Published November 21, 2018

Genome-Wide Association and Genomic Prediction Models of Tocochromanols in Fresh Sweet Corn Kernels

Matheus Baseggio, Matthew Murray, Maria Magallanes-Lundback, Nicholas Kaczmar, James Chamness, Edward S. Buckler, Margaret E. Smith, Dean DellaPenna, William F. Tracy, and Michael A. Gore*

M. Baseggio, N. Kaczmar, J. Chamness, E.S. Buckler, M.E. Smith, M.A. Gore, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell Univ., Ithaca, NY 14853; M. Murray, W.F. Tracy, Dep. of Agronomy, Univ. of Wisconsin-Madison, Madison, WI, 53706; M. Magallanes-Lundback, D. DellaPenna, Dep. of Biochemistry and Molecular Biology, Michigan State Univ., East Lansing, MI 48824; E.S. Buckler, Inst. for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; E.S. Buckler, USDA-ARS, Robert W. Holley Center for Agriculture and Health, NY 14853.

ABSTRACT Sweet corn (*Zea mays* L.), a highly consumed fresh vegetable in the United States, varies for tocochromanol (tocopherol and tocotrienol) levels but makes only a limited contribution to daily intake of vitamin E and antioxidants. We performed a genome-wide association study of six tocochromanol compounds and 14 derivative traits across a sweet corn inbred line association panel to identify genes associated with natural variation for tocochromanols and vitamin E in fresh kernels. Concordant with prior studies in mature maize kernels, an association was detected between γ-tocopherol methyltransferase (*vte4*) and α-tocopherol content, along with *tocopherol cyclase* (*vte1*) and *homogentisate geranylgeranyltransferase* (*hggt1*) for tocotrienol variation. Additionally, two kernel starch synthesis genes, *shrunken2* (*sh2*) and *sugary1* (*su1*), were associated with tocotrienols, with the strongest evidence for *sh2,* in combination with fixed, strong *vte1* and *hggt1* alleles, accounting for the greater amount of tocotrienols in *su1sh2* and *sh2* lines. In prediction models with genome-wide markers, predictive abilities were higher for tocotrienols than tocopherols, and these models were superior to those that used marker sets targeting a priori genes involved in the biosynthesis and/or genetic control of tocochromanols. Through this quantitative genetic analysis, we have established a key step for increasing tocochromanols in fresh kernels of sweet corn for human health and nutrition.

Abbreviations: αT, α-tocopherol; αT3, α-tocotrienol; γT, γ-tocopherol; γT3, γ-tocotrienol; δT, δ-tocopherol; δT3, δ-tocotrienol; *aeduwx*, *amylose-extender:dull:waxy*; BLUP, best linear unbiased predictor; *bt2*, *brittle2*; DAP, days after planting; FDR, false discovery rate; GBS, genotyping-by-sequencing; GBLUP, genomic best linear unbiased prediction; GWAS, genome-wide association study; $\hat{h}_i^2$, heritability on a line-mean basis; *hggt1*, *homogentisate geranylgeranyltransferase*; HPLC, high-performance liquid chromatography; JL, joint-linkage; LD, linkage disequilibrium; MEP, methylerythritol phosphate; MLMM, multi-locus mixed-model; NAM, nested association mapping; QTL, quantitative trait locus; RDA, recommended daily allowance; *sh2*, *shrunken2*; SNP, single-nucleotide polymorphism; *se1*, *sugary enhancer1*; *su1*, *sugary1*; total T, total tocopherols; total T3, total tocotrienols; total T3 + T, total tocochromanols; *vte1*, *tocopherol cyclase*; *vte4*, γ-tocopherol methyltransferase; WGP, whole-genome prediction.

CORE IDEAS

- Extensive natural variation exists for tocochromanols in fresh sweet corn kernels.
- *vte4* controls the levels of α-tocopherol, which has the highest vitamin E activity.
- *vte1* and *hggt1* contribute to the genetic control of tocotrienols.
- Sweet corn lines possessing *sh2* have the highest levels of δ- and γ-tocotrienols.
- Prediction abilities were highest for tocotrienols relative to tocopherols.

Tocochromanols, which include four tocopherols and four tocotrienols, are lipid-soluble compounds synthesized by photosynthetic organisms that function as powerful scavengers of lipid peroxy radicals and singlet oxygen quenchers (Kruk et al., 2005). In plants, tocochromanols are important for limiting the oxidation of storage lipids in the seed (Sattler et al., 2004) and providing protection against environmental stress (Liu et al., 2008). The saturated tail of tocopherols is derived from phytyl diphosphate, whereas the unsaturated tail of tocotrienols

Citation: Baseggio, M., M. Murray, M. Magallanes-Lundback, N. Kaczmar, J. Chamness, E.S. Buckler, M.E. Smith, D. DellaPenna, W.F. Tracy, and M.A. Gore. 2019. Genome-wide association and genomic prediction models of tocochromanols in fresh sweet corn kernels. Plant Genome 11:180038. doi: 10.3835/plantgenome2018.06.0038

Received 14 June 2018. Accepted 6 Aug. 2018.
*Corresponding author (mag87@cornell.edu).

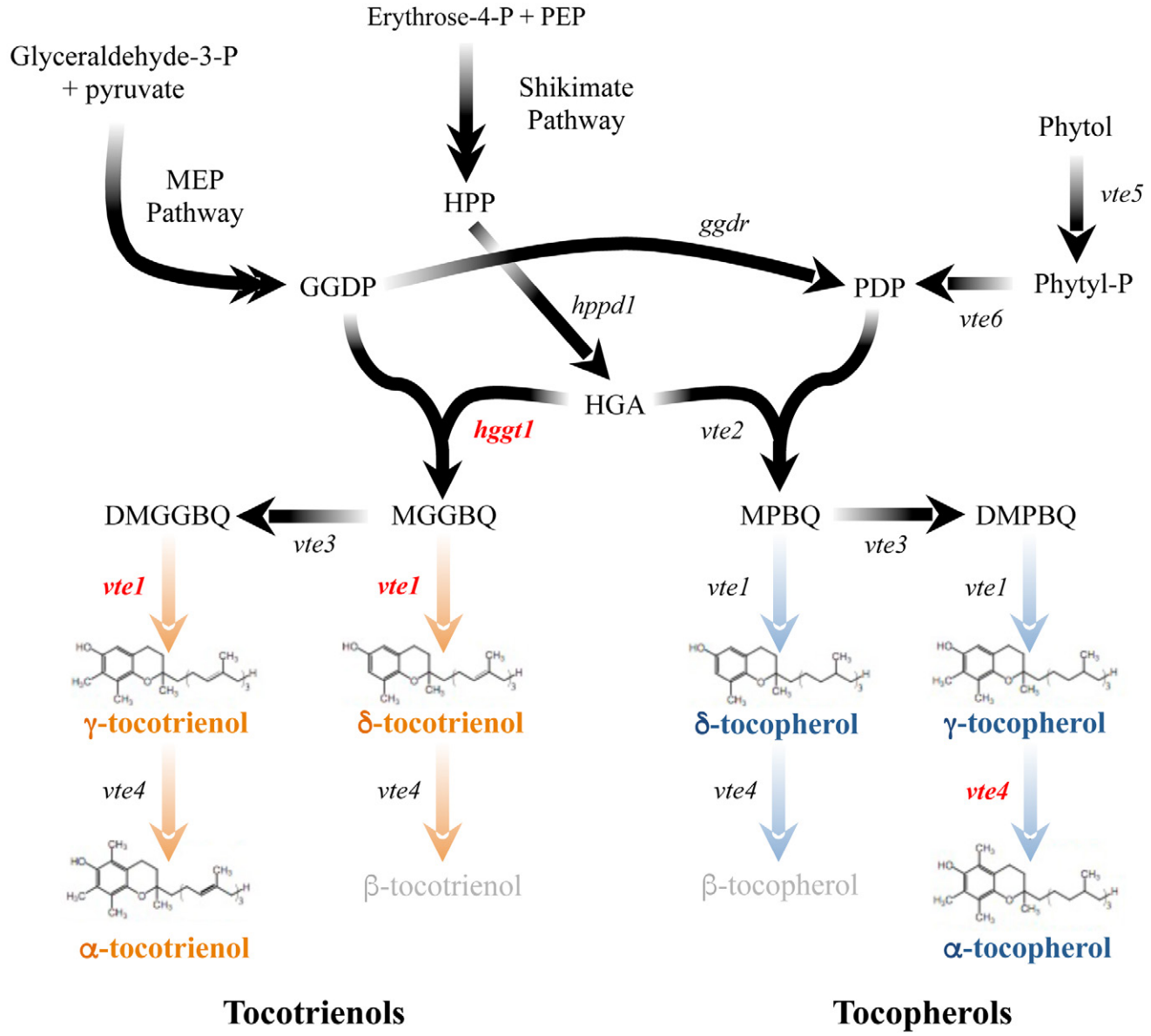

Fig. 1. Tocochromanol biosynthetic pathway in maize. The six quantified compounds are shown in bolded orange (tocotrienols) or blue (tocopherols) text. The name of genes in bolded red text correspond to genes that are within ± 250 kb of the associated single nucleotide polymorphisms (SNPs) identified in our study for an adjacent compound or derivative trait. Compound abbreviations: DMGGBQ, 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol; DMPBQ, 2,3-dimethyl-5-phytylbenzoquinol; GGDP, geranylgeranyl diphosphate; HGA, homogentisic acid; HPP, *p*-hydroxyphenylpyruvate; MGGBQ, 2-methyl-6-geranylgeranyl-1,4-benzoquinol; MPBQ, 2-methyl-6-phytyl-1,4-benzoquinol; PDP, phytyl-diphosphate; PEP, phosphoenolpyruvate; Phytyl-P, phytyl monophosphate. Gene abbreviations: *ggdr, geranylgeranyl diphosphate reductase; hggt1, homogentisate geranylgeranyltransferase; hppd1, 4-hydroxyphenylpyruvate dioxygenase; vte1, tocopherol cyclase; vte2, homogentisate phytyltransferase; vte3, MPBQ/MGGBQ methyltransferase; vte4, γ-tocopherol methyltransferase; vte5, phytol kinase; vte6, phytol phosphate kinase*. Pathway abbreviation: MEP, methylerythritol phosphate.

derives from geranylgeranyl diphosphate (Fig. 1). Within each of the two classes of tocochromanols, the four different chemical species (α, β, δ, and γ) are distinguished by the number and position of methyl groups on the aromatic ring [reviewed in DellaPenna and Mène-Saffrané (2011)]. In general, tocopherol species have greater vitamin E activity than their corresponding tocotrienol species, although tocotrienols tend to have greater antioxidant capacity (Sen et al., 2006). For both classes of compounds, vitamin E activity follows the order α > β > > γ > δ, with α-tocopherol having the highest vitamin E activity on a molar basis (Leth and Sondergaard, 1977).

Like all vitamins in the human diet, vitamin E is required at recommended daily amounts to maintain optimal health (reviewed in Mene-Saffrane, 2017). Although clinical vitamin E deficiency is rare, affecting less than 1% of the US population (Centers for Disease Control, 2006), the prevalence of suboptimal dietary intake among individuals in the United States as measured by plasma α-tocopherol levels is surprisingly high, ranging from 43

to 87%, depending on age and ethnicity (Ford et al., 2006; McBurney et al., 2015). The substandard consumption of vitamin E in the diet has been associated with increased risk of cardiovascular diseases (Knekt et al., 1994; Kushi et al., 1996). Furthermore, limited evidence exists to support the association of vitamin E intake levels with other chronic diseases (Linus Pauling Institute, 2015).

Sweet corn is the third most abundantly consumed vegetable in the United States after tomato (*Solanum lycopersicum* L.) and potato (*Solanum tuberosum* L.) (USDA, 2018b) but the vitamin E level [100% α-tocopherol + 30% α-tocotrienol + 10% γ-tocopherol; Institute of Medicine (2000)] provided from a 100-g intake (one medium to large ear of sweet corn) is only about 2.2% of the recommended daily allowance (RDA) (Xie et al., 2017). Although the vitamin E content of sweet corn is lower than that of tomato (3.6% of the RDA for 100 g) but higher than that of potato (0.3% of the RDA for 100 g) (USDA, 2018a), considerable phenotypic diversity for tocopherols has been reported in a small panel of sweet corn lines and that diversity was highly stable across growing environments (Kurilich and Juvik, 1999; Ibrahim and Juvik, 2009). This suggests that natural variation for vitamin E could be harnessed in sweet corn breeding programs to help address vitamin E insufficiencies where this vegetable is frequently consumed.

