



ORIGINAL RESEARCH

Genetic control of kernel compositional variation in a maize diversity panel

Jonathan S. Renk¹ | Amanda M. Gilbert¹ | Travis J. Hattery² |
Christine H. O'Connor^{1,3} | Patrick J. Monnahan^{1,3} | Nickolas Anderson⁴ |
Amanda J. Waters⁴ | David P. Eickholt⁴ | Sherry A. Flint-Garcia⁵  |
Marna D. Yandeu-Nelson² | Candice N. Hirsch¹ 

¹ Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

² Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA

³ Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA

⁴ PepsiCo R&D, St Paul, MN 55108, USA

⁵ United States Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA

Correspondence

Candice N. Hirsch, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, 55108.

Email: cnhirsch@umn.edu

Assigned to Associate Editor Martin Bohn.

Abstract

Maize (*Zea mays* L.) is a multi-purpose row crop grown worldwide, which, over time, has often been bred for increased yield at the detriment of lower composition grain quality. Some knowledge of the genetic factors that affect quality traits has been discovered through the study of classical maize mutants; however, much of the underlying genetic control of these traits and the interaction between these traits remains unknown. To better understand variation that exists for grain compositional traits in maize, we evaluated 501 diverse temperate maize inbred lines in five unique environments and predicted 16 compositional traits (e.g., carbohydrates, protein, and starch) based on the output of near-infrared (NIR) spectroscopy. Phenotypic analysis found substantial variation for compositional traits and the majority of variation was explained by genetic and environmental factors. Correlations and trade-offs among traits in different maize types (e.g., dent, sweetcorn, and popcorn) were explored, and significant differences and meaningful correlations were detected. In total, 22.9–71.0% of the phenotypic variation across these traits could be explained using 2,386,666 single nucleotide polymorphism (SNP) markers generated from whole-genome resequencing data. A genome-wide association study (GWAS) was conducted using these same markers and found 72 statistically significant SNPs for 11 compositional traits. This study provides valuable insights in the phenotypic variation and genetic control underlying compositional traits that can be used in breeding programs for improving maize grain quality.

Abbreviations: BLUP, best linear unbiased prediction; G×E, genotype × environment; GWAS, genome-wide association study; MLM, mixed linear model; MLM, multiple-locus linear mixed model; NIR, near-infrared; PVE, phenotypic variance explained; QTL, quantitative trait loci; RNA-seq, RNA sequencing; SNP, single nucleotide polymorphism; WiDiv, Wisconsin Diversity.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America

1 | INTRODUCTION

Maize (*Zea mays* L.) is a multi-purpose crop with a wide variety of uses including feed, fuel, food, and industrial products. In the United States, the vast majority of grain harvested does not contribute directly to human consumption or food products. In 2019, it was estimated that only ~3.5% of the grain harvested was used for cereals and other food products (USDA–National Agricultural Statistical Service, 2019). In the United States, demand and consumption of maize-based food products has tripled since the 1970s, primarily through tortillas and tortilla chips as a result of the popularity of both Hispanic and gluten-free food products (Scott et al., 2019). Maize is also a staple food crop for many other regions around the globe including Latin America and Africa. To meet increasing market demands in the United States and globally, the focus on grain quality in maize breeding programs has become paramount.

Compositional traits are important factors in determining the overall quality of the grain and, in turn, food production. However, in the United States, much of the focus of maize breeding has been on yield gain while neglecting quality traits through much of the breeding pipeline (Holmes et al., 2019). This is quite evident when looking at kernel composition over time in the United States. Investigating commercial hybrids from 1960 to 1990, Duvick and Cassman (1999) found that grain protein content has steadily decreased, and starch content has increased in a trade-off to select higher yielding hybrids. This change is a negative consequence for the cooking process known as nixtamalization, in which maize kernels are cooked in an alkaline solution to facilitate the removal of the pericarp and are ground into masa dough (Serna-Saldivar & Rooney, 2015). For example, dry matter loss during cooking is a very important trait to commercial producers as it adds to the production cost of making food products. Pflugfelder et al. (1988) found that hybrids with softer crown and endosperm in the kernel as a result of higher starch content have a higher amount of dry matter loss. The reduction in kernel hardness can also cause a higher level of kernel cracking or stress cracks during storage and drying of the grain at harvest, which leads to increased moisture uptake in the kernel (Jackson et al., 1988) with subsequent impacts on final product quality.

