium decays to nominal levels (mean $r^2 \leq 0.05$) in this association panel and potential of distant regulatory elements (Baseggio et al., 2019).

We implemented the multi-locus mixed-model (MLMM) approach of Segura et al. (2012) to control for the influence of major-effect loci on an individual chromosome basis as described previously (Lipka et al., 2013). The extended BIC (Chen & Chen, 2008) was used in the selection of the optimal model. The control of major-effect loci was also assessed by reconducting GWAS with MLMM-selected SNPs included as covariates in the mixed linear model of GEMMA.

## 2.5 | Carotenoid prediction

The prospect of genomic selection for breeding sweet corn with increased carotenoids was assessed in the 308 inbred lines using a single kernel genomic best linear unbiased prediction (GBLUP) model (VanRaden, 2008; Zhang, Todhunter, Buckler, & Van Vleck, 2007). In the R package GAPIT version 2017.08.18 (Lipka et al., 2012), method 1 from VanRaden (2008) was used to calculate genomic relationship matrices derived from three different SNP datasets varying in the number of markers: carotenoid QTL-targeted, pathway-level, and genome-wide. The carotenoid QTL-targeted set consisted of 628 SNPs within ±250 kb of eight a priori identified candidate genes underlying QTL associated with carotenoid biosynthesis and retention, whereas the pathway-level set had 4689 SNPs within ±250 kb of 60 a priori candidate genes (including the eight genes from the QTL-targeted set) involved in the biosynthesis and cleavage of carotenoids (Owens et al., 2014; Supplemental Table S4). The 172,486 high-quality SNP markers comprised the genome-wide set. These three genomic relationship matrices were used individually as a random effect for prediction of carotenoid phenotypes with the function 'emmreml' (single kernel) in the EMMREML R package (Akdemir & Okeke, 2015).

A five-fold cross-validation approach was used to estimate the predictive ability of a model for each carotenoid phenotype by calculating the Pearson's correlation between observed BLUP and genomic estimated breeding values as described in Baseggio et al. (2019). The predictive ability of each model was based on a mean of correlations from 50 replicates of the five-fold cross-validation scheme. Each fold consisted of genotype frequencies for endosperm mutants (*su1*, *sh2*, and *su1sh2*) that were representative of the association panel, and the identical cross-validation folds were used across different models. Similar to genomic prediction of tocochromanol traits (Baseggio et al., 2019), endosperm mutation type (*su1*, *sh2*, or *su1sh2*) was evaluated as a covariate in prediction models.

## 2.6 | Data availability

All raw GBS data are available from the National Center of Biotechnology Information Sequence Read Archive under accession SRP154923 and in BioProject under accession PRJNA482446. The back-transformed and transformed BLUP values of the 19 carotenoid phenotypes are provided in Supplemental Tables S2 and S3, respectively. The FILLIN imputed SNP genotype calls for the 308 inbred lines are available from the Dryad Digital Repository (https://datadryad.org/stash/dataset/doi:10.5061/dryad.vq83bk3p2).

# 3 | RESULTS

## 3.1 | Phenotypic variation

We conducted a quantitative assessment of carotenoid levels in fresh (immature stage) kernels harvested from an association panel of 308 sweet corn inbred lines with endosperm color ranging from light to dark yellow. The measurement of carotenoids by HPLC revealed that lutein and zeaxanthin represented about 65% of total carotenoids in the kernel, while the other five carotenoid phenotypes individually accounted for less than 10% of the total (Table 1). The two specifically measured compounds with provitamin A activity, β-carotene and β-cryptoxanthin, had similar concentrations, and when summed only represented approximately 6% of total carotenoids. When separating inbred lines according to their endosperm mutation type, *su1*, *sh2*, or *su1sh2*, three (antheraxanthin, β-cryptoxanthin, and lutein) of the seven individual compounds had an average amount shown to be at a significantly ($P < .05$) greater level in the *sh2* ($n = 46$) group than in the *su1* ($n = 245$) group (Table 2).

With the exception of β-carotene and β-cryptoxanthin ($r = 0.17$), the BLUP values of each carotenoid compound had a Pearson's correlation stronger than 0.5 with those of its immediate precursor in the carotenoid pathway (Supplemental Figure S1). Correlations between β-carotene and other compounds were very weak ($r = –0.05$ to 0.17). In contrast, β-cryptoxanthin had relatively much stronger correlations with all other xanthophyll compounds ($r = 0.55$ to 0.76) but violaxanthin ($r = 0.18$). The estimates of heritability on a line-mean basis for the 19 carotenoid compound, sum, and ratio traits ranged from 0.76 for violaxanthin to 0.93 for lutein and the ratio of α- to β-xanthophylls, with an average of 0.87. Such high heritability estimates suggest that these phenotypes

**TABLE 1** Means and ranges for back-transformed best linear unbiased predictors (BLUPs) of 19 fresh kernel carotenoid traits evaluated in the sweet corn association panel and estimated heritability ($\hat{h}_l^2$) on a line-mean basis across 2 yr

| Trait | Lines | BLUP | | | Heritability | |
|---|---|---|---|---|---|---|
| | | Mean | SD[a] | Range | Estimate | SE[b] |
| | | μg g$^{-1}$ fresh wt. | | | | |
| Antheraxanthin | 308 | 1.22 | 0.30 | 0.39–2.07 | 0.82 | 0.02 |
| β-Carotene | 308 | 0.54 | 0.35 | 0.16–2.83 | 0.90 | 0.01 |
| β-Cryptoxanthin | 308 | 0.47 | 0.29 | 0.11–2.29 | 0.91 | 0.01 |
| Lutein | 308 | 5.82 | 3.04 | 0.64–19.39 | 0.93 | 0.01 |
| Violaxanthin | 308 | 0.99 | 0.21 | 0.56–2.47 | 0.76 | 0.03 |
| Zeaxanthin | 308 | 4.84 | 1.75 | 1.62–10.71 | 0.87 | 0.02 |
| Zeinoxanthin | 308 | 1.20 | 1.04 | 0.06–8.88 | 0.91 | 0.01 |
| Other carotenes | 307 | 1.38 | 0.73 | 0.29–4.94 | 0.80 | 0.02 |
| α-Xanthophylls | 308 | 7.10 | 3.86 | 0.73–28.23 | 0.93 | 0.01 |
| β-Xanthophylls | 308 | 7.46 | 2.04 | 2.99–13.83 | 0.89 | 0.01 |
| Total xanthophylls | 308 | 14.43 | 4.73 | 5.54–38.23 | 0.91 | 0.01 |
| Total carotenes | 307 | 1.97 | 0.89 | 0.76–6.01 | 0.85 | 0.02 |
| Total carotenoids | 308 | 16.48 | 5.28 | 7.40–43.95 | 0.91 | 0.01 |
| β-Carotene/β-cryptoxanthin | 308 | 1.41 | 1.11 | 0.34–8.17 | 0.88 | 0.01 |
| β-Carotene/(β-cryptoxanthin+zeaxanthin) | 308 | 0.12 | 0.13 | 0.03–0.88 | 0.92 | 0.01 |
| β-Cryptoxanthin/zeaxanthin | 308 | 0.09 | 0.03 | 0.03–0.22 | 0.84 | 0.02 |
| Zeinoxanthin/lutein | 308 | 0.21 | 0.12 | 0.04–0.80 | 0.87 | 0.02 |
| β-/α-Xanthophylls | 307 | 1.35 | 0.84 | 0.32–6.29 | 0.93 | 0.01 |
| Total carotenes/total xanthophylls | 308 | 0.14 | 0.06 | 0.05–0.45 | 0.80 | 0.02 |

[a]Standard deviation of the BLUPs.
[b]Standard error of the heritabilities.

would be amenable to genetic dissection and prediction in this sweet corn association panel.

## 3.2 | Genome-wide association study

The association panel of 308 sweet corn inbred lines having yellow endosperm kernels at the fresh-eating stage, which had been scored with 172,486 genome-wide SNP markers, was used to elucidate the genetic basis of natural variation for carotenoids in fresh kernels. Through the implementation of a univariate mixed linear model that accounted for population structure, relatedness, and type of endosperm mutation, we identified 108 unique SNPs that were significantly associated with one to four phenotypes at a genome-wide FDR of 5%. The 108 SNPs were distributed across seven chromosomes, with the vast majority (92.59%) located on chromosomes 2, 8, and 10 (Supplemental Figure S2).