The tocochromanol biosynthetic pathway has been fully elucidated and involves 36 enzymatic reactions that are conserved across plant species [reviewed in DellaPenna and Mène-Saffrané (2011); Fig. 1). The aromatic head group for all tocochromanols is homogentisic acid, produced via the shikimate pathway, whereas the hydrophobic tail groups for tocochromanols are generated from isopentenyl pyrophosphate synthesized by the plastid-localized methylerithritol phosphate (MEP) pathway. Condensation of homogentisic acid with phytyl-diphosphate by homogentisate phytyltransferase VTE2 or with geranylgeranyl diphosphate by homogentisate geranylgeranyltransferase (HGGT1) produces the committed precursors for tocopherols and tocotrienols, respectively, which, in turn, are methylated (VTE3 and VTE4) and cyclized (VTE1) in various sequences and combinations to yield the α, β, γ, and δ isoforms. Geranylgeranyl diphosphate for tocotrienol synthesis in maize endosperm is produced directly from isopentenyl pyrophosphate, whereas the phytyl-diphosphate used for tocopherol synthesis in the maize embryo is generated by an indirect route involving a chlorophyll-based cycle (Diepenbrock et al., 2017).

Through several genome-wide association studies (GWAS) of mature maize kernels over recent years, a number of genes responsible for natural variation in tocochromanols and vitamin E levels have been identified. Li et al. (2012) reported a strong association between *vte4* and α-tocopherol content in maize kernels, with deeper insights into this association provided by Lipka et al. (2013). Lipka et al. (2013) also demonstrated the much weaker association of *vte1*, *hggt1*, and an arogenate/prephenate dehydratase with grain tocotrienol levels. In the US maize nested association mapping (NAM) panel, eight genes among the 81 a priori genes in the genome that encode one of the 36 enzymatic reactions were identified to be associated with natural variation of tocochromanols in maize grain (Diepenbrock et al., 2017). Another six loci encoding novel activities also were identified, including, most notably, two protochlorophyllide reductases (*por1* and *por2*), which surprisingly explained the majority of tocopherol content variation in maize grain. Most recently, Wang et al. (2018) associated the genes involved in fatty acid biosynthesis, protein import into the chloroplast, chlorophyll *b* degradation, and the regulation of chlorophyll biosynthesis with tocopherol grain traits.

These findings from GWAS of tocochromanols in maize grain at physiological maturity (dry kernel stage) serve as a starting point for identifying the genes responsible for quantitative variation of tocochromanols in developing kernels of fresh sweet corn. However, sweet corn constitutes a distinct subpopulation that has limited representation in maize diversity panels (Flint-Garcia et al., 2005; Romay et al., 2013). Indeed, Doebley et al. (1988) and Gerdes and Tracy (1994) suggested that sweet corn and dent corn represent two distinct breeding pools, with most sweet corn lines descended from three open-pollinated cultivars: 'Golden Bantam', 'Stowell's Evergreen', and 'Country Gentleman'. Therefore, the favorable alleles for increased tocochromanol content observed in dent corn studies may not be present at high frequency or even at all in the sweet corn germplasm pool. Additionally, Kurilich and Juvik (1999) and Xie et al. (2017) reported that tocopherols tend to increase as sweet corn kernels mature. Therefore, selecting for alleles of causal genes that are favorably expressed early in kernel development is critically important, especially given that fresh sweet corn is harvested around 18 to 21 d after pollination (DAP) (Jennings and McCombs, 1969).

All sweet corn has one or more mutations in the genes involved in the starch biosynthesis pathway that cause kernels to accumulate sugars in the endosperm in place of the starch accumulated in wild-type dent corn (Tracy, 1997). The homozygous *sh2* mutation results in the loss of an adenosine diphosphate-glucose pyrophosphorylase subunit and the accumulation of high levels of sucrose (Michaels and Andrew, 1986). The homozygous *su1* mutation disrupts a starch debranching enzyme that leads to higher levels of water-soluble phytoglycogen, along with increased sucrose, but at levels lower than *sh2* (Doehlert et al., 1993). Other mutations in the starch pathway used singly or in combination for breeding sweet corn include but are not limited to *sugary enhancer1* (*se1*), *brittle2* (*bt2*), and *amylose-extender:dull:waxy* (*aeduwx*) (Hannah et al., 1993). Depending on the single mutation, pleiotropic effects related to compromised biosynthetic capacity can be exerted on other enzymes from the starch biosynthesis pathway [reviewed in Tetlow et al. (2004)]. Illustrative of an interaction with a key phytohormone, ethylene has been shown to impart pleiotropic effects on maize

endosperm development of *sh2* lines (Young et al., 1997). Given the potential amplified connectedness between starch-deficient endosperm mutants and other biochemical pathways, the loci and alleles that control the levels of tocotrienols—the class of tocochromanols that are predominantly synthesized in the endosperm (Grams et al., 1970; Weber, 1987)—may differ in the high-sugar environment of sweet corn endosperm compared with those identified in studies of starch containing dent varieties.

In this study, we constructed a sweet corn association panel that captured the genetic diversity of temperate US breeding programs to dissect the genetic basis of natural variation for tocochromanol content in fresh kernels and develop genomic prediction models that could be used to enhance fresh sweet corn kernels for tocochromanol and vitamin E levels. We conducted (i) a GWAS to identify the genes involved in the genetic control of quantitative variation for tocochromanol levels in fresh (~21 DAP) kernels of sweet corn and (ii) genomic prediction studies to determine the optimal marker density needed to maximize predictive abilities for genomic selection in a sweet corn biofortification breeding program.

MATERIALS AND METHODS

Plant Materials and Experimental Design

We constructed an association panel of 411 diverse sweet corn inbred lines that were selected to sample the levels and patterns of genetic diversity found in the US sweet corn germplasm pool. The panel consisted of inbred lines homozygous for the starch-deficient endosperm mutations *su1*, *su1:se1*, *sh2*, *su1sh2*, *bt2*, and *aeduwx*. An additional 19 inbred lines included in the experiment were known at the time of inclusion or later confirmed (data not shown) not to be sweet corn. The inbred association panel was field-evaluated in the summers of 2014 and 2015 at Cornell University's Musgrave Research Farm in Aurora, NY. For each year, the panel was separated into three sets of varying numbers of lines based on plant height, and the sets were randomly partitioned into incomplete blocks. Within each set, each incomplete block of 20 experimental entries was augmented by the random assignment of two check plots depending on plant height. The incomplete blocks of sets 1 (short), 2 (medium), and 3 (tall) each included the two check lines, either 'We05407' and 'W5579', 'W5579' and 'Ia5125', or 'Ia5125' and 'IL125b', respectively. In addition, the positions of the sets within the field were randomized. Edge effects were reduced by planting a commercial sweet corn line around the perimeter of each replicate. Experimental units were one-row plots with a length of 3.05 m and inter-row spacing of 0.76 m. There was a 0.91-m alley at the end of each plot. In each plot, 24 kernels were planted and each plot was thinned to approximately 12 plants. Standard sweet corn cultivation practices for the Northeast were followed. Weather data were obtained from an automated weather station (Spectrum Technologies, Inc., Aurora, IL) located within the field.

In both years, a single complete replication of the augmented incomplete block design experiment was used for measuring tocochromanol levels. In each plot, six plants were self-pollinated by hand and the pollination dates were recorded. Two self-pollinated ears were hand-harvested from each plot at 400 growing degree-days (~21 DAP) as calculated via the NOAA 86/50 method (Barger, 1969), representing the immature milk stage of kernel development, when sweet corn is picked and eaten as a fresh vegetable. Immediately after harvest, whole ears were frozen in liquid N and shelled. For each sample, frozen kernels were randomly sampled and bulked across the two ears to produce a representative composite kernel sample, then placed in a cryogenic vial and maintained at −80°C. For each sample, 20 to 30 frozen kernels were ground to a fine powder in liquid N. Individual ground samples were transferred to a 1.5-mL tube cooled in liquid N, then transferred for storage at −80°C. Ground kernel samples were packed in dry ice and shipped to Michigan State University (East Lansing, MI) for extraction and measurement of tocochromanols.

Phenotypic Data Analysis

Tocochromanols were extracted from each ground sample, then quantified by high-performance liquid chromatography (HPLC) and fluorometry as previously described (Lipka et al., 2013), with 1 mg mL^{-1} of DL-α-tocopherol acetate added to the extraction buffer as an internal recovery control. The six quantified tocochromanol compounds were δ-tocotrienol (δT3), γ-tocotrienol (γT3), α-tocotrienol (αT3), δ-tocopherol (δT), γ-tocopherol (γT), and α-tocopherol (αT) in μg g^{-1} fresh kernel. Additionally, the following 14 sums, ratios, and proportions were calculated: total tocotrienols (total T3), total tocopherols (total T), total tocochromanols (total T3 + T), αT3/γT3, αT/γT, δT3/αT3, δT/αT, δT3/γT3, δT/γT, γT3/(γT3 + αT3), γT/(γT + αT), δT3/(γT3 + αT3), δT/(γT + αT), and total T/total T3.

To screen the raw HPLC data for significant outliers, we initially used the Box–Cox power transformation (Box and Cox, 1964) with a simple linear model with genotype, year, set within year, block within set within year, and HPLC plate within year as fixed effects to identify the most appropriate transformation that corrected for unequal variances and the non-normality of error terms. The process was conducted using the MASS package in R version 3.2.3 (R Core Team, 2015) and tested lambda values ranging from -2 to +2 in increments of 0.5 before applying the optimal convenient lambda for each phenotype (Supplemental Table S1). Next, the full mixed linear model that allowed for the estimation of genetic effects separately from field design effects, following Wolfinger et al. (1997), was fitted for each phenotype in ASReml-R version 3.0 (Gilmour et al., 2009). The full model fitted was as follows:

$$\begin{aligned} Y_{ijklmnop} = {} & \mu + check + year_j + set(year)_{jk} + block(set \times year)_{jkl} \\ & + genotype_m + genotype \times year_{jm} + plate(year)_{jn} \\ & + row(year)_{jo} + col(year)_{jp} + \varepsilon_{ijklmnop} \end{aligned} \quad [1]$$

in which $Y_{ijklmnop}$ is an individual phenotypic observation, μ is the grand mean, $check_i$ is the fixed effect for the ith check, $year_j$ is the effect of the jth year, $set(year)_{jk}$ is the effect of the kth set within the jth year, $block(set \times year)_{jkl}$ is the effect of the lth incomplete block within the kth set within the jth year, $genotype_m$ is the effect of the mth experimental genotype (noncheck line), $genotype \times year_{jm}$ is the effect of the interaction between the mth genotype and jth year, $plate(year)_{jn}$ is the laboratory effect of the nth HPLC autosampler plate within the jth year, $row(year)_{jo}$ is the effect of the oth plot grid row within the jth year, $col(year)_{jp}$ is the effect of the pth plot grid column within the jth year, and $\varepsilon_{ijklmnop}$ is the residual error effect assumed to be independent and identically distributed according to a normal distribution with a mean of zero and variance σ^2_ε. Except for the grand mean and check term, all other terms were modeled as random effects. Degrees of freedom were calculated via the Kenward–Rogers approximation (Kenward and Roger, 1997). To detect significant outliers, Studentized deleted residuals (Neter et al., 1996) obtained from these mixed linear models were examined.

Once all outliers were removed for each tocochromanol phenotype, an iterative mixed linear model fitting procedure was conducted in ASReml-R version 3.0 (Gilmour et al., 2009) with the full model. Likelihood ratio tests were conducted to remove all terms from the model fitted as random effects that were not significant at $\alpha = 0.05$ (Littell et al., 2006) to generate a final, best fitted model for each phenotype (Supplemental Fig. S1). The final model for each tocochromanol phenotype was used to generate a best linear unbiased predictor (BLUP) for each genotype. The generated BLUPs were used in a GWAS and tocochromanol prediction models (Supplemental Table S2).

Variance component estimates from the reduced model were used to estimate heritability ($\hat{h}^2_r$) on a line-mean basis (Holland et al., 2003; Hung et al., 2012) for each tocochromanol phenotype, with the SE of the estimates calculated via the delta method (Lynch and Walsh, 1998; Holland et al., 2003). Pearson's *r* was used to estimate the degree of association between back-transformed BLUP values for each pair of tocochromanol traits at $\alpha = 0.05$ via the method 'pearson' from the function 'cor.test' in R. The back-transformed BLUP values were calculated with the inverse of the given convenient lambda and used to represent the true directionality of the relationship between traits (Supplemental Table S3).