To shift the focus of breeding programs to improved quality traits as a primary breeding objective, a deep understanding of the genetic factors that contribute to variation for compositional traits is beneficial. Much of the current knowledge on the genetics of quality traits comes from the study of classical mutants in maize. Some examples of known mutants include *sugary1* (*su1*) (James et al., 1995), *sugary2* (*su2*) (Zhang et al., 2004), and *waxy1* (*wx1*) (Tsai, 1974), which are involved in the starch biosynthesis pathway and are critical for sweet corn breeding. The *opaque2* (*o2*) (Mertz et al., 1964), *floury2*

Core Ideas

- Understanding kernel compositional variation is critical for food-grade corn applications.
- Genetic and environmental factors account for most of the variation in compositional traits.
- A broad range in trait heritabilities was observed across compositional traits.
- Compositional trade-offs are important to consider when conducting multiple-trait breeding.
- Compositional traits are mostly controlled by a large number of small effect loci.

(*fl2*) (Nelson et al., 1965), and *floury1* (Holding et al., 2007) mutants affect the amino acid content and zein protein composition in the kernel. The *o2* mutant has been used in quality protein maize breeding programs to increase lysine and tryptophan content in animal feed from maize, with much of the work being done by the International Maize and Wheat Improvement Center (CIMMYT) (Babu et al., 2005). The use of these mutations has been difficult to deploy because of the negative effects on yield (Prasanna et al., 2001). Furthermore, long-term studies, such as the Illinois Long-Term Selection experiment, on oil and protein content in kernels suggests the genetic control of these traits is quite complex with many small-effect loci contributing to phenotypic variation (Dudley & Lambert, 2004).

Genome-wide association study (GWAS) is a powerful tool that allows researchers to identify regions of the genome that correlate with a trait of interest and determine the genetic control of a trait within a particular germplasm set. The resolution of association mapping depends on the historical recombination that, over time, produces smaller linkage blocks that can be associated with traits (Zhu et al., 2008). A limited number of GWAS have been performed for kernel compositional traits in maize. For example, protein, starch, and oil estimates generated from near-infrared (NIR) spectroscopy prediction equations in the maize nested association mapping population (McMullen et al., 2009; Yu et al., 2008) and a panel of 302 diverse inbred lines (Flint-Garcia et al., 2005) found 26, 21, and 22 quantitative trait loci (QTL) respectively through joint-linkage mapping (Cook et al., 2012). A study of oil and fatty acid composition measured in the AM508 diversity panel of tropical, subtropical, and temperate inbred lines (Yang et al., 2011) revealed 74 QTL underlying these traits (Li et al., 2013). A subsequent study of this same diversity panel detected 247 and 281 statistically significant QTL for 17 proteinogenic amino acid traits in two separate environments (Deng et al., 2017). Finally, the genetic control of carotenoid-related traits was explored in CIMMYT's 380

inbred carotenoid association panel, and 40 QTL were identified for these traits (Suwarno et al., 2015), and Wang et al. (2020) identified *vitreous endosperm 1* (*ven1*) as a major QTL for influencing vitreous endosperm in the kernel.

Although there have been a handful of studies examining compositional traits in large germplasm panels, there is still a need for better understanding of the compositional trade-offs between traits as well as their underlying genetic control in order to breed for enhanced composition important for human consumption and food products. Here, we evaluated the Wisconsin Diversity (WiDiv) population (Hansey et al., 2011; Hirsch et al., 2014; Mazaheri et al., 2019), across five environments and evaluated 16 compositional traits obtained from NIR spectroscopy prediction equations to investigate (a) compositional trade-offs within the population, (b) the role of genotype \times environment (G \times E) interaction in trait variation, (c) how variation partitions differently across sets of germplasm that are adapted to grow in the upper Midwest, and (d) the genetic control of these traits.

2 | MATERIALS AND METHODS

2.1 | Germplasm and genomic data

A subset of 501 lines from the full WiDiv population was used for this study (Supplemental Table S1). These lines all have available whole-genome resequencing data with $\sim 20\times$ sequence depth, and SNPs from these data were previously identified (O'Connor et al., 2020). Briefly, reads were aligned to the B73 v4 reference genome assembly (Jiao et al., 2017), and joint SNP calling was performed using freebayes v1.3.1-17 (Garrison & Marth, 2012).

