Supplementary Methods

Statistical modeling and filtering of phenotypic data

To screen the raw high-performance liquid chromatography (HPLC) data for significant outliers, we fitted a mixed linear model for each tocochromanol phenotype in ASReml-R version 3.0 (Gilmour *et al.* 2009). The full model (Equation 1) fitted to the data was as follows:

$Y_{ijklmn} = \mu + check_i + genotype_i + year_j + group \times year_{ij} + genotype \times year_{ij} + tier(year)_{jk} + pass(tier \times year)_{jkl} + range(tier \times year)_{jkm} + plate(year)_{jn} + \varepsilon_{ijklmn}$(Equation 1) in which Y_{ijklmn} is an individual phenotypic observation; μ is the grand mean; $check_i$ is the fixed effect for the B73 check, where it is set to 0 if the genotype is a non-check line; $genotype_i$ is the fixed effect of the ith genotype (non-check line), where it is set as 0 and omitted if the observation is of the B73 check; $year_j$ is the effect of the jth year; group × $year_{ij}$ is the interaction term between the ith group and jth year, where $group_i$ is an indicator variable with two levels that indicates whether the observation is of a B73 check or non-check line; genotype × $year_{ij}$ is the effect of the interaction between the ith genotype (non-check line) and jth year, which is not included in the model for the B73 check; $tier(year)_{jk}$ is the effect of the kth tier within the jth year; $pass(tier \times year)_{jkl}$ is the effect of the lth pass within the kth tier within the jth year; $range(tier \times year)_{jkm}$ is the effect of the mth range within the kth tier within the jth year; $plate(year)_{jn}$ is the effect of the nth HPLC autosampler plate within the jth year; and ε_{ijklmn} is the residual error effect assumed to be independently and identically distributed (i.i.d.) according to a normal distribution with mean zero and variance σ_ε^2, that is ~iid $N(0, \sigma_\varepsilon^2)$. Of these terms, μ, check, and genotype were modeled as fixed effects, while all other terms were modeled as random effects. Studentized deleted residuals (Neter *et al.* 1996) generated by the model were used to remove 147 significant outliers (where an outlier is a single

plot observation for a single trait) at a Bonferroni adjusted significance threshold of $\alpha = 0.05$. With the outlier-screened raw HPLC data set, we generated best linear unbiased estimator (BLUE) values for 1,762 inbred lines across years (Supplementary Table S1) by fitting the full model (Equation 1) in ASReml-R version 3.0 (Gilmour *et al.* 2009). The full model was refitted with genotype as a random effect to generate variance component estimates for the calculation of heritability on a line-mean basis (Holland *et al.* 2003; Hung *et al.* 2012).

Genotype data processing and imputation

The marker genotype imputation approach implemented in Wu *et al.* (2021) was used to generate a high-density single-nucleotide polymorphism (SNP) marker set in B73 RefGen_v4 coordinates for the Ames panel. To construct the target SNP genotype set, unimputed genotyping-by-sequencing (GBS) SNP genotypes scored at 943,455 loci in the Ames panel by Romay *et al*. (2013) were downloaded from CyVerse (ZeaGBSv27_publicSamples_raw_AGPv4-181023.vcf.gz, available at http://datacommons.cyverse.org/browse/iplant/home/shared/panzea/genotypes/GBS/v27), which provided 1,779 GBS samples for 1,493 of the 1,497 lines that had best linear unbiased estimator (BLUE) values for tocochromanol phenotypes. Given that there were 220 lines with more than one corresponding GBS sample having a call rate ≥ 20%, we followed the approach of Wu *et al.* (2021) to merge two or more GBS samples from the same line. Briefly, a stringently filtered SNP set (call rate ≥ 50%, % heterozygosity ≤ 10%, index of panmixia $F_{IT} \geq 0.8$, minor allele frequency ≥ 0.01 and linkage disequilibrium $r^2 \leq 0.2$) of 32,267 SNPs derived from the Romay *et al.* (2013) unimputed marker data set was used to calculate average pairwise identity-by-state (IBS) between multiple samples of the same line using PLINK version 1.9 (Purcell *et al.* 2007).