The most significant association was identified for the ratio of β-carotene to β-cryptoxanthin+zeaxanthin on chromosome 10 (Figure 2a). The peak SNP locus (S10_135801334; *P*-value $1.11 \times 10^{-13}$) for this association signal was located within the open reading frame (ORF) of a gene that encodes *GRAS-transcription factor 22* (*grass22*, GRMZM2G173429), but ~255 kb away from β-*carotene hydroxylase 1* (*crtRB1,* GRMZM2G152135)—a gene encoding a nonheme dioxygenase that hydroxylates β-rings of carotenoids. This SNP was also the peak association for β-carotene (*P*-value $2.04 \times 10^{-11}$) and the ratio of β-carotene to β-cryptoxanthin (*P*-value $3.03 \times 10^{-11}$), while S10_135683780 had the strongest association with violaxanthin (*P*-value $7.94 \times 10^{-7}$). In total, 61 SNPs spanning a 5.09-Mb interval (133.7–138.8 Mb) on chromosome 10 were significantly associated with one or more of these four phenotypes (Supplemental Table S5), with 38 of the 61 associated SNPs (*P*-values $3.00 \times 10^{-12}$ to $1.13 \times 10^{-5}$) located ±250 kb of the ORF for *crtRB1*.

We used a chromosome-wide multi-locus mixed-model (MLMM) procedure to better resolve the complex of association signals within the 5.09-Mb region on chromosome 10. Each optimal model for β-carotene, β-carotene/(β-cryptoxanthin+zeaxanthin), and β-carotene/β-cryptoxanthin included the peak SNP S10_135801334 (Supplemental Table S6). Additionally, the optimal models obtained for β-carotene and β-carotene/(β-cryptoxanthin+zeaxanthin) selected a second SNP (S10_136086332) that was located ~26 kb away from *crtRB1* and in very weak linkage disequilibrium ($r^2 = 0.11$) with S10_135801334. Indicative of relatively weaker significant associations, no SNPs were

**TABLE 2** Back-transformed estimated effects of endosperm mutation type for 19 fresh kernel carotenoid traits evaluated in the sweet corn association panel

| Trait | *su1*[a] | *sh2* | *su1sh2* | *P*-value[b] |
|---|---|---|---|---|
| | μg $g^{-1}$ fresh wt. | | | |
| Antheraxanthin | 1.15b[c] | 1.32a | 1.25ab | 0.003* |
| β-Carotene | 0.47 | 0.48 | 0.41 | 0.434 |
| β-Cryptoxanthin | 0.38b | 0.51a | 0.48ab | 0.002* |
| Lutein | 4.52b | 8.57a | 5.90b | <0.0001* |
| Violaxanthin | 0.97 | 0.98 | 1.03 | 0.419 |
| Zeaxanthin | 4.64 | 4.91 | 4.83 | 0.574 |
| Zeinoxanthin | 0.82 | 1.11 | 1.11 | 0.036* |
| Other carotenes | 1.19b | 1.91a | 1.57a | <0.0001* |
| α-Xanthophylls | 5.49b | 10.09a | 7.19ab | <0.0001* |
| β-Xanthophylls | 7.07 | 7.63 | 7.62 | 0.164 |
| Total xanthophylls | 13.00b | 17.33a | 15.40ab | <0.0001* |
| Total carotenes | 1.76b | 2.58a | 2.07ab | <0.0001* |
| Total carotenoids | 14.89b | 20.03a | 17.57ab | <0.0001* |
| β-Carotene/β-cryptoxanthin | 1.24a | 0.96b | 0.87b | 0.002* |
| β-Carotene/(β-cryptoxanthin+zeaxanthin) | 0.10 | 0.09 | 0.08 | 0.299 |
| β-Cryptoxanthin/zeaxanthin | 0.08b | 0.10a | 0.09ab | <0.0001* |
| Zeinoxanthin/lutein | 0.20a | 0.16b | 0.21ab | 0.028* |
| β-/α-Xanthophylls | 1.30a | 0.76b | 0.92b | <0.0001* |
| Total carotenes/total xanthophylls | 0.14 | 0.14 | 0.13 | 0.621 |

[a]The *su1* group includes both *su1Se1* and *su1se1* lines because marker genotypes were not informative for unambiguously distinguishing alleles at the *se1* locus.
[b]*P*-value from one-way ANOVA *F*-test for the endosperm mutation type effect. *P*-values with asterisk (*) indicate a statistically significant difference between two or more endosperm mutation type groups ($P < .05$).
[c]Sweet corn lines grouped by endosperm mutation type having labels with a common letter are not significantly different according to the Tukey-Kramer honest significant difference test ($P < .05$). The test was only performed for traits that had a significant *F*-test.

selected by the MLMM for violaxanthin. When GWAS was reconducted with either one or two MLMM-selected SNPs, depending on the phenotype, included as covariates in the mixed linear model for β-carotene and its two derived phenotypes, all other signals at this 5.09-Mb segment and elsewhere on chromosome 10 were no longer significant at a genome-wide FDR of 5% (Figure 2b). Additionally, 24 SNPs on chromosomes 1, 2, 3, and 8 that were associated with β-carotene/β-cryptoxanthin and/or β-carotene/(β-cryptoxanthin+zeaxanthin) were no longer significant. Conversely, only a single SNP (S6_58455321) from within the pericentromeric region of chromosome 6 remained significantly associated (*P*-value $1.83 \times 10^{-7}$) with β-carotene (Supplemental Figure S3).

When considering the two MLMM-selected SNPs at the haplotype level, the most favorable (TT) of the four observed haplotypes for increasing β-carotene concentration had an average effect estimate (0.81 μg $g^{-1}$ fresh weight) that was twofold greater than the least favorable GC haplotype (Supplemental Table S7). Not only was the most favorable haplotype found to exist at very low frequency in the association panel, but it also was not equally distributed among the endosperm mutation type groups. Only 13 *su1* and two *sh2* lines had the TT haplotype, whereas none of the *su1sh2* lines possessed this haplotype. In contrast, the least favorable haplotype had the highest occurrence in the panel, with more than 70% of the lines in each endosperm mutation type having the GC haplotype. Indicative of only informativeness for the concentration of β-carotene, individually these two SNPs linked to *crtRB1* were not significantly associated with total carotenoids (Supplemental Figure S2), and both the TT and GC haplotypes had nearly the same average effect estimate for total carotenoids (Supplemental Table S7).

The *lcyE* gene (GRMZM2G012966) on chromosome 8 had a SNP (S8_138888278) within its ORF that significantly associated with the ratio of β- to α-xanthophylls (Figure 3a; *P*-value $1.01 \times 10^{-12}$). The *lcyE* gene encodes lycopene ε-cyclase, which determines whether α- or β-carotene is produced. Thus, *lcyE* controls flux down the α- versus β-branches of the carotenoid pathway (Cunningham et al., 1996). An additional nine SNPs spanning a 3.69-Mb interval on chromosome 8, including two SNPs located within the ORF of *lcyE* (S8_138888328 and S8_138888990; *P*-values $7.32 \times 10^{-9}$ and $5.41 \times 10^{-8}$, respectively), as well as three SNPs from