DNA Extraction, Sequencing, and Genotyping

For each inbred line, a leaf tissue sample consisting of young leaves was collected from a single representative plant. The tissue samples were lyophilized and ground with a GenoGrinder (Spex SamplePrep, Metuchen, NJ). Total genomic DNA was isolated from powdered lyophilized leaf tissue with the DNeasy 96 Plant Kit (Qiagen Incorporated, Valencia, CA). The DNA samples were sent for genotyping-by-sequencing (GBS) at the Cornell Biotechnology Resource Center (Cornell University, Ithaca, NY, USA) following the procedure of Elshire et al. (2011) with *Ape*KI as the restriction enzyme. Genotyping-by-sequencing libraries were constructed in 192- or 384-plex and sequenced on an NextSeq 500 or Illumina HiSeq 2500, respectively (Illumina Incorporated, San Diego, CA). Sequence data that supported the findings of this study have been deposited in the National Center of Biotechnology Information Sequence Read Archive under accession number SRP154923 and in BioProject under accession PRJNA482446.

With the raw GBS sequencing data, the genotypes at 955,690 high confidence single-nucleotide polymorphism (SNP) loci were called with the default parameters in the TASSEL 5 GBSv1 production pipeline with the ZeaGBSv2.7 Production TagsOnPhysicalMap file in B73 RefGen_v2 coordinates (AllZeaGBSv2.7_ProdTOPM_20130605.topm.h5, available at panzea.org, accessed 25 Sept. 2018) (Glaubitz et al., 2014). There were inbred lines from the sweet corn association panel that had also been included in the comprehensive genotyping study of the USDA-ARS North Central Regional Plant Introduction Station collection, conducted by Romay et al. (2013). Therefore, existing raw unimputed SNP genotypic data for 45 sweet corn lines (ZeaGBSv27_publicSamples_rawGenos_AGPv2–150114.h5, available at www.panzea.org, accessed 25 Sept. 2018) that had been phenotyped for tocochromanols in this study were used instead of generating a redundant GBS sequencing dataset. The SNP genotype calls from this study and those of the 45 lines from Romay et al. (2013) were combined and filtered to retain only biallelic SNPs with a call rate greater than 10% (i.e., the percentage of lines successfully genotyped per SNP), as specified by Romay et al. (2013). Missing SNP genotypes were imputed with FILLIN (Swarts et al., 2014) with an available set of maize haplotype donors that had a window size of 4 kb (AllZeaGBSv2.7impV5_AnonDonors4k.tar.gz, available at panzea.org, accessed 25 Sept. 2018). This haplotype-based imputation method is not able to impute all missing data (Swarts et al., 2014) and thus some missing genotype data still remained and had to be filtered.

Upon completion of the imputation procedure, the inbred lines known to have *bt2* ($n = 4$) or *aeduwx* ($n = 2$) were removed from the dataset, allowing the panel to consist of the endosperm mutations most commonly found in the US sweet corn germplasm pool. In TASSEL version 5.2.39, additional quality filters imposed following haplotype-based imputation included removing SNPs with a call rate less than 70%, a minor allele frequency lower than 5%, heterozygosity greater than 10%, an inbreeding coefficient lower than 80%, or a mean read depth greater than 15. Additionally, lines with lower than a 40% call rate (i.e., the percentage of SNPs successfully genotyped for each line) were excluded. The imposition of these quality filters resulted in a final dataset of 174,996 high-quality SNPs scored on 384 lines that had a BLUP value for at least one tocochromanol trait. The raw unimputed SNP genotype calls for the 384 lines are

available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.jd5716f, accessed 8 Oct. 2018).

Genome-Wide Association Study

To conduct a GWAS for each phenotype that controlled for population structure and familial relatedness, a mixed linear model that included the population parameters previously determined approximation for enhanced computing efficiency (Zhang et al., 2010) was used to test for an association between the genotypes of each of the 174,996 SNPs and BLUP values in the R package GAPIT version 2017.08.18 (Lipka et al., 2012). The fitted mixed linear models included four principal components (PCs) (Price et al., 2006) and a kinship matrix based on VanRaden's Method 1 (VanRaden, 2008) that were calculated from a subset of 11,448 genome-wide SNP markers from the complete dataset that had not been imputed and had a call rate higher than 90%, a minor allele frequency greater than 5%, heterozygosity less than 10%, an inbreeding coefficient greater than 80%, and a mean read depth lower than 15. Missing genotypes remaining for all SNP markers (subset and complete marker datasets) were conservatively imputed as a 'middle' (heterozygous) value in GAPIT. The optimal number of PCs to include as covariates in the mixed linear model was determined with the Bayesian information criterion (Schwarz, 1978). A likelihood-ratio-based R^2 statistic (Sun et al., 2010) denoted as R^2_{LR} was used to estimate the amount of phenotypic variation accounted for by the model. The method of Benjamini and Hochberg (1995) was used to account for multiple testing by controlling the false discovery rate (FDR) at 5%.

A chromosome-wide approach for implementing a multi-locus mixed-model (MLMM) (Segura et al., 2012) to resolve association signals involving large-effect genes has been previously described (Lipka et al., 2013). Briefly, the MLMM method used a stepwise mixed-model regression procedure with forward selection and backward elimination. In the first step of this chromosome-wide implementation, only SNP markers on the same chromosome with a major-effect gene were tested as explanatory variables for selection in the optimal model via the extended Bayesian information criterion (Chen and Chen, 2008). The impact of controlling for the influence of a large-effect gene on association signals was then assessed by reconducting the GWAS with the MLMM-selected SNP markers included as covariates in mixed linear models.

Linkage Disequilibrium

The squared allele-frequency correlation (r^2) method of Hill and Weir (1988) was used to estimate linkage disequilibrium (LD) between pairs of SNP loci in TASSEL version 5.2.39 (Bradbury et al., 2007). The dataset of 174,996 high-quality SNP markers was used to estimate LD, with the exception that the remaining missing SNP genotypes were not imputed with the 'middle' value prior to LD analysis.

Visual Classification of Lines for Endosperm Mutations

The sweet corn lines were evaluated for which endosperm mutations they possessed to help us better understand the differences in the content and composition of tocochromanols in fresh sweet corn kernels among the endosperm mutation group types. In 2014, two self-pollinated ears per plot were harvested at physiological maturity and dried to ~15% moisture content. For each plot, an image of two mature ears on the 1KK green background (https://wheatgenetics.org/download/category/21–1kk, accessed 25 Sept. 2018) was taken by hand with a digital camera (Sony DSC-W730, Sony Corporation, Tokyo, Japan). Of the 384 sweet corn inbred lines, 333 of them had images that allowed for visual classification of endosperm mutation type. To classify the inbred lines as having either the recessive *su1* or *sh2* endosperm mutation, the ears in each image were visually scored by one person (Matheus Baseggio) as having kernels with either one of two phenotypes: (i) wrinkled and glassy (*su1* mutation) or (ii) shrunken and opaque to translucent (*sh2* mutation) (Boyer and Shannon, 1983). Given the more visually complex kernel phenotypes that can result from double mutant combinations of endosperm genes (Boyer and Shannon, 1983), it was only possible to confidently score *su1sh2* inbreds as *sh2* and *su1se1* inbreds as *su1*.

Marker-Based Classification of Lines for Endosperm Mutations

The visual classifications resulted in two binomial phenotypes (presence or absence of *su1*; presence or absence of *sh2*) that were used to train genomic prediction models separately for classifying the remaining 51 inbred lines. The 333 inbred lines with visual kernel scores for the presence or absence of *sh2* were randomly divided into two groups: 284 lines (85%) used as a set for training and cross-validating the genomic prediction models and 49 lines (15%) comprising a test set to assess the error of the final selected model. An 85:15 split was also used for modeling the presence or absence of *su1*, with the exception that 15 previously known *su1sh2* inbreds were excluded. This resulted in 271 lines (85%) for training and cross-validating the genomic prediction models and 47 lines comprising the test set for the *su1* locus. Given that the implemented GBS approach did not target specific loci, statistical models were evaluated for their accuracy in predicting the presence or absence of the *su1* or *sh2* endosperm mutations with the following variable sized marker datasets: SNP markers ± 100, 250, 500, 750, or 1000 kb of *su1* (chromosome 4; 41,369,510–41,378,299 bp) or *sh2* (chromosome 3; 216,414,684–216,424,048 bp). The marker datasets consisted of markers selected from the same 174,996 high-quality SNP markers that were also used for GWAS. Prediction of the binomial phenotypes with SNP markers was evaluated with genomic best linear unbiased prediction (GBLUP) (Zhang et al., 2007; VanRaden, 2008). To conduct the GBLUP method, a realized relationship matrix based on VanRaden's Method 1 (VanRaden, 2008) calculated from SNP markers was

fitted with the binomial family and logit link function in ASReml-R version 3.0 (Gilmour et al., 2009). For each locus, the realized relationship matrices were derived from the five sets of the variable-sized marker datasets.

The probability of a homozygous recessive genotype for each locus was obtained from the cumulative distribution function of the logistic distribution using the fivefold cross-validation approach reported in Owens et al. (2014). The average probability for each kernel phenotype was assessed by repeating this process 50 times, with inbred lines classified as having a homozygous recessive genotype for a given locus if the average probability was greater than 0.5. Sensitivity (the proportion of true positives) and specificity (the proportion of true negatives) were calculated for each model, and the SNP marker dataset that maximized the sum of sensitivity and specificity was used to train the GBLUP model with all lines from the training set. The error rate of both final optimal models, which used the 1000-kb marker dataset for each locus, was then assessed on the test set (Supplemental Table S4). Next, the same two final models were used to predict the presence or absence of the *su1* and *sh2* endosperm mutations for the 51 uncharacterized lines and the previously known 15 *su1sh2* inbreds that had been excluded from the *su1* training and test sets.

Tocochromanol Prediction

The ability of SNP markers to predict each of the 20 tocochromanol phenotypes from the 384 inbred lines was evaluated with GBLUP (Zhang et al., 2007; VanRaden, 2008). The GBLUP method was conducted by calculating a realized relationship matrix based on VanRaden's Method 1 (VanRaden, 2008) from SNP markers, followed by model fitting in ASReml-R version 3.0 (Gilmour et al., 2009). The realized relationship matrices were derived from three different sets of SNPs that varied in marker number: genome-wide, pathway-level, and tocochromanol quantitative trait locus (QTL)-targeted. The genome-wide dataset included the 174,996 high-quality SNP markers, whereas the pathway-level dataset consisted of 4819 SNP markers within ± 250 kb of the 81 a priori candidate genes based on prior knowledge of the tocochromanol pathway and its precursors and on homology with *Arabidopsis thaliana* (L.) Heynh. (Diepenbrock et al., 2017; Supplemental Table S5). The tocochromanol QTL-targeted dataset included 946 SNP markers within ± 250 kb of the 14 a priori identified genes (Supplemental Table S6) underlying joint-linkage (JL) QTL associated with grain tocochromanol levels in the US maize NAM panel (Diepenbrock et al., 2017). The fivefold cross-validation approach described in Owens et al. (2014) was used to assess the predictive ability obtained for each phenotype by assessing the Pearson's correlation between observed and genomic estimated breeding values. This process was repeated 50 times for each phenotype, with the mean of these correlations reported as the predictive ability. The same cross-validation folds allowed for a direct comparison among models. Additionally, the stratified sampling approach enabled each fold to be representative of the genotype frequencies for endosperm mutants (*su1*, *sh2*, and *su1sh2*) observed in the entire population. All prediction analyses were performed with and without a covariate accounting for the type of endosperm mutation (*su1*, *sh2*, or *su1sh2*).

Table 1. Means and ranges for back-transformed best linear unbiased predictors (BLUPs) of 20 fresh kernel tocochromanol traits evaluated in the sweet corn association panel and estimated heritability ($\hat{h}_i^2$) on a line-mean basis across 2 yr.