The SNP calls generated from freebayes were converted to HapMap format with the software package TASSEL version 5.2.64 (Bradbury et al., 2007). Within TASSEL, an overall genetic summary of genotype and site information was obtained. A pairwise genetic distance matrix was calculated by subtracting identity by state from 1 to create a dendrogram that was exported in Newick Tree format, which was visualized using Figtree version 1.4.4 (Cummings, 2014). The resulting dendrogram was visually inspected for consistency with known pedigree relationships and shared heterotic groups. Samples with $>15\%$ heterozygosity and SNPs with $>15\%$ heterozygosity and 25% missing data were removed using the TASSEL filter menu for sites and taxa. This pruning resulted in the removal of six genotypes (PHJ90, Pa778, N534, 4N506, Va22, and PHR63) and 733,462 SNPs, resulting in a 2,412,791 final SNP set (Supplemental File S1).

The SNP calls from the genomic resequencing data were compared with previous RNA-sequencing (RNA-seq)-generated SNPs on the WiDiv association panel (Hirsch et al., 2014; Mazaheri et al., 2019). Pairwise genetic distance matrices

were made for both datasets of common genotypes. A scatterplot comparing the genetic distances of the datasets was used to identify genotypes with inconsistencies between these two data sets (Supplemental Figure S1). Four additional samples (PHR55, B91, PHG86, and B14A) were removed following this criteria. A dendrogram with this final set of genotypes and SNPs was once again generated using the same methods described above (Supplemental Figure S2).

2.2 | Field experimental design and phenotypic data collection

The 501 genotypes were grown in field trials in the summers of 2016 and 2017. All experiments were planted as a randomized complete-block design with two replications at each location. Within replicates, genotypes were further blocked by flowering time. The first block consisted of the earlier flowering lines (flowering at 71–80 d after planting based on a previous planting of this population in St. Paul, MN), while the second block consisted of the later-flowering lines (flowering at 80–87 d after planting). Inbred lines B73 and PH207 were planted as checks within each block, with five entries of each check per block. Plots were grown as single rows (3.35 m long and 0.76 m apart) at $\sim 70,000$ plants ha^{-1} . In the summer of 2016, trials were grown at the Minnesota Agricultural Experiment Station (St. Paul, MN) and the Iowa State University Agricultural Engineering and Agronomy Research Farm (Boone County, IA). The trial in St. Paul was planted on 17 May, and ears were harvested 19–25 October. The trial in Boone County was planted on 16 May, and ears were harvested 16–19 October. In 2017, trials were grown again at the Minnesota Agricultural Experiment Station, the Iowa State University Agricultural Engineering and Agronomy Research Farm, and also at the Bradford Research Center (Columbia, MO). The trial in St. Paul was planted on 11 May, and ears were harvested 24–25 October. The trial in Boone County was planted on 8 May, and ears were harvested on 27 September and 4 October. The trial in Columbia was planted on 15 May, and ears were harvested on 28 September and 12 October.

Ears from 10 representative plants in the middle of each plot were harvested and a random sample of 120 grams of kernels was ground using a Mill Feeder 3170 (Perten Instruments). Ground samples were scanned using a Perten DA 7250 NIR spectroscopy scanner, and global prediction equations developed by Perten Instruments (www.perten.com) were used to predict ash as is, fat as is, Fiber as is, moisture, protein as is, and starch as is. 'As is' is defined as the trait value without correction for grain moisture at the time of the test. These global prediction equations are developed from a mixture of methods and instrumentation from various laboratories making them more robust than equations developed from a single method (D.E. Honigs, Perten Instruments, personal communication,

2020). Local equations were also developed for the Perten DA 7250 NIR scanner for 10 additional traits with wet chemistry run at Eurofins Scientific, Inc. (Des Moines, IA) using AOAC and AOCS standard protocols: ash (AOAC, 2019a), Ankom crude fiber (AOCS, 2020a), crude fat (AOAC, 2019b), crude fiber (AOAC, 2019c; AOCS, 2020b), fructose (AOAC, 2019d), glucose (AOAC, 2019d), N combustion (AOAC, 2019e, 2019g; AOCS, 2020c), Kjeltec nitrogen (hereafter referred to as N Kjeltec; AOAC, 2019f), sucrose (AOAC, 2019d), and total sugars (AOAC, 2019d). Ankom crude fiber was measured by an Ankom fiber analyzer. Each sample was run in triplicate for equation development. These equations were developed using a set of 100 spectrally diverse inbred lines from the WiDiv association panel (Supplemental Table S2). Wet-lab data were calibrated to the Perten DA 7250 NIR scanner via the Honigs regression method within the Perten NIR software (Supplemental Figure S3) (Honigs et al., 1985). Raw data for the six global equations and 10 local equations on the 4,886 plots are contained in Supplemental Table S3. It should be noted that for both the global and local equations, some negative values are observed because these values are predictions from regression equations.