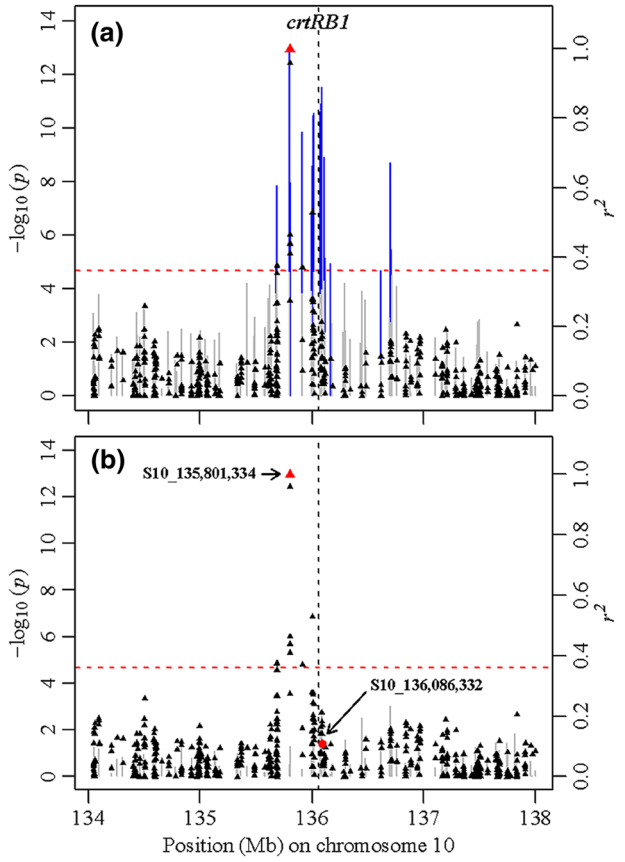


**FIGURE 2** Genome-wide association study for the ratio of β-carotene to the sum of β-cryptoxanthin and zeaxanthin [β-carotene/(β-cryptoxanthin+zeaxanthin)] in fresh kernels of sweet corn. (a) Scatter plot of association results from a mixed linear model analysis and linkage disequilibrium estimates ($r^2$). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNPs) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the $r^2$ values of each SNP relative to the peak SNP (indicated as a red triangle) at 135,801,334 bp (B73 RefGen_v2) on chromosome 10. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical dashed line indicates the genomic position of the β-*carotene hydroxylase 1* (*crtRB1*) gene. (b) Scatter plot of association results from a conditional mixed linear model analysis and linkage disequilibrium estimates ($r^2$). The SNPs (S10_135801334, red triangle; and S10_136086332, red circle) from the optimal multi-locus mixed-model were included as covariates in the mixed linear model to control for the *crtRB1* effect

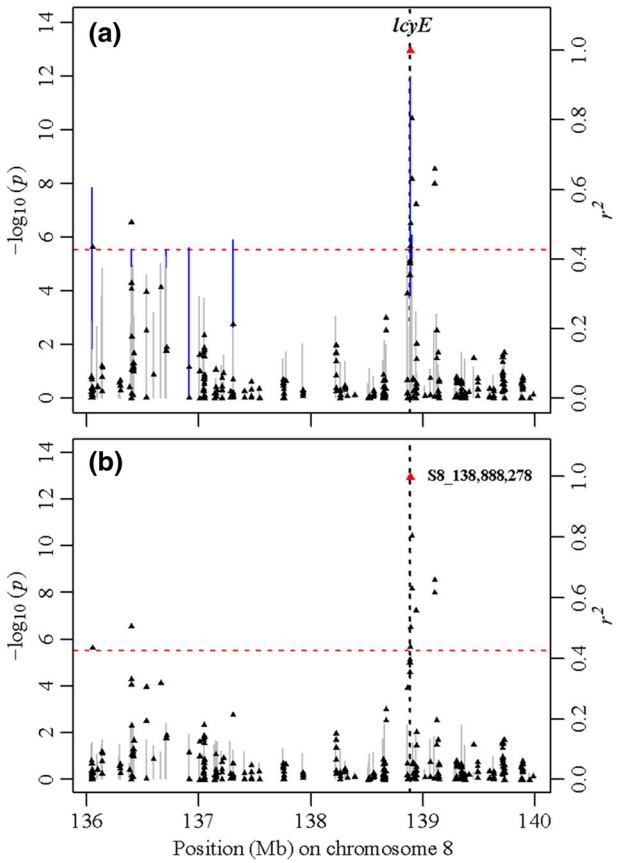


**FIGURE 3** Genome-wide association study for the ratio of β- to α-xanthophylls in fresh kernels of sweet corn. (a) Scatter plot of association results from a mixed linear model analysis and linkage disequilibrium estimates ($r^2$). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNP) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the $r^2$ values of each SNP relative to the peak SNP (indicated in red) at 138,888,278 bp (B73 RefGen_v2) on chromosome 8. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical dashed line indicates the genomic position of the *lycopene ε-cyclase* gene (*lcyE*). (b) Scatter plot of association results from a conditional mixed linear model analysis and linkage disequilibrium estimates ($r^2$). The SNP from the optimal multi-locus mixed-model (S8_138888278) was included as a covariate in the mixed linear model to control for the *lcyE* effect

chromosome 9 were found to be associated with the ratio of β-xanthophylls to α-xanthophylls. However, only the peak SNP S8_138888278 was selected in the optimal model obtained by the MLMM for the xanthophyll ratio phenotype (Supplemental Table S6). When the peak SNP was fitted as a covariate in the mixed linear model, all other SNPs from chromosomes 8 and 9 were no longer significantly associated with β-xanthophylls/α-xanthophylls at a 5% FDR (Figure 3b). Interestingly, only one line each in the *sh2* and *su1sh2* endosperm mutation type groups was homozygous for the allele of the peak SNP associated with a larger average value of the β- to α-xanthophylls ratio (i.e., greater amount of β-xanthophylls), whereas the same SNP allele was found to be homozygous at a relatively higher frequency (12.7%) among *su1* lines (Supplemental Table S7).

We detected seven SNPs covering a 748.29-Kb interval on chromosome 2 that were significantly associated with total xanthophylls. Of these SNPs, six of them were also associated with total carotenoids. The peak association signal for both traits was SNP S2_222880454 (*P*-values $2.74 \times 10^{-8}$,

**TABLE 3** Mean predictive ability (standard deviation) of genomic prediction models for 19 fresh kernel carotenoid traits using three marker sets as predictors

| | GBLUP[a] | | | GBLUP with endosperm mutation type covariate | | |
|---|---|---|---|---|---|---|
| **Trait** | **QTL targeted[b]** | **Pathway-level[c]** | **Genome-wide[d]** | **QTL targeted** | **Pathway-level** | **Genome-wide** |
| Antheraxanthin | 0.35 (0.03) | 0.35 (0.03) | 0.45 (0.02) | 0.39 (0.02) | 0.41 (0.02) | 0.48 (0.02) |
| β-Carotene | 0.41 (0.03) | 0.48 (0.03) | 0.49 (0.04) | 0.41 (0.03) | 0.47 (0.04) | 0.49 (0.04) |
| β-Cryptoxanthin | 0.36 (0.03) | 0.49 (0.03) | 0.61 (0.02) | 0.39 (0.02) | 0.53 (0.02) | 0.63 (0.02) |
| Lutein | 0.54 (0.02) | 0.64 (0.02) | 0.73 (0.01) | 0.65 (0.01) | 0.72 (0.01) | 0.75 (0.01) |
| Violaxanthin | 0.25 (0.03) | 0.27 (0.03) | 0.28 (0.03) | 0.25 (0.02) | 0.27 (0.03) | 0.27 (0.03) |
| Zeaxanthin | 0.24 (0.03) | 0.35 (0.03) | 0.42 (0.03) | 0.24 (0.03) | 0.36 (0.03) | 0.44 (0.03) |
| Zeinoxanthin | 0.42 (0.02) | 0.49 (0.02) | 0.54 (0.02) | 0.43 (0.02) | 0.50 (0.02) | 0.54 (0.02) |
| Other carotenes | 0.21 (0.03) | 0.29 (0.03) | 0.49 (0.02) | 0.42 (0.02) | 0.46 (0.02) | 0.52 (0.02) |
| α-Xanthophylls | 0.53 (0.02) | 0.64 (0.02) | 0.71 (0.01) | 0.63 (0.01) | 0.70 (0.01) | 0.73 (0.01) |
| β-Xanthophylls | 0.26 (0.03) | 0.38 (0.03) | 0.46 (0.03) | 0.26 (0.03) | 0.40 (0.03) | 0.47 (0.03) |
| Total xanthophylls | 0.41 (0.02) | 0.56 (0.02) | 0.67 (0.02) | 0.52 (0.02) | 0.64 (0.01) | 0.70 (0.02) |
| Total carotenes | 0.23 (0.03) | 0.34 (0.03) | 0.53 (0.02) | 0.40 (0.02) | 0.47 (0.03) | 0.55 (0.02) |
| Total carotenoids | 0.40 (0.02) | 0.56 (0.02) | 0.68 (0.02) | 0.52 (0.02) | 0.65 (0.01) | 0.71 (0.02) |
| β-Carotene/β-cryptoxanthin | 0.48 (0.02) | 0.51 (0.03) | 0.54 (0.03) | 0.48 (0.02) | 0.53 (0.03) | 0.55 (0.03) |
| β-Carotene/(β-cryptoxanthin +Zeaxanthin) | 0.44 (0.03) | 0.46 (0.04) | 0.42 (0.05) | 0.43 (0.03) | 0.46 (0.04) | 0.42 (0.05) |
| β-Cryptoxanthin/zeaxanthin | 0.35 (0.02) | 0.44 (0.02) | 0.54 (0.02) | 0.40 (0.02) | 0.50 (0.02) | 0.55 (0.02) |
| Zeinoxanthin/lutein | 0.26 (0.03) | 0.25 (0.03) | 0.33 (0.03) | 0.27 (0.03) | 0.26 (0.03) | 0.33 (0.03) |
| β-/α-Xanthophylls | 0.57 (0.02) | 0.57 (0.02) | 0.62 (0.02) | 0.65 (0.01) | 0.63 (0.02) | 0.64 (0.01) |
| Total carotenes/total xanthophylls | 0.30 (0.03) | 0.32 (0.03) | 0.39 (0.04) | 0.29 (0.04) | 0.31 (0.03) | 0.38 (0.04) |
| Average | 0.37 | 0.44 | 0.52 | 0.42 | 0.49 | 0.53 |