		BLUPs			Heritabilities	
Trait†	Lines	Mean	SD‡	Range	Estimate	SE§
		—μg g^{-1} fresh weight —				
αT	383	1.53	0.70	0.33–4.72	0.87	0.01
αT3	384	1.91	0.43	1.04–3.99	0.68	0.03
δT	383	0.34	0.20	0.04–1.51	0.82	0.02
δT3	384	0.50	0.49	0.07–3.35	0.89	0.01
γT	383	8.67	2.65	1.82–17.4	0.78	0.02
γT3	384	10.41	4.92	2.11–36.96	0.84	0.02
Total T	383	9.61	2.80	2.21–23.48	0.79	0.02
Total T3	384	12.60	4.95	3.48–40.16	0.81	0.02
Total T + T3	383	22.67	6.00	8.34–49.67	0.81	0.02
αT/γT	383	0.20	0.13	0.034–0.98	0.87	0.01
δT/αT	384	0.28	0.24	0.015–1.63	0.85	0.02
δT/γT	384	0.04	0.02	0.015–0.18	0.88	0.01
δT/(γT + αT)	384	0.04	0.02	0.007–0.16	0.86	0.01
γT/(γT + αT)	383	0.84	0.09	0.513–0.97	0.86	0.01
αT3/γT3	384	0.23	0.13	0.061–0.95	0.83	0.02
δT3/αT3	383	0.27	0.25	0.030–1.59	0.87	0.01
δT3/γT3	383	0.04	0.02	0.012–0.18	0.87	0.01
δT3/(γT3 + αT3)	383	0.04	0.02	0.010–0.15	0.88	0.01
γT3/(γT3 + αT3)	384	0.82	0.07	0.528–0.94	0.82	0.02
Total T/Total T3	384	0.86	0.40	0.211–3.09	0.81	0.02

† αT, α-tocopherol; αT3, α-tocotrienol; δT, δ-tocopherol; δT3, δ-tocotrienol; γT, γ-tocopherol; γT3, γ-tocotrienol; Total T3, total tocotrienols; Total T, total tocopherols; Total T + T3, total tocochromanols.

‡ Standard deviation of the BLUPs.

§ Standard error of the heritabilities.

RESULTS

Phenotypic Variation

Phenotypic variation for tocochromanol traits was evaluated in an association panel that was constructed to comprehensively represent the genetic diversity that exists in temperate US sweet corn breeding programs. The measurement of six tocochromanol compounds by HPLC in kernels sampled at the immature milk stage from 384 inbred lines revealed γ-species to be the most abundant, followed by α- and δ-species for both tocopherols and tocotrienols (Table 1). On average, the amount of γT3 moderately exceeded the sum total of all three tocopherol species in the sweet corn population. When lines were separated by the presence or absence of endosperm mutations through the combination of visual and marker-based classifications (Supplemental Table S2 and Supplemental Table S4), the average quantities of γT3 and δT3 were at significantly ($P < 0.0001$) greater levels

Table 2. Back-transformed estimated effects of endosperm mutation type for 20 fresh sweet corn kernel tocochromanol traits.

Trait†	*su1*	*sh2*	*su1sh2*	*P*-value‡
	μg g^{-1} fresh weight			
αT	1.36	1.39	1.48	0.763
αT3	1.84 b	1.98 a	1.70 b	**0.006**
δT	0.30	0.27	0.32	0.407
δT3	0.29 b§	0.72 a	0.69 a	**<0.0001**
γT	8.59	8.08	8.13	0.276
γT3	8.53 b	14.77 a	13.89 a	**<0.0001**
Total T	9.32	8.83	9.08	0.374
Total T3	10.52 b	16.58 a	15.68 a	**<0.0001**
Total T + T3	20.70 b	26.22 a	25.32 a	**<0.0001**
αT/γT	0.16	0.17	0.18	0.464
δT/αT	0.21	0.20	0.22	0.728
δT/γT	0.04	0.04	0.05	0.116
δT/(γT + αT)	0.03	0.03	0.04	0.287
γT/(γT + αT)	0.84	0.84	0.82	0.495
αT3/γT3	0.23 b	0.15 a	0.13 a	**<0.0001**
δT3/αT3	0.16 b	0.35 a	0.41 a	**<0.0001**
δT3/γT3	0.04 b	0.05 a	0.05 a	**<0.0001**
δT3/(γT3 + αT3)	0.03 b	0.04 a	0.04 a	**<0.0001**
γT3/(γT3 + αT3)	0.80 b	0.86 a	0.87 a	**<0.0001**
Total T/Total T3	0.89 b	0.53 a	0.58 a	**<0.0001**

† αT, α-tocopherol; αT3, α-tocotrienol; δT, δ-tocopherol; δT3, δ-tocotrienol; γT, γ-tocopherol; γT3, γ-tocotrienol; Total T3, total tocotrienols; Total T, total tocopherols; Total T + T3, total tocochromanols.

‡ *P*-value from a one-way ANOVA *F*-test for the endosperm mutation type effect. A bolded *P*-value indicates a statistically significant difference between two or more endosperm mutation type groups ($P < 0.05$).

§ Sweet corn lines grouped by endosperm mutation type having labels with the same letter are not significantly different according to the Tukey–Kramer honest significant difference test ($P < 0.05$). The test was only performed for traits that had a significant *F*-test.

in the *sh2* ($n = 76$) and *su1sh2* ($n = 19$) groups than in the *su1* ($n = 289$) group (Table 2). Interestingly, none of the other four tocochromanol compounds showed a similar pattern, with the exception that *sh2* had a significantly greater level of αT3 relative to the *su1* and *su1sh2* groups.

Even though the six compounds are the product of a shared biosynthetic pathway, correlations between compounds only exceeded ~0.20 for two pairs of compounds (Supplemental Fig. S2). The correlation between δT3 and γT3 was 0.77, whereas the correlation of δT with γT was 0.76. The nonexistent to weak correlations between other compound pairs imply the lack of a deeply shared genetic architecture, reflected by tocotrienols being synthesized predominantly in endosperm and tocopherols predominantly in embryo. As shown in Table 1, the six tocochromanol compounds and 14 sum, ratio, and proportion traits had an average heritability of 0.83, with estimates ranging from 0.68 (αT3) to 0.89 (δT3). These high heritabilities suggest that the extent of phenotypic variation for the tocochromanol traits is mostly under genetic control and would be responsive to selection in breeding programs.

Genome-wide Association Study

The genetic architecture of natural variation for tocochromanols in kernels harvested at the immature milk stage was dissected in an association panel of 384 sweet

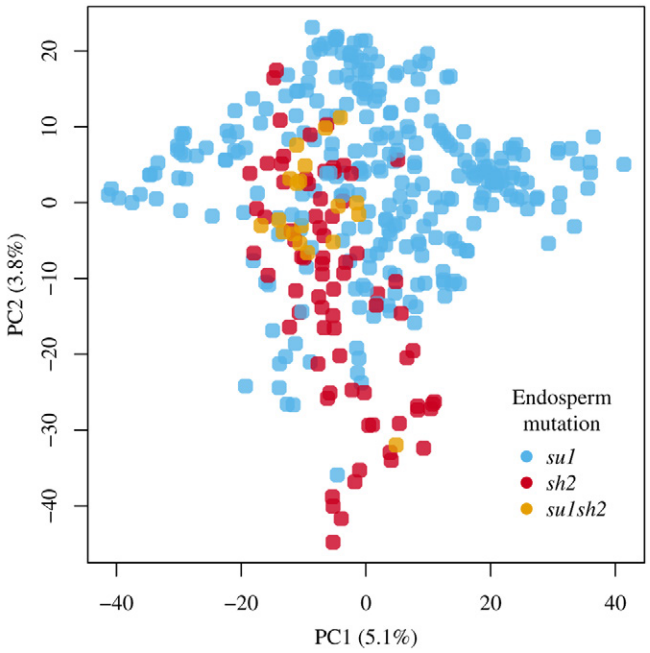

Fig. 2. Principal component analysis of the sweet corn diversity panel. Genetic differentiation of 384 sweet corn inbred lines as revealed by the first two principal components from a principal component analysis of the single nucleotide polymorphism marker data.

corn inbred lines scored with 174,996 high-quality SNP markers at genome-wide coverage. Even though sweet corn is a distinct subpopulation of maize (Romay et al., 2013), moderately weak patterns of population structure appeared to be defined by starch-deficient endosperm mutations within the association panel as inferred by a principal component analysis of the SNP genotypic data (Fig. 2). In the sweet corn association panel, the median LD (50th percentile) estimated with the 174,996 genome-wide SNP markers decayed to low (background) levels ($r^2 < 0.1$) by ~12 kb, but with a large variance in LD structure (Supplemental Fig. S3). Given the persistence of LD at higher percentile cutoffs, the potential importance of distant regulatory elements in the genetic control of kernel tocochromanols (Li et al., 2012), and to intersect with the GWAS results of Lipka et al. (2013), the candidate gene search space was limited to ± 250 kb (median $r^2 \leq 0.05$) of GWAS-detected SNP markers. Through implementation of a mixed linear model that controlled for population stratification and unequal relatedness, 336 unique SNPs were found to significantly associate with one or more phenotypes at a genome-wide FDR of 5% (Supplemental Table S7 and Supplemental Fig. S4). At least one significant SNP was found on every maize chromosome, with the exception of chromosome 10, but 87.2% of the 336 unique SNPs were localized to chromosomes 2, 3, 4, and 5.

The only significant association signals detected for αT were on chromosome 5 (Fig. 3A), with the signal peak defined by two SNPs (S5_200369243 and S5_200369213; *P*-values 2.40×10^{-8} and 3.10×10^{-8}, respectively). The two SNPs were separated by a distance of 30 bp, in perfect LD with each other, and located within an intron

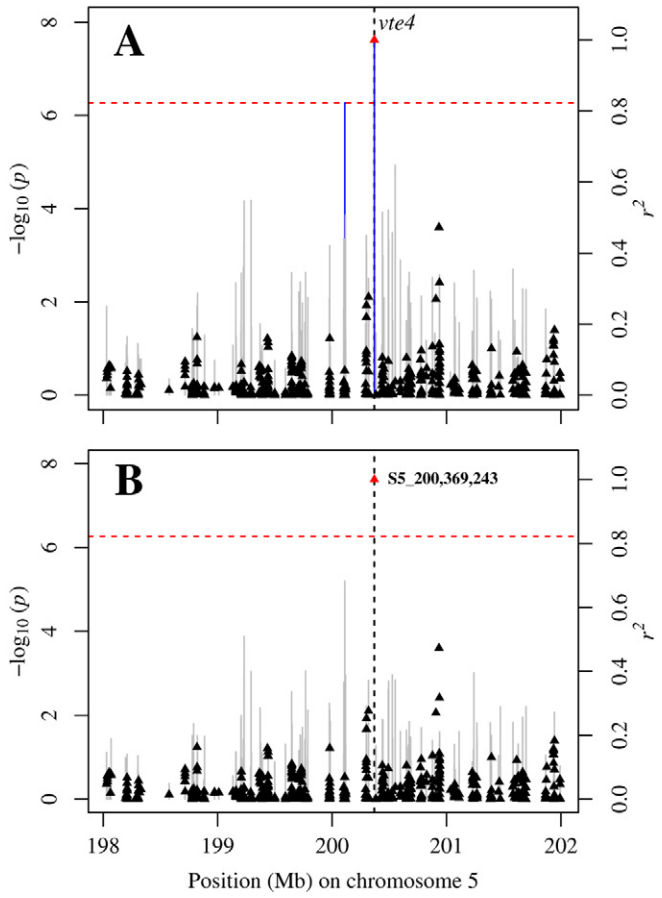

Fig. 3. Genome-wide association study for α- tocopherol (αT) content in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNPs) 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 in red) at 200,369,243 bp (B73 RefGen_v2) on chromosome 5. 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 γ-*tocopherol methyltransferase* gene (*vte4*). (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The peak SNP (S5_200369243) from the optimal multi-locus mixed-model was included as a covariate in the mixed linear model to control for the *vte4* effect.

of the gene encoding γ-tocopherol methyltransferase (*vte4*, GRMZM2G035213), which catalyzes the conversion of γT to αT (Fig. 1). An additional significant SNP positioned more than 250 kb away from the start of the open reading frame for the *vte4* gene and in linkage equilibrium ($r^2 < 0.01$) with the other two significant intronic *vte4* SNPs was associated with αT and the αT/γT ratio (S5_200111824, *P*-values 5.38×10^{-7} and 1.65×10^{-7}, respectively) (Supplemental Fig. S5).