A total of 19 lines with a mean IBS value < 0.95 for all within-line sample comparisons were removed from the analysis, followed by consensus genotype calling for the remaining 201 lines. Collectively, the final target data set consisted of 443,419 biallelic GBS SNPs scored on a retained 1,462 lines with a call rate ≥ 0.2%, heterozygosity ≤ 10%, and inbreeding coefficient (F) ≥ 0.8. All heterozygous genotype calls were set to missing prior to imputation.

The reference SNP genotype set, which was identical to that constructed in Wu *et al*. (2021), consisted of 14,613,169 SNPs derived from maize HapMap 3.2.1 (Bukowski *et al.* 2018). In BEAGLE v5.0 (Browning *et al.* 2018) with parameters as previously specified in Wu *et al.* (2021), the genotypes at the 14,613,169 SNP loci were imputed based on 443,419 GBS SNPs (target set) in the 1,462 Ames panel lines. This data set of 14,613,169 loci served as the foundation for the subsetting of markers for all of the quantitative genetic analyses conducted in this study.

Expression data set quality control

To verify the quality and integrity of the samples, SNPs were called using the 3′ QuantSeq read alignments and compared to SNP calls from a 942 maize line RNA-Seq data set (WiDiv-942 panel) (Gage *et al.* 2019). In total, 375 lines overlapped with the WiDiv-942 panel, for a total of 430 3′ QuantSeq samples and 54 positive controls. First, 3′ QuantSeq reads were mapped to the B73 RefGen_v4 assembly (Jiao *et al.* 2017) following the HISAT2 mapping protocol indicated above. Duplicate reads were identified and marked using Picard tools MarkDuplicates version 2.20.8 (https://broadinstitute.github.io/picard/). Output was sorted using SAMTools sort version 1.9 and a pileup file created using SAMTools mpileup with BAQ computation disabled (-B) and alignments with a mapQ less than 60 were omitted (-q 60), allowing for only unique alignments

to be processed. Only positions with a base quality of greater than or equal to 20 were included and all insertions and deletions were discarded. Genotype calls were made at a position in an individual if the coverage was at least five reads, but not greater than 500 reads, and the allele made up greater than 3% of the calls at that position in the individual. If more than two alleles passed the coverage and frequency cutoff, the position was scored as heterozygous and set to missing data only when calculating percent identity. After removing positions from the WiDiv-942 SNP matrix that were not called in the 3′ QuantSeq data set, there were 919,074 remaining positions. Percent identity between the same line in the two data sets was calculated by taking the number of positions that had the same genotype call at a position divided by the total number of positions excluding missing data positions in either data set.

Stringent filtering was employed to curate the final expression data set and ensure it contained high quality data. Samples were filtered out based on the following criteria: sampling concerns such as moldy kernels etc. (12 samples removed), number of cleaned reads were below 5 million (1 sample removed), a HISAT2 alignment rate of less than or equal to 65% (17 samples removed), a Pearson's correlation value (r) less than 0.90 with 40 or more samples (3 samples removed), samples that had less than 95% identity when compared to their high confidence WiDiv-942 panel counterpart during genotype confirmation assessment (15 samples removed), and finally removal of samples that had an heterozygosity greater than or equal to 10% (339 samples removed). This final heterozygosity filter was employed to remove samples that were contaminated by spillover during library construction at the Cornell Institute of Biotechnology's Genomics Facility. This stringent heterozygosity filtering was employed to ensure the final data set was free of contaminating reads that may have impacted downstream analysis. The final data set of 784 high confidence, high quality samples included 43 B73 positive control samples and

741 collected field samples of check and noncheck lines. The B73 positive controls were used during data processing for quality control. Only the 741 check and noncheck samples were used for downstream expression data analysis.

Statistical modeling of expression data

To account for the potential effect of different amounts of accumulated heat units on kernel development, growing degree days (GDD, Bollero *et al.* 1996) from pollination to fresh-harvest at the ~23 DAP time point for each ear was included as a model term when calculating BLUE expression values.