[a]GBLUP, genomic best linear unbiased prediction.
[b]628 markers within ±250 kb of eight a priori genes underlying quantitative trait loci associated with grain carotenoid biosynthesis and retention.
[c]4689 markers within ±250 kb of 60 a priori candidate genes.
[d]172,486 genome-wide markers.

total xanthophylls; $1.15 \times 10^{-7}$, total carotenoids), which was located within the ORF of a putative seryl-tRNA synthetase (GRMZM2G172101). Through a MLMM analysis on a chromosome-wide level, this peak SNP (S2_222880454) was selected in the optimal model obtained for total xanthophylls and total carotenoids. The peak SNP was found to be ~172 Kb from a gene encoding a member of the SWEET (Sugars Will Eventually be Exported Transporter) sucrose-efflux transporter family (*sweet14a*; GRMZM2G094955) that is strongly expressed in the endosperm of developing maize kernels at 16–24 d after pollination (Stelpflug et al., 2016). In a conditional univariate mixed model analysis that included the peak SNP as a covariate, SNPs on chromosome 2 were no longer found to be associated with xanthophyll and carotenoid totals (Supplemental Figure S4). This conditional analysis resulted in the detection of six additional SNPs on chromosome 1 (peak SNP S1_290953298; *P*-value $6.52 \times 10^{-7}$) significantly associated with total carotenoids at an FDR of 5% (Supplemental Table S8), with two of the six SNPs located within a gene that encodes *alkaline galactosidase 4* (*aga4*, GRMZM2G077181). However, within ±250 kb of these six SNPs there were no genes with an encoded protein that had an obvious role in the genetic control of kernel carotenoid levels.

## 3.3 | Carotenoid prediction

To evaluate the potential of genomic selection for enhancing the level of carotenoids in fresh kernels, we evaluated whole-genome prediction (WGP) using the 172,486 SNP markers for all 19 carotenoid phenotypes that had been measured on the 308 sweet corn inbred lines. The average predictive ability across the 19 carotenoid phenotypes was 0.52, with a range in abilities of 0.28 for violaxanthin to 0.73 for lutein (Table 3). The two measured compounds with provitamin A activity, β-carotene and β-cryptoxanthin, had moderately high predictive abilities of 0.49 and 0.61, respectively. A moderately strong positive correlation ($r_{sp} = 0.71$, *P*-value < .001) was found

between heritability estimates and predictive abilities for the 19 carotenoid phenotypes. Conversely, predictive abilities had essentially no correlation ($r_{sp} = -0.036$, $P$-value = 0.883) with the number of significant markers detected in GWAS at 5% FDR.

Carotenoid grain traits in maize show patterns of oligogenic inheritance (Chander et al., 2008; Kandianis et al., 2013; Wong, Lambert, Wurtzel, & Rocheford, 2004), with variability for content and composition mostly under the genetic control of several moderate- to large-effect loci involved in the synthesis or cleavage of carotenoids (Owens et al., 2014). Therefore, we evaluated the predictive ability of two marker datasets that included SNPs within ±250 kb of eight candidate genes underpinning QTL associated with variation for carotenoid levels in maize grain (carotenoid QTL-targeted) or 60 candidate genes involved in carotenoid biosynthesis and retention in maize (pathway-level) (Supplemental Table S4). When compared to the genome-wide marker dataset, on average, the predictive abilities of the 19 phenotypes were 15 and 8 percentage points lower for the carotenoid QTL-targeted and pathway-level marker sets, respectively (Table 3). The predictive ability for β-carotene with the pathway-level set was 7 percentage points higher than that of the QTL-targeted set, but it was essentially equivalent to the predictive ability of the genome-wide marker set (0.49). In contrast, the decrease in predictive abilities for β-cryptoxanthin, lutein, zeaxanthin, and total carotenoids with the carotenoid QTL-targeted dataset ranged from 10–28 percentage points compared to abilities obtained with the pathway-level and genome-wide marker sets.

Given that there were significant differences in variation among endosperm mutation type groups for more than half of the carotenoid phenotypes (Table 2), we evaluated the extent to which predictive abilities would improve from the inclusion of a covariate for the type of endosperm mutation in prediction models that varied for marker coverage of the genome. On average, predictive ability across the 19 phenotypes only increased by a single percentage point when including the endosperm mutation type covariate (Table 3) in the WGP model. Illustrative of the impact of including this covariate for both less dense marker datasets, the improvement in predictive abilities ranged from 5 to 21 percentage points for the eight phenotypes with a highly significant endosperm mutation type effect ($P < .0001$; Table 2) when using the carotenoid QTL-targeted marker dataset, whereas the improvement for the same phenotypes was a slightly narrower range of 6 to 17 percentage points with the pathway-level marker set. The improvements in predictive abilities across both reduced marker datasets were far more modest or negligible for the phenotypes with a weaker significant ($.0001 < P < .05$; range: 0 to 6 percentage points) or nonsignificant ($P > .05$; range: –1 to 2 percentage points) endosperm mutation type effect (Table 2).

# 4 | DISCUSSION

The consumption of sweet corn enhanced for carotenoids, especially lutein and zeaxanthin, has the potential to help reduce the risk of AMD that is prevalent among the elderly in the Western world (Congdon et al., 2004; Friedman et al., 2004). Favorable alleles and haplotypes of genes associated with the genetic control of carotenoid levels and genomic selection models optimized for predictive abilities could be used together to accelerate progress in breeding for higher levels of AMD-impacting carotenoids in sweet corn at the fresh-eating stage. To establish a key step for biofortification efforts in sweet corn, we conducted a GWAS to elucidate the genetic basis of natural variation for 19 highly heritable ($\hat{h}_l^2 = 0.76$–$0.93$) carotenoid phenotypes in fresh kernels with a range of light to dark yellow endosperm color from a panel of 308 inbred lines. Additionally, the predictive ability of genomic prediction models varying in marker densities and the genes they target were tested on the same set of carotenoid phenotypes to provide insights into the potential effectiveness of genomic selection. To our knowledge, this work is the most comprehensive quantitative genetic analysis of carotenoid variation in fresh sweet corn kernels.

The two most abundant carotenoids found in fresh kernels were lutein and zeaxanthin, which is consistent with the previously studied carotenoid profiles of yellow kernels from maize (dent/flint/sweet corn) inbred lines (Kurilich & Juvik, 1999; Owens et al., 2014). The sweet corn association panel showed a 30.3- and 6.61-fold range in variation for lutein and zeaxanthin, respectively. If targeting an intake of 6 mg of lutein and zeaxanthin per day—an amount associated with reduced risk for the development of AMD (reviewed in Mares, 2016)—13 lines (11 *sh2* and 2 *su1* lines) from our panel would provide more than 30% (maximum of 48%) of this amount for lutein+zeaxanthin with only 100 g of fresh sweet corn (Supplemental Figure S5). Although a maximum zeaxanthin value of only 10.71 μg $g^{-1}$ was observed in our sweet corn association panel, this highly heritable phenotype has been shown to be responsive to selection. Through phenotypic selection for elevated zeaxanthin levels and concomitant deeper orange kernel color at the fresh eating stage in tropical sweet corn, O'Hare, Fanning, and Martin (2015) increased zeaxanthin up to 25 μg $g^{-1}$ and later to ∼30 μg $g^{-1}$ (Calvo-Brenes et al., 2019). The maximum β-carotene content observed in our association panel (2.83 μg $g^{-1}$) is 1.67-fold lower than that observed by Fanning et al. (2010) in fresh kernels sampled from a population of 385 tropical sweet corn breeding lines (4.72 μg $g^{-1}$) that had been selected for increased zeaxanthin—a β-branch compound. When considering β-carotene and β-cryptoxanthin together, our sweet corn association panel could provide a maximal 3.6% (women) or 2.8% (men) of the recommended daily allowance for vitamin

A with 100 g (medium size ear) of fresh sweet corn (Supplemental Figure S5). Although the HPLC chromatogram peaks for the compounds summed to comprise the 'other carotenes' phenotype are too numerous and variable to quantify reproducibly in the sweet corn association population (data not shown), some of the peaks had spectra indicative of the presence of a β-ring and, though unidentified, are provitamin A active. Thus, the provitamin A content calculated from summing β-carotene and β-cryptoxanthin represents the minimal estimate of provitamin A in the population. The true content is likely to be somewhat higher.