When a chromosome-wide MLMM was used to clarify the association signals in this genomic interval, the optimal model obtained by the MLMM for αT contained the peak SNP S5_200369243, whereas the peak SNP S5_200111824 was selected in the optimal model for the αT/γT ratio (Supplemental Table S8). For each of these two tocochromanol traits, there were no longer any significant associations at a 5% FDR when GWAS was conducted with their MLMM-selected SNP as a covariate in the mixed linear model (Fig. 3B). Notably, the allele of the peak SNP S5_200369243 associated with higher levels of αT was fixed in *su1sh2* lines and all but two *sh2* lines (Supplemental Table S9). The results from this conditional analysis suggest that *vte4* and, potentially, a distant regulatory element are responsible for variation in the αT-related traits.

Within the pericentromeric region of chromosome 5, significant associations were identified between 40 SNPs that spanned a 3.3-Mb interval and at least one of the two tocotrienol-related traits δT3/(γT3 + αT3) and δT3/γT3 (Fig. 4A). The most significant SNP for δT3/γT3 was S5_131738084 (*P*-value 7.64×10^{-9}); S5_133512770 (*P*-value 3.55×10^{-8}) was the peak SNP for δT3/(γT3 + αT3). The latter of the two SNPs was located within an intron of the gene encoding tocopherol cyclase (*vte1*, GRMZM2G009785), an enzyme that catalyzes the synthesis of both γT3 and δT3. Two additional significant SNPs associated with both δT3/γT3 and δT3/(γT3 + αT3) (S5_133505829, *P*-values 5.58×10^{-8} and 1.67×10^{-7}; S5_133501992, *P*-values 6.83×10^{-8} and 1.94×10^{-7}, respectively) were also located within the *vte1* gene. To further resolve signals in the recombination-suppressed *vte1* region, all SNPs on chromosome 5 were considered when conducting the MLMM approach for the two tocotrienol traits. Only the peak SNP was selected in the optimal model for each trait. When each peak SNP was fitted separately as a covariate in the mixed linear model, all other significant associations in the vicinity of *vte1* were eliminated for δT3/(γT3 + αT3) and δT3/γT3 (Fig. 4B and Supplemental Fig. S6).

Signals of association were also identified for tocotrienol traits in the pericentromeric region of chromosome 9. The start of the open reading frame for a gene encoding a homogentisate geranylgeranyltransferase (*hggt1*, GRMZM2G173358), the committed step in tocotrienol biosynthesis (Cahoon et al., 2003), was 138 kb downstream from a SNP (S9_92345469, *P*-value 4.47×10^{-5}; Supplemental Fig. S7) on chromosome 9 that was significantly associated with γT3, the most abundant tocotrienol. Additionally, one SNP each was found to be significantly associated with total T3 (S9_91476108, *P*-value 4.28×10^{-5}) and the total T/total T3 ratio (S9_90663281, *P*-value 2.14×10^{-5}). Indicative of the long-range LD in this region, these two SNPs were in strong (S9_91476108, $r^2 = 0.74$) to moderate (S9_90663281, $r^2 = 0.34$) LD with SNP S9_92345469, although both SNPs were more than 1 Mb away from *hggt1*. Neither these nor any other SNPs were selected by the MLMM for the three tocotrienol traits. This was not unexpected, given that these were among the relatively weaker significant associations for tocochromanol traits in the sweet corn association panel.

Extensive association signals were identified for multiple tocotrienol traits that colocalized with two genes encoding enzymes involved in kernel starch biosynthesis,

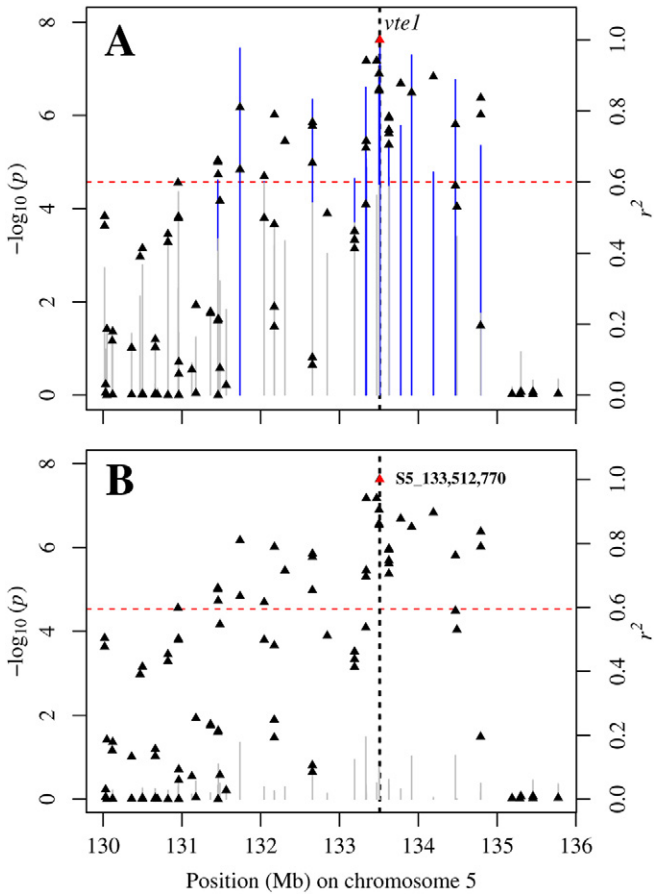

Fig. 4. Genome-wide association study for the ratio of δ-tocotrienol (δT3) to the sum of γ-tocotrienol (γT3) and αT3 [δT3/(γT3 + αT3)] in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed linear model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical 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 in red) at 133,512,770 bp (B73 RefGen_v2) on chromosome 5. 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 *tocopherol cyclase* gene (*vte1*). (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The peak SNP (S5_133512770) from the optimal multi-locus mixed-model was included as a covariate in the mixed linear model to control for the *vte1* effect.

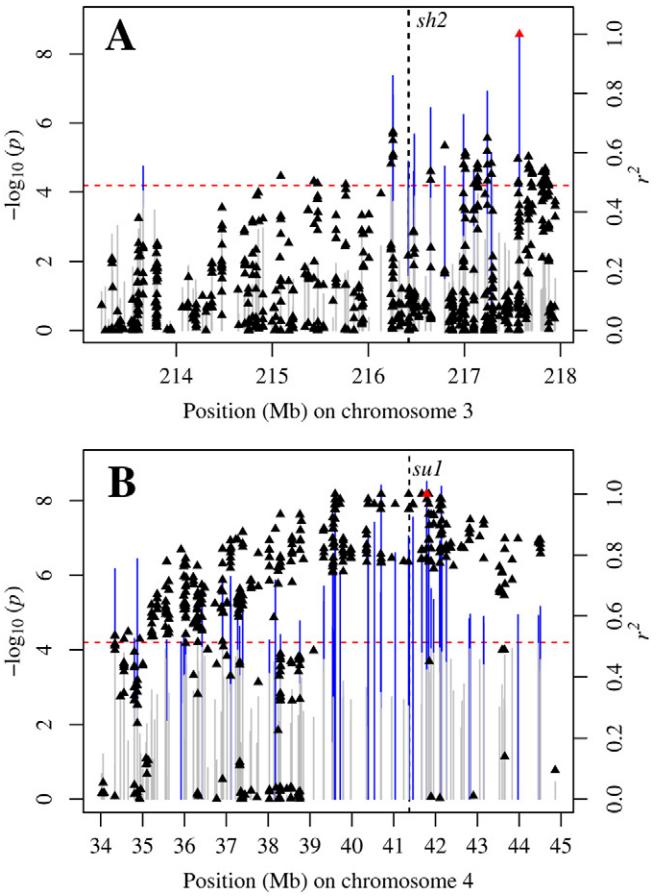

Fig. 5. Genome-wide association study for δ-tocotrienol (δT3) content in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed linear model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are -$\log_{10}$ *P*-values of single nucleotide polymorphisms (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 in red) at 217,572,130 bp (B73 RefGen_v2) on chromosome 3. 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 gene *shrunken2* (*sh2*). (B) Scatter plot of association results from a mixed linear model analysis and LD estimates (r^2). The vertical lines are -$\log_{10}$ *P*-values of SNPs and blue color represents SNPs that are statistically significant at a 5% FDR. Triangles are the r^2 values of each SNP relative to the peak SNP (indicated in red) at 41,789,076 bp (B73 RefGen_v2) on chromosome 4. The black vertical dashed line indicates the position of the gene *sugary1* (*su1*).

adenosine diphosphate-glucose pyrophosphorylase, large subunit (*sh2*, GRMZM2G429899) and isoamylase-type starch debranching enzyme 1 (*su1*, GRMZM2G138060). Recessive mutations for either of these genes inhibit starch formation and increase sugar levels in the endosperm (Creech, 1965), the tissue in which tocotrienols are synthesized (Grams et al., 1970; Weber, 1987). On chromosome 3, 38 SNPs spanning a ~4 Mb interval that included *sh2* were significantly associated with at least one of eight tocotrienol-related traits (Fig. 5A). Similarly, 179 SNPs that covered a 10.2-Mb region encompassing *su1* on chromosome 4 were found to be significantly associated with one or more of seven tocotrienol-related traits (Fig. 5B). In concordance with the findings of Wilson et al. (2004), long-range patterns of LD defined both the *sh2* and *su1* genomic regions (Fig. 5A and B), thus limiting the mapping resolution. Of these 217 total SNPs from chromosomes 3 and 4, all but one of them were also significantly associated with the type of endosperm mutation when used as a phenotype (*su1*, *sh2*, or *su1sh2*) for the 384 lines in GWAS (Supplemental Fig. S8).

Given the strong diffuse signals of association at *su1* and *sh2* and their potential contribution to distant complex associations (interchromosomal LD of $r^2 = 0.61$ between the peak SNPs S3_216256039 and S4_41789076 at *sh2* and *su1*, respectively), GWAS was reconducted for all 20 tocochromanol traits with the type of endosperm

Table 3. Predictive abilities of genomic prediction models using three marker sets as predictors and significant marker associations for 20 fresh sweet corn kernel tocochromanol traits.

	GBLUP						GBLUP with endosperm mutation type covariate						
	Genome-wide†		Pathway-level‡		QTL targeted§		Genome-wide		Pathway-level		QTL targeted		Significant marker–trait associations¶
Trait	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
αT#	0.38	0.02	0.37	0.02	0.36	0.02	0.38	0.02	0.37	0.02	0.36	0.02	3
δT	0.40	0.03	0.34	0.02	0.27	0.02	0.39	0.03	0.33	0.02	0.26	0.03	0
γT	0.39	0.02	0.36	0.02	0.30	0.02	0.40	0.02	0.36	0.02	0.30	0.02	0
Total T	0.42	0.01	0.39	0.02	0.33	0.02	0.43	0.01	0.39	0.02	0.33	0.02	0
αT/γT	0.32	0.03	0.32	0.02	0.31	0.02	0.32	0.03	0.31	0.02	0.30	0.02	1
δT/αT	0.35	0.03	0.30	0.02	0.29	0.02	0.34	0.03	0.28	0.02	0.27	0.02	0
δT/γT	0.45	0.03	0.35	0.02	0.28	0.02	0.44	0.03	0.34	0.02	0.27	0.03	0
δT/(γT + αT)	0.41	0.03	0.30	0.02	0.24	0.02	0.40	0.04	0.29	0.02	0.23	0.03	0
γT/(γT + αT)	0.30	0.03	0.29	0.02	0.29	0.02	0.29	0.03	0.28	0.02	0.28	0.02	0
Average T	0.38		0.34		0.30		0.38		0.33		0.29		0.44
αT3	0.44	0.02	0.40	0.02	0.31	0.02	0.44	0.02	0.40	0.02	0.31	0.02	0
δT3	0.65	0.01	0.53	0.02	0.41	0.02	0.70	0.01	0.66	0.01	0.59	0.01	230
γT3	0.62	0.01	0.53	0.02	0.46	0.02	0.67	0.01	0.64	0.01	0.62	0.01	174
Total T3	0.61	0.01	0.52	0.02	0.44	0.02	0.65	0.01	0.63	0.01	0.59	0.01	165
αT3/γT3	0.59	0.01	0.49	0.02	0.46	0.02	0.61	0.01	0.55	0.01	0.56	0.01	5
δT3/αT3	0.68	0.01	0.55	0.02	0.46	0.02	0.71	0.01	0.64	0.01	0.61	0.01	83
δT3/γT3	0.57	0.01	0.47	0.02	0.27	0.03	0.61	0.01	0.55	0.02	0.41	0.02	0
δT3/(γT3 + αT3)	0.62	0.01	0.51	0.02	0.33	0.02	0.66	0.01	0.60	0.01	0.48	0.02	95
γT3/(γT3 + αT3)	0.57	0.01	0.49	0.02	0.45	0.02	0.59	0.01	0.53	0.02	0.54	0.02	2
Average T3	0.59		0.50		0.40		0.63		0.58		0.52		83.78
Total T3 + T	0.55	0.01	0.49	0.02	0.43	0.02	0.57	0.01	0.55	0.01	0.52	0.01	0
Total T/Total T3	0.49	0.02	0.38	0.02	0.28	0.02	0.56	0.01	0.52	0.01	0.49	0.01	120
Overall average	0.49		0.42		0.35		0.51		0.46		0.42		43.90

† 174,996 genome-wide markers.

