## Supporting Information

**TITLE:** Simultaneous Dissection of Grain Carotenoid Levels and Kernel Color in Biparental Maize Populations with Yellow-to-Orange Grain

**AUTHORS AND ADDRESSES:**

Mary-Francis LaPorte1, Mishi Vachev1†, Matthew Fenn2†, Christine Diepenbrock1\*

1University of California, Davis, Department of Plant Sciences, Davis, CA 95616.

2Cornell University, Plant Breeding and Genetics Section, School of Integrative Plant Science, Ithaca, NY 14853.

†Indicates equal contribution.

\*For correspondence: [chdiepenbrock@ucdavis.edu](mailto:chdiepenbrock@ucdavis.edu), +1(530)754-0666.

**Table S1** Optimal lambda for calculated BLUEs.

|  |  |
| --- | --- |
| Trait | lambda |
| α-carotene | -0.05 |
| β-carotene | -0.6 |
| β-cryptoxanthin | -0.8 |
| lutein | 0.6 |
| phytofluene | -1 |
| zeaxanthin | 0 |
| zeinoxanthin | -0.35 |
| total carotenoids | 0.55 |
| kernel color | 0.7 |

**Table S2** Transformed and untransformed best linear unbiased estimators (BLUEs) for grain carotenoid traits and visually scored kernel color in 10 families of the US maize nested association mapping (NAM) panel. Uploaded as a separate file due to file size.

**Table S3** Consensus genetic linkage map with resolution of 0.1 centiMorgans (cM) that was used for joint-linkage analysis in this study. Uploaded as a separate file due to file size.

**Table S4** Allelic effect estimates for grain carotenoid and kernel color traits from joint-linkage analysis on transformed BLUEs, in the families analyzed in this study.Uploaded as a separate file due to file size.

**Table S5** Summary of joint-linkage analysis results.Uploaded as a separate file due to file size.

**Table S6** Summary of individual-trait support intervals in joint-linkage analysis results for which one or more support interval bounds did not uplift to RefGen\_v4 coordinates.Uploaded as a separate file due to file size.

**Table S7** Linear model performance for prediction of kernel color given combinations of carotenoid traits. The penalized likelihood criteria Akaike’s and Bayesian Information Criterion (AIC and BIC), Pearson’s correlation coefficients (R2, R2 adjusted), root mean squared error (RMSE), and residual standard deviation (Sigma) are reported as criteria for model selection. The rows are ordered from optimal (top) to least optimal (bottom) model. Trait abbreviations: phytofluene (PHYF), α-carotene (ACAR), β-carotene (BCAR), zeinoxanthin (ZEI), β-cryptoxanthin (BCRY), lutein (LUT), zeaxanthin (ZEA), total carotenoids (TOTCAR), kernel color (KCOL). Family is a categorical variable (factor) that contains a unique identifier for each of the 10 families included in this study.

Shape

Description automatically generated with medium confidence

**Table S8** Performance of the best-performing linear model compared to the Random Forest model. Model performance is defined as the correlation between predicted and test values (Pearson’s correlation coefficient R2), Baseline Error (averaged over 30 replicates for each linear prediction), and Percent Accuracy. The equation for the best-performing linear model is KCOL~Family+ACAR+BCAR+BCRY+LUT+PHYF+ZEA+ZEI+TOTCAR.



**Table S9** Results from the genome-wide association study conducted herein for grain carotenoid and kernel color traits.Uploaded as a separate file due to file size.

**Table S10** List of 58 genes with *a priori* roles in carotenoid biosynthesis and retention.Uploaded as a separate file due to file size.

**Fig. S1** Linkage disequilibrium (LD) estimates in the 10 families analyzed in this study, for GWAS variants (solid black lines) and 250,000 randomly selected variants in the genome (dashed grey lines). These lines depict percentiles, as labeled at right, representing the distribution of variants. LD is depicted in units of squared allele-frequency correlations (*r*2) in pairwise examinations of GWAS or randomly selected variants and variants proximal to those. Physical distance is measured in kilobases.

**Diagram

Description automatically generated**

**Fig. S2** Significant pleiotropic relationships (α = 0.05) between traits at each common support interval. SIs with no significant relationships are not included. Red denotes a negative pleiotropic relationship, and blue denotes a positive pleiotropic relationship. Uploaded as a separate file due to resolution.

**Fig. S3** Importance of each variable calculated using the Random Forest model. Trait abbreviations: phytofluene (PHYF), α-carotene (ACAR), β-carotene (BCAR), zeinoxanthin (ZEI), β-cryptoxanthin (BCRY), lutein (LUT), zeaxanthin (ZEA), total carotenoids (TOTCAR), kernel color (KCOL). Family is a categorical variable, attached to the individual family number label. Results are ranked in order of importance.

**Chart, histogram

Description automatically generated**