The GWAS of quantitative variation for carotenoids in fresh sweet corn kernels resulted in the detection of significant associations at the genome-wide level for two core carotenoid biosynthetic pathway genes, *crtRB1* (β-carotene hydroxylase) and *lcyE* (lycopene ε-cyclase). These two genes have been previously associated at the candidate gene (Harjes et al., 2008; Yan et al., 2010) and genome-wide (Azmach, Menkir, Spillane, & Gedil, 2018; Owens et al., 2014; Suwarno, Pixley, Palacios-Rojas, Kaeppler, & Babu, 2015) levels with carotenoids in maize (non-sweet corn) grain at physiological maturity. Within a 5.09-Mb genomic interval that included *crtRB1* on chromosome 10, there was a cluster of 38 SNPs within ± 250 kb of *crtRB1* that significantly associated with β-carotene, β-carotene/(β-cryptoxanthin+zeaxanthin), β-carotene/β-cryptoxanthin, and/or violaxanthin. In an effort to resolve this association complex, the implemented MLMM approach optimally selected the peak SNP S10_135801334, which was located ∼255 kb from *crtRB1*, for β-carotene and its two derivative traits. A second SNP positioned only ∼26 kb away from *crtRB1* was additionally selected by the MLMM for β-carotene and β-carotene/(β-cryptoxanthin+zeaxanthin). Despite the existence of complex linkage disequilibrium patterns and the absence of SNPs scored within *crtRB1*—a gene whose expression levels control β-carotene concentration in the maize endosperm (Yan et al., 2010) and a favorable allele of which has been used for increasing provitamin A content in grain of tropical maize (Azmach, Gedil, Menkir, & Spillane, 2013; Muthusamy et al., 2014), our findings support a role of *crtRB1* as well in the genetic control of β-carotene concentration and its conversion in fresh sweet corn kernels.

Resembling the complex association signal at the *crtRB1* locus, 10 SNPs that collectively covered a 3.69-Mb region on chromosome 8 that included the *lcyE* gene significantly associated with the ratio of β- to α-xanthophylls. However, in contrast to the results from the statistical modeling effort to resolve the expansive signal at *crtRB1*, only the peak SNP (S8_138888278) located within *lcyE* was optimally selected by the MLMM for β-xanthophylls/α-xanthophylls. Comparatively, Harjes et al. (2008) showed through a candidate gene association analysis of yellow/orange colored endosperm lines from the Goodman-Buckler maize association panel (Flint-Garcia et al., 2005) that a multi-allelic promoter indel (5′ transposable element polymorphisms) and a non-synonymous SNP in exon 1 of *lycE* together explained most of the phenotypic variation (5.2-fold effect) for the ratio of flux between the α- and β-carotene branches [(α-carotene + lutein)/(β-carotene + β-cryptoxanthin + zeaxanthin)] of the carotenoid pathway in maize grain, followed by a 3.3-fold effect attributed to a second 8 bp indel in the 3′-untranslated region of *lycE*. Notably, the weaker effect 3′ UTR-indel is only about ∼1 kb from the MLMM-selected SNP S8_138888278 shown to associate with β-xanthophylls/α-xanthophylls in the sweet corn association panel. Additionally, Harjes et al. (2008) demonstrated that the haplotype most favorable for increased levels of β-branch carotenoids had both the 5′ promoter-indel and the 3′ UTR-indel. Therefore, the 5′ promoter-indel could be eventually targeted with the PCR-based genotyping assay developed by Harjes et al. (2008) to more deeply assess the existence of an allelic series at *lycE* across sweet corn inbred lines.

In the sweet corn association panel, the most favorable *crtRB1* and *lcyE* alleles for the MLMM-selected SNPs, which presumably have decreased enzymatic activities to promote accumulation of β-carotene (Harjes et al., 2008; Yan et al., 2010), were at low frequencies among the three endosperm mutation type groups, especially *sh2* and *su1sh2*. An analysis of the four observed *crtRB1* haplotypes showed that only 13 *su1* and two *sh2* lines had the most favorable haplotype (TT) that confers more than twice as much accumulated β-carotene compared to the least favorable haplotype (GC) (Supplemental Table S7), with the GC haplotype present at a frequency ranging from 70.6–79.1% across the three endosperm mutation type groups. Through a haplotype analysis, we also showed that *crtRB1* had no effect on total carotenoids in concordance with previous association mapping studies involving diverse maize (non-sweet corn) inbred lines (Azmach et al., 2018; Yan et al., 2010). Similarly, the favorable allele (T) of the peak SNP marker located within the *lycE* gene that was associated with higher β-branch carotenoids was found at an overall frequency of ∼11% in the association panel (Supplemental Table S7), but only one line each from the *sh2* and *su1sh2* groups was homozygous for the T allele. When considering the entire association panel, we observed only two *su1* lines homozygous for both the favorable TT haplotype and T allele of *crtRB1* and *lycE*, respectively. This is in stark contrast to our findings from an earlier study in the same sweet corn association panel that revealed the most favorable *hggt1* and *vte1* alleles for increased levels of tocotrienols in fresh kernels were almost entirely fixed for *sh2* and *su1sh2* lines (Baseggio et al., 2019). Taken together, stacking the favorable but low frequency *crtRB1* haplotype and *lycE* allele into a single genetic background for *sh2* and *su1sh2* lines is urgently needed for further

improvements in the accumulation of carotenoid compounds from the β-branch such as β-carotene and zeaxanthin in fresh kernels. A similar strategy to combine favorable alleles of *o2* (starchy endosperm mutant), *crtRB1*, *lycE* in a single genetic background has been shown to be successful for increased provitamin A concentration in combination with elevated levels of lysine and tryptophan in yellow grain maize hybrids (Zunjare et al., 2018).

We detected an association signal for total xanthophylls and total carotenoids consisting of seven significant SNPs that spanned a ~750-Kb region on chromosome 2. None of the significant SNPs that comprised this signal were within ±250 kb of the 60 a priori candidate genes involved in the synthesis or retention of carotenoids. Even though this complex association signal for both phenotypes was resolved by the MLMM approach down to the same single peak SNP, the long-range linkage disequilibrium in this region—exemplified by an average $r^2$ of 0.83 between the seven significant SNPs—limited mapping resolution. Nonetheless, the investigated genes contained within this interval did not have a function known to be involved in carotenoid production or regulation of the carotenoid pathway. Interestingly, this region had been previously shown by Baseggio et al. (2019) to also associate with total tocotrienol levels in the same sweet corn association panel. Specifically, the SNP S2_222219441 significantly associated with total xanthophylls (includes lutein and zeaxanthin), total carotenoids, and total tocotrienols, which are all sum traits consisting of metabolites predominantly synthesized in the endosperm (Grams et al., 1970; Weber, 1987). Not only is there a very strong correlation between total xanthophylls and total carotenoids ($r = 0.99$) as expected, but there is also a strong correlation of total tocotrienols with total xanthophylls ($r = 0.61$) and total carotenoids ($r = 0.59$). If not the product of cryptic population structure, these findings suggest the possible presence of a novel locus influencing the production of shared precursor substrates, such as GGDP, which is used by committed pathway steps for the formation of phytoene for carotenoids and 2-methyl-6-geranylgeranyl-1,4-benzoquinol (MGGBQ) for tocotrienols. This hypothesis and the other novel loci on chromosomes 1 (total carotenoids) and 6 (β-carotene) (Supplemental Table S8) will need to be further investigated by a combination of approaches, such as gene expression profiling and mutagenesis, to assess their potential novel contribution to the genetic control of carotenoid levels in fresh sweet corn kernels.