‡ 4819 markers within ± 250 kb of 81 a priori candidate genes.

§ 946 markers within ± 250 kb of 14 a priori genes underlying joint-linkage quantitative trait loci (QTL) associated with grain tocochromanol levels in the US maize nested association mapping panel.

¶ The number of significant marker associations for each trait in a genome-wide association study without covariates at a genome-wide false discovery rate of 5%.

αT, α-tocopherol; αT3, α-tocotrienol; δT, δ-tocopherol; δT3, δ-tocotrienol; γT, γ-tocopherol; γT3, γ-tocotrienol; Total T3, total tocotrienols; Total T, total tocopherols; Total T + T3, total tocochromanols; GBLUP, genomic best linear unbiased prediction.

mutation (*su1*, *sh2*, or *su1sh2*) as a covariate in the mixed linear model. When we controlled for association signals at *su1* and *sh2*, the significant association of SNPs (*P*-values 3.37×10^{-10} to 1.43×10^{-5}) encompassing *vte1* with δT3/(γT3 + αT3) and δT3/γT3 still remained (Supplemental Fig. S9 and Supplemental Table S10). Additionally, the two SNPs within *vte4* (S5_200369243 and S5_200369213) were still found to be significantly associated with αT (*P*-values 2.21×10^{-8} and 2.81×10^{-8}, respectively). In contrast, SNP S9_92345469 (*P*-value 6.34×10^{-5}) within 200 kb of *hggt1*, which was associated with γT3 when not controlling for endosperm mutation type was no longer significant at 5% FDR (Supplemental Fig. S9). Interestingly, the alleles of this SNP were not equally distributed between *su1* and *sh2* lines, such that all but one of the 76 *sh2* lines were fixed for the SNP allele associated with higher levels of γT3 (Supplemental Table S9).

To further clarify the association signal at *vte1* and of additional loci potentially masked by the endosperm effect, an MLMM analysis with endosperm mutation type as a covariate was conducted on a chromosome-wide level for tocotrienol-related traits. The optimal models obtained for δT3/γT3 and δT3/(γT3 + αT3) both included the SNP S5_131738084 (Supplemental Table S8), which was also the MLMM-selected peak SNP for δT3/γT3 when not controlling for endosperm mutation type. Although 1.76 Mb from *vte1*, this SNP was in high LD with four SNPs contained in *vte1* ($r^2 = 0.60–0.81$) that were significantly associated with δT3/γT3 and δT3/(γT3 + αT3) when controlling for endosperm mutation type. Additionally, the MLMM analysis resulted in the selection of the same single SNP (S5_214707875) for αT3/γT3 and γT3/(γT3 + αT3), representing a novel association for these two tocotrienol traits on chromosome 5 (Supplemental Table S8). This SNP is within a gene encoding a zinc finger family protein (GRMZM2G178038).

With endosperm mutation type and the two MLMM-identified SNPs (S5_131738084 and S5_214707875) as covariates in the mixed linear model, GWAS was reconducted for all tocochromanol traits (Supplemental Table S11). It was found that one or more tocotrienol traits were significantly associated with a total of seven SNPs at 5% FDR (Supplemental Table S11). On chromosome 1, two SNPs (S1_279565998

and S1_279566000) in perfect LD and located within a gene encoding a peroxidase superfamily protein (GRMZM2G047456) were significantly associated (*P*-values 3.91×10^{-7} and 5.42×10^{-7}) with δT3/γT3. These two SNPs also had a slightly weaker association (FDR-adjusted *P*-value of 0.06) with δT3/(γT3 + αT3). Significant associations were also detected between total T3 and five SNPs (*P*-values 1.24×10^{-7}–1.19×10^{-6}) contained within (two SNPs) or ~225 to 590 kb away (three SNPs) from a gene that encoded an abscisic acid or stress-induced protein (HVA22, GRMZM2G311011) on chromosome 2. Additionally, one of these five SNPs was significantly associated with δT3 at the 5% FDR level. Although it is tempting to speculate on the biological involvement of these three associated genes, higher mapping resolution in combination with gene expression profiling and mutagenesis approaches are needed to assess the potential contribution of these identified novel loci to the genetic basis of tocotrienol traits more completely.

Prediction of Tocochromanols

The promise of genomic selection as an approach for the genetic improvement of fresh kernels for levels of tocochromanols and vitamin E in sweet corn breeding populations was evaluated. The predictive ability of whole-genome prediction (WGP) was assessed for all 20 tocochromanol phenotypes from the 384 inbred lines with the genome-wide dataset of 174,996 SNP markers. This analysis revealed a predictive ability of 0.49 averaged across the 20 phenotypes, with abilities ranging from 0.30 for γT/(γT + αT) to 0.68 for δT3/αT3 (Table 3). When all traits were considered, the correlation between heritabilities and predictive abilities was not statistically significant at a level of α = 5% ($r = 0.18$; *P*-value = 0.44). On average, tocotrienol-related traits had a higher predictive ability (average = 0.59) than tocopherol-related traits (average = 0.38). There was a strong, positive correlation ($r = 0.65$, *P*-value < 0.01) between the number of significant markers observed in GWAS at 5% FDR and predictive abilities (Table 3), which could partly account for the difference in predictive abilities between tocopherol and tocotrienol phenotypes.

Given the oligogenic nature of tocochromanol grain traits in maize, where most of the phenotypic variation is explained by a few moderate- to large-effect loci associated with biosynthetic pathways (Diepenbrock et al., 2017), pathway-level and tocochromanol QTL-targeted marker datasets were also evaluated for their predictive abilities. These two marker datasets included SNPs in proximity of either 81 a priori candidate genes from the precursor and core tocochromanol pathways in maize (Supplemental Table S5) or 14 genes underlying the QTL responsible for 56 to 93% of tocochromanol variation in maize grain (Supplemental Table S6; Diepenbrock et al., 2017). On average, the predictive abilities of both the pathway-level (0.42) and tocochromanol QTL-targeted (0.35) marker datasets for the 20 tocochromanol phenotypes were lower than that obtained with the genome-wide marker dataset (0.49; Table 3). Specifically, tocotrienol-related traits showed the highest accuracy reduction, with an average decrease of 9 percentage points (81 candidate gene set) and 19 percentage points (14 QTL gene set) relative to the genome-wide marker dataset. In contrast, there was an average decrease of only 4 to 8 percentage points for tocopherol-related traits across the two loci-focused marker sets, suggesting that the inclusion of additional genes are more critically needed to predict tocotrienol levels in sweet corn accurately.

In an effort to improve predictive ability, the endosperm mutation type (*su1*, *sh2*, or *su1sh2*) was assessed as a covariate in prediction models evaluating the marker datasets with three different levels of genome coverage. With the genome-wide marker dataset, the inclusion of the endosperm mutation type covariate improved predictive ability by five percentage points for both γT3 and δT3, the only two compounds that had a significant association with SNPs within and nearby *su1* and *sh2* (Supplemental Table S7). When the same covariate was included in prediction models with the pathway-level and tocochromanol QTL-targeted marker datasets, there were similarly improvements in accuracy for γT3 and δT3, but the increase in predictive abilities was higher and ranged from 11 to 18 percentage points across both marker sets. This allowed the predictive abilities of γT3 and δT3 with the pathway-level marker dataset to nearly equal that obtained with genome-wide markers when this covariate was not included. Conversely, the predictive abilities for the other three tocopherol compounds, which are synthesized in the embryo, and αT3 were essentially unchanged across all three marker datasets when the covariate for endosperm mutation type was included. Taken together, these results suggest that capturing the genetic information associated with *su1* and *sh2* is important to improving the predictive ability of γT3, δT3, and their related derivative traits when selecting in breeding populations that are segregating for both the *su1* and *sh2* endosperm mutations.

DISCUSSION

Improving the nutritional quality of fresh sweet corn through genetic improvement offers an avenue to help address vitamin E insufficiencies where this vegetable is frequently consumed. Such biofortification efforts would be enhanced by association studies that identify the loci underlying phenotypic variation for tocochromanol levels in sweet corn kernels. As a complement to GWAS for the genetic dissection of these nutritional kernel traits, the optimization of predictive abilities for genomic selection models with marker sets that are genome-wide or that more directly target genes controlling tocochromanol phenotypes would also provide insight into the genetic gains that could be expected under selection in a breeding program. In that light, we conducted a GWAS to identify the genetic controllers of natural variation for 20 tocochromanol kernel traits and, under different marker set scenarios, assessed the accuracy of genomic

prediction models that could be used for the biofortification of sweet corn. This work represents the first GWAS conducted in a sweet corn association panel, the most extensive assessment of natural variation for tocochromanol levels in fresh kernels, and the first genomic prediction analysis of tocochromanol traits in maize.

The extent of variation for α-, δ-, and γ- tocopherols and tocotrienols in fresh kernels from an association panel of diverse sweet corn lines was evaluated. The 3.8- to 47.8-fold range in variation, calculated as the maximum divided by the minimum BLUP value for each trait, revealed for these six tocochromanol compounds represents extensive phenotypic variability at the fresh kernel stage, with αT (the highest vitamin E activity compound) having a 14.3-fold range in variation. In contrast, there was a 30-fold higher range of variation for αT in physiologically mature dry kernels of temperate and tropical non-sweet corn lines comprising the Goodman–Buckler association panel (Lipka et al., 2013) relative to our sweet corn panel. The likely drivers explaining these differences are the higher levels of allelic diversity captured by the Goodman–Buckler association panel (Flint-Garcia et al., 2005) and the continued accumulation of tocochromanols to higher levels in the kernel beyond the ~21 DAP analyzed in this study (Kurilich and Juvik, 1999; Xie et al., 2017), the time point when sweet corn is typically harvested for consumption. Irrespective of these limitations, the observed wide range of phenotypic variation in the sweet corn association panel was found to be highly heritable ($\hat{h}_i^2$ = 0.68–0.89) and capture a more biologically relevant topmost RDA of 4.4% for vitamin E (Supplemental Fig. S10).

Through a GWAS of the sweet corn association panel, significant associations for three core tocochromanol pathway genes (*vte4*, *vte1*, and *hggt1*) were identified at the genome-wide level, which are in agreement with GWAS results from prior studies of mature kernels in maize (Li et al., 2012; Lipka et al., 2013; Diepenbrock et al., 2017; Wang et al., 2018). The significant association between two nonindependent SNPs within *vte4* and αT content confirms the critical role of this biosynthetic gene in compositional profiles for fresh sweet corn kernels. This association signal defined by two SNPs in complete LD was resolved to a single SNP selected by the MLMM, suggesting a lack of allelic heterogeneity, as was implicated for *vte4* in the Goodman–Buckler association panel (Lipka et al., 2013). However, this could be attributed to the relatively fewer number of scored SNPs and the expected lower haplotype diversity at *vte4* in the sweet corn association panel. Evidence was found to support the hypothesis of a distant upstream regulatory element at *vte4*, with an association signal ~170 kb away from a putative regulatory element previously shown to be associated with αT levels in maize grain (Li et al., 2012).