2.3 | Statistical analysis

Statistical analyses were conducted using R version 4.0.2 (R Core Team, 2020). Raw data were plotted for each trait as a whole and by environment to visually identify outliers. Extreme outliers were removed from the dataset if points were beyond five standard deviations from the mean of a given trait, which resulted in the removal of two data points from downstream analyses and from Supplemental Table S3. Each compositional trait was then analyzed using a random-effects model $y = \mu + g + e + r(e) + b[r(e)] + (g \times e) + \epsilon$, where y is the phenotypic response, g is the genotypic effect of an inbred line, e is the environmental effect of the year \times location combination, $r(e)$ is the effect of replicate within environment, $b[r(e)]$ is the effect of block nested within replicate nested within environment, $g \times e$ is the genotype \times environment interaction, and ϵ is the residual. Models were fit with the restricted maximum likelihood method using the R package lme4 (Bates et al., 2014), and no violations were detected while checking model assumptions (independence of residuals, residual normality, and homoscedastic variance of residuals) (Supplemental Figure S4).

Heritability was calculated on a line-mean basis for each trait on an entry-mean basis using the following equation:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{g \times e}^2}{e} + \frac{\sigma_e^2}{re}}$$

where σ_g^2 is the genotypic variance, $\sigma_{g \times e}^2$ is the genotype \times environment interaction variance, σ_e^2 is the error variance, e is the number of environments, and r is the number of replications. Pearson correlations between NIR compositional predictions were calculated using the cor function in R. Correlations were conducted on the data set as a whole and then partitioned out based on heterotic groups and types of corn (e.g., dent, sweet corn, and popcorn). Meaningful correlations were interpreted based on the magnitude of the correlation with a threshold of ± 0.20 , rather than the significance of the correlation as previously suggested (Taylor, 1990; Kozak et al., 2012).

2.4 | Genome-wide association analysis

To investigate whether compositional traits should be analyzed separately for each environment or combined across environments, Spearman rank correlations were calculated between environments. Spearman rank correlations revealed that the environments were significantly correlated with each other ($p < .05$) and thus environments were combined. Best linear unbiased predictions (BLUPs) were extracted as the genotypic random effect from the model for each trait. All BLUPs can be found in Supplemental Table S4.

The HapMap data generated in TASSEL was transformed into numerical format in R version 4.0.2 (R Core Team, 2020) using the package GAPIT version 3 (Lipka et al., 2012) by setting the numerical function flag equal to TRUE. GAPIT was also used to generate principal components to be used as covariates in GWAS (Supplemental Figure S5). The GWAS for the WiDiv BLUPs was conducted in R version 4.0.2 (R Core Team, 2020) using the multiple-locus linear mixed model (MLMM) in the package FarmCPU version 1.02 (Liu et al., 2016). In the model, the first five principal components were included as covariates to account for population structure (Supplemental Figure S5, Supplemental Table S1), and a threshold for initial SNP inclusion into the model was set for each trait by allowing a false entry rate of 1% based on 100 permutations. The SNPs with minor allele frequency < 0.05 were discarded, resulting in a final SNP total for the GWAS analysis of 2,386,666. Optimal bin size and quantitative trait nucleotide number were selected for each trait by selecting ‘optimum’ under the method.bin parameter. The SNPs were deemed statistically significant with a Bonferroni-adjusted p value of $< .05$. A GWAS was conducted using 446 samples (out of the initial 501) after the removal of inbreds with high heterozygosity (6 genotypes), discordance with previous RNA-seq data (4 genotypes), and lack of genomic sequence data (45 genotypes).