Through WGP in the sweet corn panel, we observed an average predictive ability of 0.52 across the 19 carotenoid phenotypes, with predictive abilities having a strong positive Spearman correlation (0.71) with heritability estimates. With regards to compounds of importance to human health and nutrition, moderate to high predictive abilities (0.42–0.73) were shown for carotenoids central to reducing the risk of AMD (lutein and zeaxanthin) and alleviating vitamin A deficiency (β-carotene and β-cryptoxanthin). When standardized by heritability to allow for a comparison to Owens et al. (2014), prediction accuracies were 1.45 to 2.16-fold higher for β-carotene, β-cryptoxanthin, and lutein but 1.15-fold lower for zeaxanthin in fresh kernels of the sweet corn association panel relative to physiologically mature dry kernels (dent/flint) from the Goodman-Buckler association panel. The markedly higher prediction accuracies for these carotenoid compounds in fresh kernels is probably attributable to the overall higher genetic relatedness between lines in the sweet corn panel—a factor known to elevate trait prediction accuracies in populations of maize and other crops (Albrecht et al., 2014; Gowda et al., 2014; Ly et al., 2013; Wang et al., 2014). Given the overall findings from WGP, it is plausible that genomic selection would be more accurate and offer higher genetic gain than phenotypic selection, particularly for the carotenoid phenotypes with relatively lower heritability (Calus, Meuwissen, de Roos, & Veerkamp, 2008), but this would need to be empirically evaluated in sweet corn breeding populations.

On average, improvement to predictive abilities was negligible when the endosperm mutation type covariate was included in WGP models, but its inclusion would enhance predictive abilities for phenotypes with an endosperm mutation type effect (Table 2) when using either the pathway-level or carotenoid QTL-targeted model. In our same sweet corn association panel, Baseggio et al. (2019) observed similar improvements in predictive abilities for pathway-level and QTL-targeted genomic prediction models for tocotrienols but not tocopherols when accounting for endosperm mutation type. Both tocotrienols and carotenoids are synthesized in the endosperm, whereas the synthesis of tocopherols occurs in the embryo (Grams et al., 1970; Weber, 1987). On average, the *sh2* lines included in our study had significantly higher levels of carotenoids compared to *su1* lines, suggesting the possible involvement of *sh2* in the genetic control of carotenoids as was also postulated by Baseggio et al. (2019) for tocotrienols. However, Baseggio et al. (2019) showed that *sh2* lines were essentially fixed for strong alleles of *vte1* and *hggt1* that likely had an additional contribution to the higher tocotrienol levels in fresh kernels. In contrast, the most favorable *crtRB1* and *lcyE* variants for the synthesis of β-carotene and β-branch carotenoids, respectively, occurred at lower frequencies for the *sh2* lines relative to the *su1* lines, suggesting that other genetic factors outside of the carotenoid pathway could be involved such as the *sweet14a* locus that was less than 175 Kb from the peak SNP associated with total xanthophylls and total carotenoids on chromosome 2. Although the influence of sugar signaling on carotenoid synthesis is largely unknown for the high-sucrose endosperm of *sh2* kernels, sucrose supplementation is associated with an increased accumulation of chlorophylls and carotenoids in wild-type *Arabidopsis* seedlings (Flores-Pérez et al., 2010). Despite these observations, it is still plausible that the inclusion of the endosperm

mutation type covariate in genomic prediction models instead accounts for population structure and relatedness patterns that happen to correlate with the allele frequencies of causal variants controlling quantitative variation for carotenoid traits.

Even though the extent of heritable variation for each carotenoid phenotype is expected to be predominantly controlled by a few genes with moderate to large effects (Owens et al., 2014), on average, lower predictive abilities were achieved with the pathway-level and carotenoid QTL-targeted marker datasets compared to the genome-wide markers used for WGP. With essentially the same sweet corn association panel, Baseggio et al. (2019) showed a comparable pattern in predictive abilities of tocochromanol fresh kernel traits with three marker datasets that similarly varied in scope. A plausible explanation would be that SNPs included in the reduced marker datasets, which target a priori candidate gene loci, were not at a density to be in sufficiently strong linkage disequilibrium with causal variants throughout the genome. Contrastingly, the work of Owens et al. (2014) showed that the carotenoid QTL-targeted marker dataset of eight genes used in our study predicted carotenoid traits with accuracies statistically equivalent to a genome-wide set of markers. However, in their study, four (*crtRB1*, *lycE*, *lut1*, and *zep1*) of the eight genes were detected as strong signals in GWAS, whereas only two (*crtRB1* and *lycE*) were strongly associated with carotenoids at the genome-wide level in our sweet corn association panel. In light of this, transcriptional networks could be integrated to identify novel candidate genes showing weak associations with carotenoids, and these genes combined in prediction models to potentially improve their predictive ability (Chan, Rowe, Corwin, Joseph, & Kliebenstein, 2011; Owens et al., 2014; Schaefer et al., 2018).

## 5 | CONCLUSIONS

We detected a significant association of *crtRB1* with β-carotene concentration and *lcyE* with the ratio of β-xanthophylls to α-xanthophylls in kernels of fresh sweet corn. The most favorable *lcyE* allele and *crtRB1* haplotype for increasing β-branch carotenoids (β-carotene and zeaxanthin) and β-carotene, respectively, were found to be uncommon (≤5%) for lines possessing *sh2*. Therefore, lines with these favorable uncommon variants are potential donors for increasing β-carotene and/or zeaxanthin in *sh2* and *su1sh2* sweet corn breeding populations. Of the 19 carotenoid phenotypes, 13 were significantly different among endosperm mutation types, but further experimental studies involving a combination of higher resolution segregating biparental populations and isogenic lines are needed to identify the biological factors driving these differences. The eventual profiling of carotenoid and gene expression levels across multiple developing kernel stages of select sweet corn lines combined with their whole-genome resequencing data will help to better characterize functional haplotypes at *crtRB1*, *lcyE*, and those yet to be detected because of limitations in statistical power. Furthermore, these findings could then be used to help inform genomic prediction models as to which combinations of haplotypes such as from *crtRB1*, *lcyE*, *zep1*, and *lut1* are most optimal for increasing zeaxanthin and lutein levels for prevention of AMD without diminishing β-carotene levels. Genomic prediction models that used the genome-wide marker dataset showed the most promise for selecting sweet corn lines with elevated levels of lutein, zeaxanthin, and provitamin A carotenoids, but increasing the number of markers in the vicinity of candidate genes associated with carotenoid synthesis and retention is a warranted pursuit having potential benefit for both GWAS and genomic selecton. Collectively, our work in sweet corn represents an important step toward resolving the genetic control of carotenoid synthesis and accumulation at the fresh eating stage, as well as developing optimal genomic prediction models for effectively increasing carotenoids in fresh kernels to levels that will better benefit human health and nutrition.

### AUTHOR CONTRIBUTIONS

M.B. and M.A.G. co-wrote the manuscript; M.B. led the data analysis; M.M.-L. performed the HPLC analyses and metabolite quantifications; N.K. provided overall management of panel growth (planting, pollination, harvesting); M.B., M.M., and J.C. generated the marker datasets; D.D.P. oversaw the metabolite analyses and biological interpretation; W.F.T., M.M., and E.S.B. constructed the association panel; E.S.B., M.E.S., W.F.T., and M.A.G. conceived and designed the project; M.A.G. oversaw the data analysis, project management, design, and coordination.

### CONFLICT OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

### ACKNOWLEDGMENTS

This research was supported by the National Institute of Food and Agriculture; the USDA Hatch under accession numbers 100397 (M.A.G.), 1010428 (M.A.G.), 1013637 (M.A.G.), 1013641 (M.A.G.), and 142 AAC6861 072600 4 (W.F.T.); the National Science Foundation (IOS-1546657 to D.D.P and M.A.G.); Cornell University startup funds (M.A.G.); the USDA-ARS (E.S.B.); and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (M.B.). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. We thank the current and past members of the Tracy and Gore labs for their efforts in pollination, harvest,

and sample preparation. We also thank Jenna Hershberger for providing comments on an earlier version of the manuscript.