For each gene, BLUE values were generated for each of the 664 non-check lines in ASReml-R version 3.0 (Gilmour *et al.* 2009) as follows:

$$Y_{ijklmn} = \mu + check_i + genotype_i + \alpha \times GDD_j + tier_k + block(tier)_{kl} + plate_m + lane_n + \varepsilon_{ijklmn} \quad \text{(Equation 2)}$$

in which Y_{ijklmn} is an individual *rlog*-transformed value; μ is the grand mean; $check_i$ is the fixed effect for the ith check, where it is set to 0 if the genotype is a non-check line; $genotype_i$ is the fixed effect of the ith genotype (non-check line), where it is set as 0 and omitted if the ith observation is of a check line; α is a scalar regression coefficient for the GDD value of ears harvested on the jth day (GDD_j); $tier_k$ is the kth tier; $block(tier)_{kl}$ is the lth block in the kth tier; $plate_m$ is the mth RNA sample plate; $lane_n$ is the nth lane on an Illumina sequencer (minimum unit of the RNAseq run); and ε_{ijklmn} is the residual error effect assumed to be ~iid $N(0, \sigma_\varepsilon^2)$. With the exception of the grand mean, check, genotype and GDD, all terms were fitted as random effects.

Expression data analysis for *vte7*

Given that the *vte7* locus consists of tandemly duplicated genes (Zm00001d006778 and

Zm00001d006779) with high pairwise nucleotide sequence identity (> 99%) in the B73 RefGen_v4 assembly (Jiao *et al.* 2017), the reads pertaining to these two genes were not uniquely mappable with our standard expression abundance determination bioinformatic pipeline. Therefore, to calculate the transcript abundances at the *vte7* locus, the number of read alignments to the two gene models using multi-mapping reads were summed and normalized to counts per million alignments (CPMA) as (total count within both loci/total reads aligned)*1,000,000. Next, the CPMA values were fitted with the Equation 2 model to generate BLUE values, which were then screened for outliers with Studentized deleted residuals (Neter *et al.* 1996) to produce the final *vte7* data set (Supplementary Data Set 3).

Literature cited

Bollero G. A., D. G. Bullock, and S. E. Hollinger, 1996 Soil temperature and planting date effects on corn yield, leaf area, and plant development. Agron. J. 88: 385–390.

Browning B. L., Y. Zhou, and S. R. Browning, 2018 A one-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. 103: 338–348.

Bukowski R., X. Guo, Y. Lu, C. Zou, B. He, *et al.*, 2018 Construction of the third-generation *Zea mays* haplotype map. Gigascience 7: 1–12.

Gage J. L., B. Vaillancourt, J. P. Hamilton, N. C. Manrique-Carpintero, T. J. Gustafson, *et al.*, 2019 Multiple maize reference genomes impact the identification of variants by genome-wide association study in a diverse inbred panel. Plant Genome 12. https://doi.org/10.3835/plantgenome2018.09.0069

Gilmour A. R., B. J. Gogel, B. R. Cullis, R. Thompson, D. Butler, *et al.*, 2009 ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Holland J. B., W. E. Nyquist, and C. T. Cervantes-Martínez, 2003 Estimating and interpreting heritability for plant breeding: an update. Plant Breed. Rev. 22: 9–112.

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

Jiao Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, *et al.*, 2017 Improved maize reference genome with single-molecule technologies. Nature 546: 524–527.

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

Purcell S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, *et al.*, 2007 PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575.

Romay M. C., M. J. Millard, J. C. Glaubitz, J. A. Peiffer, K. L. Swarts, *et al.*, 2013 Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14: R55.

Wu D., R. Tanaka, X. Li, G. P. Ramstein, S. Cu, *et al.*, 2021 High-resolution genome-wide association study pinpoints metal transporter and chelator genes involved in the genetic control of element levels in maize grain. G3. https://doi.org/10.1093/g3journal/jkab059