In concordance with Lipka et al. (2013) and Diepenbrock et al. (2017), moderately strong associations were identified between SNPs spanning a recombination-ally suppressed pericentromeric region that encompassed *vte1* and tocotrienol traits. In these two previous studies, the clear attribution of the association signal to *vte1* was confounded by complex patterns of LD in the genomic interval. In our study, however, the optimal MLMM model was able to resolve the detected association signal to within the *vte1* gene for δT3/(γT3 + αT3). The enzyme encoded by *vte1*, tocopherol cyclase, converts 2,3-dimethyl-5-geranylgeranylbenzoquinol and 2-methyl-6-geranylgeranylbenzoquinol to γT3 and δT3, respectively, which is in agreement with the association of *vte1* with δT3/(γT3 + αT3) and δT3/γT3 in the sweet corn association panel. Furthermore, *vte1* is expressed at low levels in the tocotrienol-rich endosperm (Stelpflug et al., 2016) and therefore is more likely to be a limiting factor for tocotrienol than tocopherol biosynthesis (Lipka et al., 2013). Taken together, to date this is the strongest support that implicates *vte1* as contributing to the natural variation of tocotrienols in kernels.

A second gene related to tocotrienol biosynthesis, *hggt1*, was shown to be associated with tocotrienol traits. The HGGT enzyme catalyzes the first committed step of tocotrienol synthesis and is expressed strongly in the endosperm (Stelpflug et al., 2016), the site of tocotrienol accumulation (Grams et al., 1970; Weber, 1987). Within this pericentromeric region on chromosome 9, the most significant SNP associated with γT3 was 138 kb away from *hggt1*, which is within the range of physical distances for SNPs upstream of *hggt1* that had a significant association with γT3 and total tocotrienols in maize grain through a pathway-level analysis (Lipka et al., 2013). The two significant SNP associations detected at 1 and 1.8 Mb upstream of *hggt1* for total T and the total T/total T3 ratio, respectively, are most likely to have resulted from long-range LD patterns rather than very distant regulatory elements. In the US maize NAM panel, *hggt1* was shown to explain the highest phenotypic variation (24.0–40.2%) for δT3, γT3, and total T3. Although it was not the most significant locus for tocotrienol-related traits in this study and that of Lipka et al. (2013), the difference is probably because these two mapping panels have weaker statistical power from their smaller sample size, rarer allele frequencies, and fewer scored SNP markers relative to the NAM panel.

In contrast to the GWAS of tocotrienols conducted by Lipka et al. (2013), which excluded the few sweet corn lines included in the Goodman–Buckler association panel, two genes involved in kernel starch biosynthesis (*su1* and *sh2*) were found to be associated with tocotrienol traits in our study. Given that *sh2* kernels (~80%) have a higher percentage of moisture relative to *su1* kernels (~75%) arising from differences in sugar and water-soluble polysaccharide content at the fresh-eating stage (Creech, 1965; Soberalske and Andrew, 1978), it was posited that the differences in the percentage of moisture could explain the association of *su1* and *sh2* with tocotrienols. However, the significant associations between these endosperm-expressed genes and the tocotrienol-related traits still remained even after conducting a more stringent GWAS with a mixed linear model

that had a kinship matrix and the first four PCs derived from all 174,996 SNP markers plus fresh kernel weight as a covariate (results not shown). When we considered the JL-QTL mapping results of three tocotrienol-related traits (γT3, total T3, and total T + T3) for grain from the US maize NAM panel, only one of the two sweet corn families ('B73 × P39') that segregate for the *su1* mutation had a significant allelic effect estimate for a JL-QTL with a support interval that included *su1* (Diepenbrock et al., 2017). However, this JL-QTL could not be resolved down to the gene level by GWAS in the NAM panel. Therefore, strong independent evidence is lacking for the implication of *su1* in the genetic control of tocotrienol levels in maize grain. The role of *sh2* in the regulation of tocotrienols, however, could not be assessed by Diepenbrock et al. (2017), because neither of the two sweet corn families segregated for the *sh2* mutation.

Of the six measured tocochromanol compounds in kernels, only γT3 and δT3 were at significantly greater levels ($P < 0.01$) for *sh2* and *su1sh2* lines relative to *su1* lines (Table 2). This could result from the unintentional fixation of causal alleles associated with the increased levels of these two tocotrienols, which might have arisen early in the breeding process if only a limited number of highly related *sh2* lines were used as donor parents (Tracy, 1997). Indeed, *hggt1* (S9_92345469) and *vte1* (S5_131738084) alleles that increased the level of γT3 were fixed in all *sh2* and *su1sh2* lines, as inferred by the peak SNPs, with the exception of one *sh2* line that had the weaker *hggt1* allele (Supplemental Table S9). The involvement of *hggt1* and *vte1* would directly influence γT3 levels, as HGGT1 condenses homogentisic acid and GGDP to produce 2-methyl-6-geranylgeranylbenzoquinol, which, after being methylated to 2,3-dimethyl-5-geranylgeranyl-1,4-benzoquinol by 2-methyl-6-phytyl-1,4-benzoquinol/2-methyl-6-geranylgeranyl-1,4-benzoquinol methyltransferase (VTE3), is then converted to γT3 by VTE1 (Fig. 1). Given that *hggt1* and *vte1* do not explain all of the variation between *su1* versus *sh2* and *su1sh2* for these two tocotrienols (Supplemental Table S8 and Supplemental Table S9), it is likely that one or more of the other nine genes identified to be underlying JL-QTL for γT3 and δT3 levels in maize grain (Diepenbrock et al., 2017) account for the missing heritability. Further exploration to evaluate the genetic contribution of these nine undetected genes and their variant allele frequencies within groups of *su1*, *sh2*, and *su1sh2* lines would be enhanced through the higher statistical power attainable with a larger, more densely genotyped sweet corn association panel.

The genes most strongly associated with both γT3 and δT3 are the two endosperm-expressed genes, *su1* and *sh2*. However, neither of these genes was associated with tocopherol (embryo) traits (Grams et al., 1970; Weber, 1987). If either of these genes is responsible for the greater levels of γT3 and δT3 in kernels of *sh2* and *su1sh2* lines, it is most plausibly driven by the increased sugar content in the endosperm, especially that of the *su1sh2* and *sh2* genotypes that have two- to threefold and seven- to eightfold more sucrose at 20 DAP than *su1* and dent corn genotypes, respectively (Creech, 1965). In leaf tissue from *A. thaliana* plants grown on media supplemented with 3% sucrose, Hsieh and Goodman (2005) observed moderate increases in the expression of MEP pathway genes, including *1-deoxy-d-xylulose 5-phosphate synthase*, which is a gene (*dxs2*) that underlies a major JL-QTL for levels of γT3, δT3 and total T3 (but not tocopherol traits) in grain from the US maize NAM panel (Diepenbrock et al., 2017). Therefore, we hypothesize that the increased sucrose concentration in kernels of *su1sh2* and *sh2* lines stimulates the synthesis of tocotrienols in the endosperm through upregulation of the MEP pathway that provides isopentenyl pyrophosphate for biosynthesis of the tocotrienol tail groups. This could synergistically enhance tocotrienol production in the presence of the strongly expressed *hggt1* allele that is essentially fixed in lines with the *sh2* mutation. Taken together, with the lack of evidence for an association between *su1* and tocotrienols in the US maize NAM panel and the relatively modest accumulation of sucrose, γT3, and δT3 in kernels of *su1* lines, *sh2* becomes the most probable genetic contributor to tocotrienol levels through its production of high sucrose in the endosperm. This hypothesis would be further supported if the association between *su1* and tocotrienols is eventually found to be spurious because of the high interchromosome LD ($r^2 = 0.61$) between *su1* and *sh2*.

Through the implementation of WGP via the GBLUP method, moderate (tocopherols) to moderately high (tocotrienols) predictive abilities were shown for tocochromanol phenotypes in the panel, suggesting that genomic selection could be used to improve genetic gain for tocochromanols and vitamin E in sweet corn breeding programs. The pathway-level and tocochromanol QTL-targeted marker datasets were found to have lower average predictive abilities than those from WGP, although tocochromanol traits are mostly explained by several moderate- to large-effect loci in the NAM panel (Diepenbrock et al., 2017). The most probable explanation for these lower predictive abilities is that the genotyped SNP markers (common variants) at the targeted loci were not in strong LD with causative variants. In support of this theory, the number of significantly associated markers at a genome-wide FDR of 5% from GWAS had a strong, positive correlation with the predictive abilities of tocochromanols. Additionally, 11 of the 14 causal loci controlling grain tocochromanols have significant allelic effect estimates in at least one of the two sweet corn families of the US maize NAM panel (Diepenbrock et al., 2017) but these would have escaped detection in models used for GWAS and WGP if the causal variants at these loci were rarer, weaker effects in the less densely genotyped, smaller sweet corn association panel. These findings are in contrast to the work of Owens et al. (2014), which showed that an eight gene QTL-targeted set is as effective as genome-wide markers for the prediction of the highly oligogenic carotenoid grain traits in the Goodman–Buckler

association panel. However, four of these eight genes were detected via GWAS in the Goodman–Buckler association panel, thus ensuring the capture of large-effect genes that are critical for modifying grain carotenoid composition in the prediction model. Conversely, only 2 of the 14 causal genes underlying QTL associated with grain tocochromanols (Diepenbrock et al., 2017) were identified in the sweet corn association panel.

CONCLUSIONS

We found natural variation for αT (the tocochromanol with the highest vitamin E activity) in sweet corn kernels at the fresh-eating stage to be predominantly under the genetic control of *vte4*, while *vte1* and *hggt1* are involved in controlling the content and composition of tocotrienols. Of the two starch biosynthesis genes found to associate with tocotrienols, the strongest evidence exists for the involvement of *sh2* rather than *su1* in modifying tocotrienol levels. However, additional experiments are needed to develop and evaluate a set of near isogenic lines that capture the different alleles of *su1* and *sh2* in several genetic backgrounds to determine the contribution of these two genes, if any, to heritable differences in tocotrienol levels. The majority of lines with the *sh2* mutation are fixed for alleles at *vte4*, *vte1*, and *hggt1* that collectively increase the levels of αT and γT3. In light of this finding, targeted resequencing and characterization of allelic variation at these three and the other undetected loci previously identified are needed to better assess if the extant sweet corn germplasm pool captures the most favorable variants that exist for maize as a species, especially given that sweet corn experienced a postdomestication genetic bottleneck and recent founder events (Tracy, 1997; Whitt et al., 2002). Whether it be through selection on existing or introgressed allelic variation in breeding programs, our work constitutes an important step for the necessary genomics-assisted breeding efforts to enhance vitamin E to a level that meets or exceeds an RDA of 4.4% for 100 g of fresh sweet corn kernels.

Supplemental Information

Supplemental Table S1: Lambda values used in Box–Cox transformation of 20 fresh kernel tocochromanol traits.

Supplemental Table S2: Transformed best linear unbiased predictors of the 20 fresh kernel tocochromanol traits.

Supplemental Table S3: Back-transformed best linear unbiased predictors of the 20 fresh kernel tocochromanol traits.

Supplemental Table S4: Comparison of genomic prediction models for the presence or absence of two endosperm mutations (*su1* and *sh2*) using marker datasets with different levels of coverage.

Supplemental Table S5: Genomic information (RefGen_v2) for the 81 a priori candidate genes.

Supplemental Table S6: Genomic information (RefGen_v2) for the 14 a priori genes underlying joint-linkage quantitative trait loci associated with grain tocochromanol levels in the US maize nested association mapping panel.

Supplemental Table S7: Statistically significant results from a genome-wide association study of 20 fresh kernel tocochromanol traits.

Supplemental Table S8: Multi-locus mixed-model results from an analysis of tocochromanol traits for chromosomes 3, 4, and 5 with and without endosperm mutation type as a covariate.

Supplemental Table S9: Back-transformed effect estimates for *vte4*, *vte1*, *hggt1*-related SNPs selected with an optimal multi-locus mixed-model.

Supplemental Table S10: Statistically significant results from a genome-wide association study of 20 fresh kernel tocochromanol traits when including endosperm mutation type as a covariate in the mixed linear model.

Supplemental Table S11: Statistically significant results from a genome-wide association study of 20 fresh kernel tocochromanol traits in sweet corn when using endosperm mutation type and the two SNPs selected by multi-locus mixed-models (S5_131738084 and S5_214707875) as covariates in the mixed linear model.

Supplemental Fig. S1: Sources of variation for tocochromanol traits in fresh sweet corn kernels.

Supplemental Fig. S2: Correlation matrix for back-transformed BLUPs of the 20 tocochromanol fresh kernel traits.

Supplemental Fig. S3: Linkage disequilibrium estimates in the sweet corn diversity panel.

Supplemental Fig. S4: Genome-wide association study of 20 fresh kernel tocochromanol traits in sweet corn.