## ORCID

*Matheus Baseggio*
https://orcid.org/0000-0002-5789-1663
*Edward S. Buckler*
https://orcid.org/0000-0002-3100-371X
*Michael A. Gore* https://orcid.org/0000-0001-6896-8024

## REFERENCES

Akdemir, D., & Okeke, U. G. (2015). *EMMREML: Fitting mixed models with known covariance structures*. Vienna, Austria: R Core Development Team. https://CRAN.R-project.org/package=EMMREML

Albrecht, T., Auinger, H.-J., Wimmer, V., Ogutu, J. O., Knaak, C., Ouzunova, M., … Schön, C.-C. (2014). Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theoretical and Applied Genetics*, *127*, 1375–1386. https://doi.org/10.1007/s00122-014-2305-z

Azmach, G., Gedil, M., Menkir, A., & Spillane, C. (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. *BMC Plant Biology*, *13*, 227. https://doi.org/10.1186/1471-2229-13-227

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

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

Baseggio, M., Murray, M., Magallanes-Lundback, M., Kaczmar, N., Chamness, J., Buckler, E. S., … Gore, M. A. (2019). Genome-wide association and genomic prediction models of tocochromanols in fresh sweet corn kernels. *The Plant Genome*, *12*. https://doi.org/10.3835/plantgenome2018.06.0038

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

Bone, R. A., Landrum, J. T., Hime, G. W., Cains, A., & Zamor, J. (1993). Stereochemistry of the human macular carotenoids. *Investigative Ophthalmology & Visual Science*, *34*, 2033–2040.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*, 211–252.

Buckner, B., Kelson, T. L., & Robertson, D. S. (1990). Cloning of the *y1* locus of maize, a gene involved in the biosynthesis of carotenoids. *Plant Cell*, *2*, 867–876. https://doi.org/10.2307/3869323

Burt, A. J., Grainger, C. M., Smid, M. P., Shelp, B. J., & Lee, E. A. (2011). Allele mining of exotic maize germplasm to enhance macular carotenoids. *Crop Science*, *51*, 991–1004. https://doi.org/10.2135/cropsci2010.06.0335

Calus, M. P. L., Meuwissen, T. H. E., de Roos, A. P. W., & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, *178*, 553–561. https://doi.org/10.1534/genetics.107.080838

Calvo-Brenes, P., Fanning, K., & O'Hare, T. (2019). Does kernel position on the cob affect zeaxanthin, lutein and total carotenoid contents or quality parameters, in zeaxanthin-biofortified sweet-corn? *Food Chemistry*, *277*, 490–495. https://doi.org/10.1016/j.foodchem.2018.10.141

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

Chan, E. K. F., Rowe, H. C., Corwin, J. A., Joseph, B., & Kliebenstein, D. J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *Plos Biology*, *9*, E1001125. https://doi.org/10.1371/journal.pbio.1001125

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

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*, 759–771. https://doi.org/10.1093/biomet/asn034

Chew, E. Y., Clemons, T. E., Sangiovanni, J. P., Danis, R. P., Ferris, F. L. 3rd, Elman, M. J., … Age-Related Eye Disease Study 2 (AREDS2) Research Group. (2014). Secondary analyses of the effects of lutein/zeaxanthin on age-related macular degeneration progression: AREDS2 report No. 3. *JAMA Ophthalmology*, *132*, 142–149. https://doi.org/10.1001/jamaophthalmol.2013.7376

Combs, G. F. (2012). Vitamin A. In *Vitamins: Fundamental aspects in nutrition and health* (4th ed., pp. 93–138). New York: Academic Press.

Congdon, N., O'Colmain, B., Klaver, C. C. W., Klein, R., Munoz, B., Friedman, D. S., … Eye Diseases Prevalence Research Group. (2004) Causes and prevalence of visual impairment among adults in the United States. *Archives of Ophthalmology*, *122*, 477–485. https://doi.org/10.1001/archopht.122.4.477

Cunningham, F. X., Jr., Pogson, B., Sun, Z., McDonald, K. A., DellaPenna, D., & Gantt, E. (1996). Functional analysis of the beta and epsilon lycopene cyclase enzymes of *Arabidopsis* reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell*, *8*, 1613–1626. https://doi.org/10.1105/tpc.8.9.1613

Cuttriss, A. J., Cazzonelli, C. I., Wurtzel, E. T., & Pogson, B. J. (2011). Carotenoids. *Advances in Botanical Research*, *58*, 1–36. https://doi.org/10.1016/B978-0-12-386479-6.00005-6

DellaPenna, D., & Pogson, B. J. (2006). Vitamin synthesis in plants: Tocopherols and carotenoids. *Annual Review of Plant Biology*, *57*, 711–738. https://doi.org/10.1146/annurev.arplant.56.032604.144301

Egesel, C. O., Wong, J. C., Lambert, R. J., & Rocheford, T. R. (2003). Gene dosage effects on carotenoid concentration in maize grain. *Maydica*, *48*, 183–190.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One*, *6*, e19379. https://doi.org/10.1371/journal.pone.0019379

Fanning, K. J., Martin, I., Wong, L., Keating, V., Pun, S., & O'Hare, T. (2010). Screening sweetcorn for enhanced zeaxanthin concentra-

tion. *Journal of the Science of Food and Agriculture*, *90*, 91–96. https://doi.org/10.1002/jsfa.3787

Flint-Garcia, S. A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., … Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant Journal*, *44*, 1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Flores-Pérez, Ú., Pérez-Gil, J., Closa, M., Wright, L. P., Botella-Pavía, P., Phillips, M. A., … Rodríguez-Concepción, M. (2010). Pleiotropic regulatory locus 1 (PRL1) integrates the regulation of sugar responses with isoprenoid metabolism in *Arabidopsis*. *Molecular Plant*, *3*, 101–112. https://doi.org/10.1093/mp/ssp100

Friedman, D. S., O'Colmain, B. J., Muñoz, B., Tomany, S. C., McCarty, C., de Jong, P. T., … Eye Diseases Prevalence Research Group. (2004) Prevalence of age-related macular degeneration in the United States. *Archives of Ophthalmology*, *122*, 564–572. https://doi.org/10.1001/archopht.122.4.564

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

Fulgoni, V. L., 3rd, Keast, D. R., Bailey, R. L., & Dwyer, J. (2011). Foods, fortificants, and supplements: Where do Americans get their nutrients? *Journal of Nutrition*, *141*, 1847–1854. https://doi.org/10.3945/jn.111.142257

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

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

Gowda, M., Zhao, Y., Würschum, T., Longin, C. F. H., Miedaner, T., Ebmeyer, E., … Reif, J. C. (2014). Relatedness severely impacts accuracy of marker-assisted selection for disease resistance in hybrid wheat. *Heredity*, *112*, 552–561. https://doi.org/10.1038/hdy.2013.139

Grams, G. W., Blessin, C. W., & Inglett, G. E. (1970). Distribution of tocopherols within the corn kernel. *Journal of the American Oil Chemists Society*, *47*, 337–339. https://doi.org/10.1007/BF02638997

Hannah, L. C., Giroux, M., & Boyer, C. (1993). Biotechnological modification of carbohydrates for sweet corn and maize improvement. *Scientia Horticulturae*, *55*, 177–197. https://doi.org/10.1016/0304-4238(93)90031-K

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

Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: An update. *Plant Breeding Reviews*, *22*, 9–111.

Howitt, C. A., & Pogson, B. J. (2006). Carotenoid accumulation and function in seeds and non-green tissues. *Plant, Cell and Environment*, *29*, 435–445. https://doi.org/10.1111/j.1365-3040.2005.01492.x

Hung, H. Y., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., … Holland, J. B. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity*, *108*, 490–499. https://doi.org/10.1038/hdy.2011.103

Ibrahim, K. E., & Juvik, J. A. (2009). Feasibility for improving phytonutrient content in vegetable crops using conventional breeding strategies: Case study with carotenoids and tocopherols in sweet corn and broccoli. *Journal of Agricultural and Food Chemistry*, *57*, 4636–4644. https://doi.org/10.1021/jf900260d

Institute of Medicine. (2000). *Dietary reference intakes for vitamin C, vitamin E, selenium, and carotenoids*. Washington, DC: The National Academies Press.