FarmCPU was chosen as the preferred method to conduct GWAS for a few reasons. First, a major issue with GWAS

is that many false positives and false negatives appear in the results. To mitigate this issue, kinship and population structure of the association panel are often used as covariates in a mixed linear model (MLM) to control false positives, which, in turn, weakens the signals of the associated SNPs. FarmCPU uses a MLMM that incorporates multiple markers as covariates in a stepwise MLM to limit confounding markers and kinship (Liu et al., 2016). A complete elimination of any confounding from markers and kinship is achieved through the use of a MLMM in two parts in an iterative fashion with a fixed-effects model and a random-effects model, which result in increased statistical power, computational efficiency, and control for both false positives and negatives (Liu et al., 2016). Kaler et al. (2020) conducted a study of simulated soybean [*Glycine max* (L.) Merr.] and maize data comparing different statistical models used in GWAS and found that FarmCPU was able to control for false positives and false negatives and was consistently able to identify statistically significant SNPs closest to known published genes. It was also concluded that FarmCPU is a robust model for GWAS of complex traits in plants (Kaler et al., 2020).

Candidate genes for statistically significant SNP hits corresponded to physical positions based on annotation of the AGPv4 reference assembly of inbred B73 (Jiao et al., 2017). Functional annotations of candidate genes were obtained from MaizeGDB (Andorf et al., 2016), BLAST alignments to *Z. mays* (Madden, 2013), and Pfam domains (Finn et al., 2016).

Phenotypic variance explained (PVE) by the 2,386,666 SNPs used for GWAS was calculated as previously described using the software GCTA version 1.93 (Yang et al., 2010). Briefly, in TASSEL version 5.2.64 (Bradbury et al., 2007), SNPs were forced into a biallelic state and then the HapMap format was exported into PLINK ped and map files. In PLINK version 1.07 (Purcell et al., 2007), the two files were used to create the binary ped files used in GCTA. In GCTA, the genome-based restricted maximum likelihood pipeline was used to calculate the genetic relationship matrix for all SNPs, and PVE was estimated for all 16 traits based on the BLUPs generated from the random-effects model as the phenotypic input file. In genome-based restricted maximum likelihood, SNPs are fit jointly in a random-effect model, and the only estimates of the model are the variance components, which is much smaller than the sample size so there is no overfitting.

2.5 | Code availability

Code for statistical analysis and GWAS can be found at https://github.com/HirschLabUMN/GWAS_composition.

3 | RESULTS AND DISCUSSION

3.1 | Substantial variation for compositional traits exists within the WiDiv association panel

The NIR predictions for 16 traits were collected for 501 inbred lines in the WiDiv association panel grown in two environments in 2016 (Minnesota and Iowa) and three environments in 2017 (Minnesota, Iowa, and Missouri). Visualization and descriptive statistics of raw values for compositional traits of the association panel show substantial variation for each trait roughly following normal distributions (Table 1, Figure 1). The variation in compositional traits across this panel of lines is consistent with many other traits that have been previously measured in the WiDiv association panel such as flowering time (Hansey et al., 2011), tassel morphology (Gage et al., 2018), and stalk characteristics (Mazaheri et al., 2019).

Means, standard deviations, and ranges were calculated from the raw data across all environments (Table 1). Crude fat had the largest coefficient of variation (CV = 0.90) with values ranging from −4.68 to 6.24%. The smallest CV observed was for starch as is (0.03) with a range of 64.52–77.82%. For the global equation traits, protein as is ranged from 5.75 to 21.27%, fiber as is ranged from 1.09 to 4.06%, fat as is ranged from −1.28 to 7.08%, starch as is ranged from 64.52 to 77.82%, and ash as is ranged from 1.06 to 2.19% (Table 1). The ranges for the global equation traits in the WiDiv association panel are consistent with a previous study that analyzed a 282 inbred panel, in which starch, protein, and oil ranged from 59.6 to 70.3%, 11.5 to 17.5%, and 3.1 to 8.2%, respectively (Cook et al., 2012). Compared with a study by Reynolds et al. (2005), who analyzed seven contemporary maize hybrids, the proximate starch content was higher than in WiDiv (79.3–88.0 vs. 64.52–77.82% in WiDiv) and also had a smaller range in protein (8.03–15.6 vs. 5.75–21.27% in WiDiv). This observation is consistent with a general trade-off between starch and protein content that has been observed in maize breeding history (Duvick & Cassman, 1999).