Supplemental Fig. S5: Genome-wide association study for the ratio of α-tocopherol to γ-tocopherol in fresh kernels of sweet corn.

Supplemental Fig. S6: Genome-wide association study for the ratio of δ-tocotrienol to γ-tocotrienol in fresh kernels of sweet corn.

Supplemental Fig. S7: Genome-wide association study for γ-tocotrienol content in fresh kernels of sweet corn.

Supplemental Fig. S8: Genome-wide association study for endosperm mutation type of physiologically mature sweet corn kernels.

Supplemental Fig. S9: Genome-wide association study of 20 fresh kernel tocochromanol traits in sweet corn with endosperm mutation type (*su1*, *sh2*, or *su1sh2*) included as a covariate.

Supplemental Fig. S10: Distribution of the percentage of the recommended daily allowance (RDA) for vitamin E provided by inbred lines from the sweet corn association panel.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

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.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Food and Agriculture; the USDA Hatch under accession numbers 1003971 (M.A.G), 1010428 (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 current and past members of the Tracy and Gore labs for their efforts in pollination, harvest, and sample preparation. We also thank Christine Diepenbrock for providing comments on an earlier version of the manuscript.

REFERENCES

Barger, G.L. 1969. Total growing degree days. Weekly Weather & Crop Bull. 56(18):10.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57:289–300.

Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. J. R. Stat. Soc. Series B Stat. Methodol. 26:211–252.

Boyer, C.D., and J.C. Shannon. 1983. The use of endosperm genes for sweet corn improvement. In: J. Janick, editor, Plant breeding reviews. Springer, Boston, MA. p.139–161.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Cahoon, E.B., S.E. Hall, K.G. Ripp, T.S. Ganzke, W.D. Hitz, and S.J. Coughlan. 2003. Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. Nat. Biotechnol. 21:1082–1087. doi:10.1038/nbt853

Centers for Disease Control. 2006. 2nd national report on biochemical indicators of diet and nutrition in the US population. Centers for Disease Control. https://www.cdc.gov/nutritionreport/pdf/Fat.pdf (accessed 25 Sept. 2018).

Chen, J., and Z. Chen. 2008. Extended Bayesian information criteria for model selection with large model spaces. Biometrika 95:759–771. doi:10.1093/biomet/asn034

Creech, R.G. 1965. Genetic control of carbohydrate synthesis in maize endosperm. Genetics 52:1175–1186.

DellaPenna, D., and L. Mène-Saffrané. 2011. Vitamin E. In: F. Rebeille and R. Douce, editors, Advances in Botanical Research. Elsevier Ltd., Amsterdam, The Netherlands. p. 179–227.

Diepenbrock, C.H., C.B. Kandianis, A.E. Lipka, M. Magallanes-Lundback, B. Vaillancourt, E. Gongora-Castillo, et al. 2017. Novel loci underlie natural variation in vitamin E levels in maize grain. Plant Cell 29:2374–2392. doi:10.1105/tpc.17.00475

Doebley, J., J.F. Wendel, J.S.C. Smith, C.W. Stuber, and M.M. Goodman. 1988. The origin of cornbelt maize: The isozyme evidence. Econ. Bot. 42:120–131. doi:10.1007/BF02859042

Doehlert, D.C., T.M. Kuo, J.A. Juvik, E.P. Beers, and S.H. Duke. 1993. Characteristics of carbohydrate metabolism in sweet corn (*sugary-1*) endosperms. J. Am. Soc. Hortic. Sci. 118:661–666.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:E19379. doi:10.1371/journal.pone.0019379

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

Ford, E.S., R.L. Schleicher, A.H. Mokdad, U.A. Ajani, and S. Liu. 2006. Distribution of serum concentrations of alpha-tocopherol and gamma-tocopherol in the US population. Am. J. Clin. Nutr. 84:375–383. doi:10.1093/ajcn/84.2.375

Gerdes, J.T., and W.F. Tracy. 1994. Diversity of historically important sweet corn inbreds as estimated by RFLPs, morphology, isozymes, and pedigree. Crop Sci. 34:26–33. doi:10.2135/cropsci1994.0011183X003400010004x

Gilmour, A.R.G., B.B. Cullis, R. Thompson, and D. Butler. 2009. Asreml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:E90346. doi:10.1371/journal.pone.0090346

Grams, G.W., C.W. Blessin, and G.E. Inglett. 1970. Distribution of tocopherols within the corn kernel. J. Am. Oil Chem. Soc. 47:337–339. doi:10.1007/BF02638997

Hannah, L.C., M. Giroux, and C. Boyer. 1993. Biotechnological modification of carbohydrates for sweet corn and maize improvement. Sci. Hortic. (Amsterdam) 55:177–197. doi:10.1016/0304-4238(93)90031-K

Hill, W.G., and B.S. Weir. 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33:54–78. doi:10.1016/0040-5809(88)90004-4

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martínez. 2003. Estimating and interpreting heritability for plant breeding: An update. In J. Janick, editor, Plant Breeding Reviews 2. John Wiley and Sons, Hoboken, NJ. p. 9–112.

Hsieh, M.H., and H.M. Goodman. 2005. The *Arabidopsis* IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. Plant Physiol. 138:641–653. doi:10.1104/pp.104.058735

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

Ibrahim, K.E., and J.A. Juvik. 2009. Feasibility for improving phytonutrient content in vegetable crops using conventional breeding strategies: Case study with carotenoids and tocopherols in sweet corn and broccoli. J. Agric. Food Chem. 57:4636–4644. doi:10.1021/jf900260d

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

Jennings, P.H., and C.L. McCombs. 1969. Effects of *sugary-1* and *shrunken-2* loci on kernel carbohydrate contents, phosphorylase and branching enzyme activities during maize kernel ontogeny. Phytochemistry 8:1357–1363. doi:10.1016/S0031-9422(00)85898-7

Kenward, M.G., and J.H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983. doi:10.2307/2533558

Knekt, P., A. Reunanen, R. Jarvinen, R. Seppanen, M. Heliovaara, and A. Aromaa. 1994. Antioxidant vitamin intake and coronary mortality in a longitudinal population study. Am. J. Epidemiol. 139:1180–1189. doi:10.1093/oxfordjournals.aje.a116964

Kruk, J., H. Hollander-Czytko, W. Oettmeier, and A. Trebst. 2005. Tocopherol as singlet oxygen scavenger in photosystem II. J. Plant Physiol. 162:749–757. doi:10.1016/j.jplph.2005.04.020

Kurilich, A.C., and J.A. Juvik. 1999. Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. J. Agric. Food Chem. 47:1948–1955. doi:10.1021/jf981029d

Kushi, L.H., A.R. Folsom, R.J. Prineas, P.J. Mink, Y. Wu, and R.M. Bostick. 1996. Dietary antioxidant vitamins and death from coronary heart disease in postmenopausal women. N. Engl. J. Med. 334:1156–1162. doi:10.1056/NEJM199605023341803

Leth, T., and H. Sondergaard. 1977. Biological activity of vitamin E compounds and natural materials by the resorption–gestation test, and chemical determination of the vitamin E activity in foods and feeds. J. Nutr. 107:2236–2243. doi:10.1093/jn/107.12.2236

Li, Q., X. Yang, S. Xu, Y. Cai, D. Zhang, Y. Han, et al. 2012. Genome-wide association studies identified three independent polymorphisms

associated with alpha-tocopherol content in maize kernels. PLoS One 7:E36807. doi:10.1371/journal.pone.0036807

Linus Pauling Institute. 2015. Vitamin E. Oregon State Univ. http://lpi.oregonstate.edu/mic/vitamins/vitamin-E (accessed 25 Sept. 2018).

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

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

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

Liu, X., X. Hua, J. Guo, D. Qi, L. Wang, Z. Liu, et al. 2008. Enhanced tolerance to drought stress in transgenic tobacco plants overexpressing *VTE1* for increased tocopherol production from *Arabidopsis thaliana*. Biotechnol. Lett. 30:1275–1280. doi:10.1007/s10529-008-9672-y

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits, Sinauer Associates, Inc., Sunderland, MA.

McBurney, M.I., E.A. Yu, E.D. Ciappio, J.K. Bird, M. Eggersdorfer, and S. Mehta. 2015. Suboptimal serum alpha-tocopherol concentrations observed among younger adults and those depending exclusively upon food sources, NHANES 2003–2006. PLoS One 10:E0135510. doi:10.1371/journal.pone.0135510

Mene-Saffrane, L. 2017. Vitamin E biosynthesis and its regulation in plants. Antioxidants (Basel) 7(1):2. doi:10.3390/antiox7010002

Michaels, T.E., and R.H. Andrew. 1986. Sugar accumulation in *shrunken-2* sweet corn kernels. Crop Sci. 26:104–107. doi:10.2135/cropsci1986.0011183X002600010025x

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. Applied linear statistical models, McGraw-Hill, Boston, MA.

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

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. doi:10.1038/ng1847

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

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55. doi:10.1186/gb-2013-14-6-r55

Sattler, S.E., L.U. Gilliland, M. Magallanes-Lundback, M. Pollard, and D. DellaPenna. 2004. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. Plant Cell 16:1419–1432. doi:10.1105/tpc.021360

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464. doi:10.1214/aos/1176344136

Segura, V., B.J. Vilhjalmsson, A. Platt, A. Korte, U. Seren, Q. Long, et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44:825–830. doi:10.1038/ng.2314

Sen, C.K., S. Khanna, and S. Roy. 2006. Tocotrienols: Vitamin E beyond tocopherols. Life Sci. 78:2088–2098. doi:10.1016/j.lfs.2005.12.001

Soberalske, R.M., and R.H. Andrew. 1978. Gene effects on kernel moisture and sugars of near-isogenic lines of sweet corn. Crop Sci. 18:743–746. doi:10.2135/cropsci1978.0011183X001800050012x

Stelpflug, S.C., R.S. Sekhon, B. Vaillancourt, C.N. Hirsch, C.R. Buell, N. de Leon, et al. 2016. An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. Plant Genome 9:1–16. doi:10.3835/plantgenome2015.04.0025

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

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

Tetlow, I.J., M.K. Morell, and M.J. Emes. 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. J. Exp. Bot. 55:2131–2145. doi:10.1093/jxb/erh248

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

USDA. 2018a. National nutrient database for standard reference. Nutrient Data Laboratory, Beltsville Human Nutrition Research Center. https://ndb.nal.usda.gov/ndb/search/ (accessed 25 Sept. 2018).

USDA. 2018b. Vegetables 2017 summary. USDA NASS. http://usda.mannlib.cornell.edu/usda/nass/VegeSumm//2010s/2018/VegeSumm-02-13-2018.pdf (accessed 25 Sept. 2018).

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Wang, H., S. Xu, Y. Fan, N. Liu, W. Zhan, H. Liu, et al. 2018. Beyond pathways: Genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. Plant Biotechnol. J. doi:10.1111/pbi.12889

Weber, E.J. 1987. Carotenoids and tocols of corn grain determined by HPLC. J. Am. Oil Chem. Soc. 64:1129–1134. doi:10.1007/BF02612988

Whitt, S.R., L.M. Wilson, M.I. Tenaillon, B.S. Gaut, and E.S. Buckler. 2002. Genetic diversity and selection in the maize starch pathway. Proc. Natl. Acad. Sci. USA 99:12959–12962. doi:10.1073/pnas.202476999

Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman, and E.S. Buckler. 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733. doi:10.1105/tpc.104.025700

Wolfinger, R., W.T. Federer, and O. Cordero-Brana. 1997. Recovering information in augmented designs, using SAS PROC GLM and PROC Mixed. Agron. J. 89:856–859. doi:10.2134/agronj1997.00021962008900060002x

Xie, L., Y. Yu, J. Mao, H. Liu, J.G. Hu, T. Li, et al. 2017. Evaluation of biosynthesis, accumulation and antioxidant activityof vitamin E in sweet corn (*Zea mays* L.) during kernel development. Int. J. Mol. Sci. 18. doi:10.3390/ijms18122780

Young, T.E., D.R. Gallie, and D.A. DeMason. 1997. Ethylene-mediated programmed cell death during maize endosperm development of wild-type and *shrunken2* genotypes. Plant Physiol. 115:737–751. doi:10.1104/pp.115.2.737

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360. doi:10.1038/ng.546

Zhang, Z., R.J. Todhunter, E.S. Buckler, and L.D. Van Vleck. 2007. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. J. Anim. Sci. 85:881–885. doi:10.2527/jas.2006-656