Jahns, P., & Holzwarth, A. R. (2012). The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica Et Biophysica Acta*, *1817*, 182–193. https://doi.org/10.1016/j.bbabio.2011.04.012

Jerome-Morais, A., Diamond, A. M., & Wright, M. E. (2011). Dietary supplements and human health: For better or for worse? *Molecular Nutrition & Food Research*, *55*, 122–135. https://doi.org/10.1002/mnfr.201000415

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

Kermode, A. R. (2005). Role of abscisic acid in seed dormancy. *Journal of Plant Growth Regulation*, *24*, 319–344. https://doi.org/10.1007/s00344-005-0110-2

Klein, R., Chou, C.-F., Klein, B. E. K., Zhang, X., Meuer, S. M., & Saaddine, J. B. (2011). Prevalence of age-related macular degeneration in the US population. *Archives of Ophthalmology*, *129*, 75–80. https://doi.org/10.1001/archophthalmol.2010.318

Krinsky, N. I., Landrum, J. T., & Bone, R. A. (2003). Biologic mechanisms of the protective role of lutein and zeaxanthin in the eye. *Annual Review of Nutrition*, *23*, 171–201. https://doi.org/10.1146/annurev.nutr.23.011702.073307

Kundu, S., & Gantait, S. (2017). Abscisic acid signal crosstalk during abiotic stress response. *The Plant Genome*, *11*, 61–69. https://doi.org/10.1016/j.plgene.2017.04.007

Kurilich, A. C., & Juvik, J. A. (1999). Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. *Journal of Agricultural and Food Chemistry*, *47*, 1948–1955. https://doi.org/10.1021/jf981029d

Linus Pauling Institute. (2016). *Carotenoids*. Corvallis, OR: Oregon State Univ. https://lpi.oregonstate.edu/mic/dietary-factors/phytochemicals/carotenoids

Lipka, A. E., Gore, M. A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., … DellaPenna, D. (2013). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3 (Bethesda)*, *3*, 1287–1299. https://doi.org/10.1534/g3.113.006148

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

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). Appendix 1: Linear mixed model theory. *SAS for mixed models* (pp. 733–756). Cary, NC: SAS Institute Inc.

Ly, D., Hamblin, M., Rabbi, I., Melaku, G., Bakare, M., Gauch, H. G., … Jannink, J.-L. (2013). Relatedness and genotype × environment interaction affect prediction accuracies in genomic

selection: A study in cassava. *Crop Science*, *53*, 1312–1325. https://doi.org/10.2135/cropsci2012.11.0653

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates, Inc.

Mares, J. (2016). Lutein and zeaxanthin isomers in eye health and disease. *Annual Review of Nutrition*, *36*, 571–602. https://doi.org/10.1146/annurev-nutr-071715-051110

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*, 1819–1829.

Muthusamy, V., Hossain, F., Thirunavukkarasu, N., Choudhary, M., Saha, S., Bhat, J. S., … Gupta, H. S. (2014). Development of β-carotene rich maize hybrids through marker-assisted introgression of β-*carotene hydroxylase* allele. *Plos One*, *9*, e113583. https://doi.org/10.1371/journal.pone.0113583

National Center for Environmental Health. (2012). *Second national report on biochemical indicators of diet and nutrition in the US population 2012*. Atlanta, GA: CDC, Division of Laboratory Sciences. https://www.cdc.gov/nutritionreport/pdf/Nutrition_Book_complete508_final.pdf

National Health and Nutrition Examination Survey. (2016). *What we eat in America, NHANES 2013–2014*. Washington, DC.: USDA Agricultural Research Service. https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/1516/Table_1_NIN_GEN_15.pdf

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models*. Boston: McGraw-Hill.

O'Hare, T. J., Fanning, K. J., & Martin, I. F. (2015). Zeaxanthin biofortification of sweet-corn and factors affecting zeaxanthin accumulation and color change. *Archives of Biochemistry and Biophysics*, *572*, 184–187. https://doi.org/10.1016/j.abb.2015.01.015

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

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909. https://doi.org/10.1038/ng1847

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rein, D. B., Wittenborn, J. S., Zhang, X., Honeycutt, A. A., Lesesne, S. B., Saaddine, J. & Vision Health Cost-Effectiveness Study Group. (2009). Forecasting age-related macular degeneration through the year 2050: The potential impact of new treatments. *Archives of Ophthalmology*, *127*, 533–540. https://doi.org/10.1001/archophthalmol.2009.58

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., … Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, *14*, R55. https://doi.org/10.1186/gb-2013-14-6-r55

Schaefer, R. J., Michno, J.-M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., & Myers, C. L. (2018). Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell*, *30*, 2922–2942. https://doi.org/10.1105/tpc.18.00299

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

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

Sen, S., & Chakraborty, R. (2011). The role of antioxidants in human health. *ACS Symposium Series*, *1083*, 1–37.

Song, J., Li, D., He, M., Chen, J., & Liu, C. (2015). Comparison of carotenoid composition in immature and mature grains of corn (*Zea mays* L.) varieties. *International Journal of Food Properties*, *19*, 351–358. https://doi.org/10.1080/10942912.2015.1031245

Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., & Kaeppler, S. M. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *The Plant Genome*, *9*. https://doi.org/10.3835/plantgenome2015.04.0025

Sun, G., Zhu, C., Kramer, M. H., Yang, S. S., Song, W., Piepho, H. P., & Yu, J. (2010). Variation explained in mixed-model association mapping. *Heredity*, *105*, 333–340. https://doi.org/10.1038/hdy.2010.11

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

Swarts, K., Li, H., Navarro, J. A. R., An, D., Romay, M. C., Hearne, S., … Bradbury, P. J. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *The Plant Genome*, *7*, 1–12. https://doi.org/10.3835/plantgenome2014.05.0023

Tracy, W. F. (1997). History, genetics, and breeding of supersweet (*shrunken2*) sweet corn. In J. Janick (Ed.), *Plant breeding reviews 14* (pp. 189–236). Hoboken, NJ: John Wiley & Sons.

USDA. (2018a). *National nutrient database for standard reference*. Beltsville, MD: USDA, Nutrient Data Laboratory, Human Nutrition Research Center. https://fdc.nal.usda.gov/

USDA. (2018b). *Vegetables 2017 summary*. Washington, DC: USDA, National Agricultural Statistics Service. https://usda.library.cornell.edu/concern/publications/02870v86p?locale=en

Vallabhaneni, R., & Wurtzel, E. T. (2009). Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiology*, *150*, 562–572. https://doi.org/10.1104/pp.109.137042

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang, Y., Mette, M. F., Miedaner, T., Gottwald, M., Wilde, P., Reif, J. C., & Zhao, Y. (2014). The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genomics*, *15*, 556. https://doi.org/10.1186/1471-2164-15-556

Weber, E. J. (1987). Carotenoids and tocols of corn grain determined by HPLC. *Journal of the American Oil Chemists Society*, *64*, 1129–1134. https://doi.org/10.1007/BF02612988

West, K. P. (2002). Extent of vitamin A deficiency among preschool children and women of reproductive age. *Journal of Nutrition*, *132*, 2857S–2866S. https://doi.org/10.1093/jn/132.9.2857S

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

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

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

Yang, R., Yan, Z., Wang, Q., Li, X., & Feng, F. (2018). Marker-assisted backcrossing of *lcyE* for enhancement of proA in sweet corn. *Euphytica*, *214*, 130. https://doi.org/10.1007/s10681-018-2212-5

Zhang, Z., Todhunter, R. J., Buckler, E. S., & Van Vleck, L. D. (2007). Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *Journal of Animal Science*, *85*, 881–885. https://doi.org/10.2527/jas.2006-656

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, *11*, 407–409. https://doi.org/10.1038/nmeth.2848

Zhou, Y., Han, Y., Li, Z., Fu, Y., Fu, Z., Xu, S., … Yang, X. (2012). *ZmcrtRB3* encodes a carotenoid hydroxylase that affects the accumulation of α-carotene in maize kernel. *Journal of Integrative Plant Biology*, *54*, 260–269. https://doi.org/10.1111/j.1744-7909.2012.01106.x

Zunjare, R. U., Hossain, F., Muthusamy, V., Baveja, A., Chauhan, H. S., Bhat, J. S., … Gupta, H. S. (2018). Development of biofortified maize hybrids through marker-assisted stacking of β-*carotene hydroxylase, lycopene-ε-cyclase* and *opaque2* genes. *Frontiers in Plant Science*, *9*, 178. https://doi.org/10.3389/fpls.2018.00178

## SUPPORTING INFORMATION

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

**How to cite this article:** Baseggio M, Murray M, Magallanes-Lundback M, et al. Natural variation for carotenoids in fresh kernels is controlled by uncommon variants in sweet corn. *Plant Genome*. 2020;e20008. https://doi.org/10.1002/tpg2.20008