The 501 inbred lines of the WiDiv association panel consist of many different types of maize, including 145 stiff stalk, 143 nonstiff stalk, 80 unknown, 54 mixed, 50 Iodent, 12 tropical, 11 popcorn, five sweet corn, and one flint based on pedigree information (Supplemental Table S1). Given the range of types of maize that comprised the subset of the WiDiv panel included in this study, we investigated the differences in composition between types and heterotic groups. The BLUP-adjusted means were generated from a random effects model for each type and a Tukey's HSD ($\alpha = 0.05$) was used to investigate if statistically significant differences existed for compositional traits among maize types (Table 2). For starch as is, sweet corn had the lowest value at 70.86% and differed

TABLE 1 Summary of traits and heritabilities

Trait	Mean \pm SD	Range ^a	Heritability ^b
Ankom crude fiber	1.83 \pm 0.32	0.43–3.32	0.92
Ash as is	1.49 \pm 0.13	1.06–2.19	0.94
Ash	1.21 \pm 0.25	0.22–2.13	0.89
Crude fat	1.99 \pm 1.80	(–)4.68–6.24	0.79
Crude fiber	2.42 \pm 0.44	1.06–4.1	0.84
Fat as is	4.17 \pm 1.00	(–)1.28–7.08	0.88
Fiber as is	1.99 \pm 0.45	1.09–4.06	0.84
Fructose	0.52 \pm 0.21	0.04–1.06	0.50
Glucose	0.36 \pm 0.18	(–)0.54–0.98	0.54
Moisture	6.31 \pm 3.31	(–)3.64–12.71	0.84
Nitrogen combustion	1.53 \pm 0.33	0.42–2.68	0.94
Nitrogen Kjeltac	1.52 \pm 0.33	0.43–2.61	0.94
Protein as is	12.32 \pm 2.17	5.75–21.27	0.94
Starch as is	72.01 \pm 2.43	64.52–77.82	0.75
Sucrose	2.65 \pm 1.01	0.82–4.76	0.44
Total sugars	3.00 \pm 1.13	(–)1.69–6.05	0.51

^aNegative values are observed as a result of the fact that these values are predictions from regression equations.

^bHeritability is calculated on a line-mean basis.

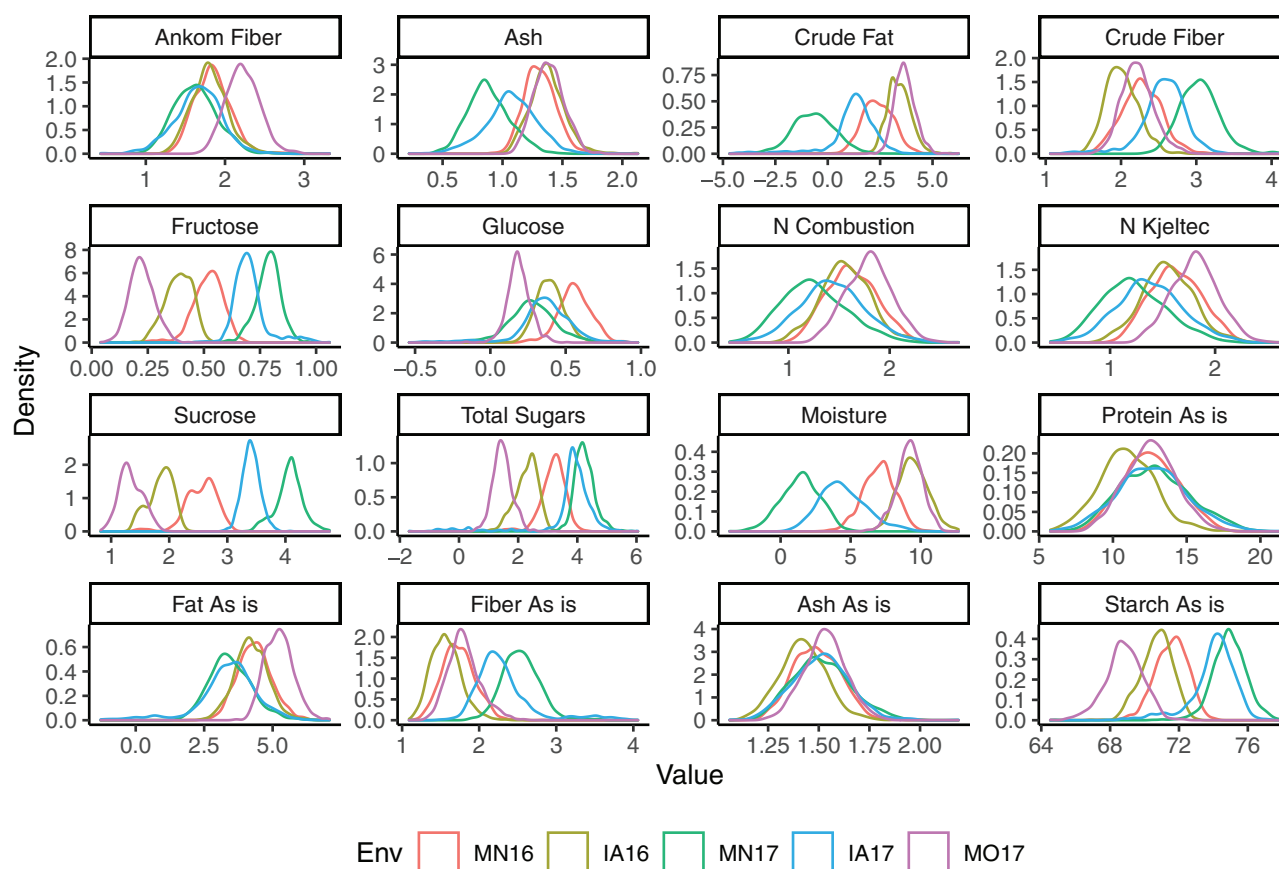


FIGURE 1 Trait distributions. All 16 near-infrared prediction trait raw value distributions in each of the five unique environments in 2016 and 2017. The density plot lines are colored according to the corresponding environment. Env, environment; MN16, St. Paul, MN, 2016; IA16, Ames, IA, 2016; MN17, St. Paul, MN, 2017; IA17, Ames, IA, 2017; MO17, Columbia, MO, 2017

TABLE 2 Summary of best linear unbiased prediction (BLUP)-adjusted mean differences for each compositional trait compared among maize types

Trait	Stiff stalk	Non-stiff stalk	Unknown	Iodent	Mixed	Popcorn	Tropical	Sweet corn	Flint
Ankom crude fiber	1.76c ^a	1.85c	1.90c	1.83c	1.85c	2.19a	1.97bc	2.15ab	2.10abc
Ash	1.17ab	1.23a	1.24a	1.18ab	1.24a	1.28a	1.27a	1.32a	1.43a
Ash as is	1.46b	1.51a	1.54a	1.46ab	1.51a	1.62a	1.58abc	1.69a	1.74a
Crude fat	1.92b	2.06ab	2.15ab	1.86b	2.05ab	2.17ab	2.41a	1.99ab	2.08ab
Crude fiber	2.39c	2.43bc	2.43bc	2.46b	2.42bc	2.68a	2.44bc	2.58ab	2.41bc
Fat as is	4.06b	4.16ab	4.35ab	3.81b	4.20ab	4.50ab	4.65a	4.77a	4.58ab
Fiber as is	1.95d	1.98cd	2.02bc	2.04bc	1.98cd	2.28a	2.02bcd	2.17ab	1.92d
Fructose	0.53a	0.52abc	0.52bc	0.53a	0.52ab	0.50d	0.52bc	0.51cd	0.51bcd
Glucose	0.35bc	0.36b	0.35c	0.38a	0.36bc	0.33c	0.35bc	0.35c	0.35bc
Moisture	6.45ab	6.23bc	6.08cd	6.57a	6.18bcd	5.57d	5.88cd	5.75cd	5.45d
N combustion	1.47ab	1.58a	1.63a	1.43ab	1.59a	1.70a	1.71a	1.78a	2.05a
N Kjeltec	1.45ab	1.56a	1.61a	1.42ab	1.57a	1.69a	1.69a	1.75a	1.99a
Protein as is	11.80ab	12.63a	13.04a	11.61ab	12.74a	13.51a	13.64a	14.47a	15.67a
Starch as is	72.08a	72.03a	71.85a	72.11a	72.02a	71.79a	71.83a	70.86b	72.37a
Sucrose	2.64b	2.64b	2.63b	2.67a	2.64b	2.64b	2.63b	2.65ab	2.64b
Total sugars	2.98bc	2.99b	2.94c	3.09a	2.99bc	2.92c	2.97bc	2.89c	2.95bc

^aType-adjusted means followed by the same letter within a trait are not statistically different according to Tukey's HSD ($p < 0.05$).

statistically from all other types. This can be attributed to the physiological differences among sweet corn kernels vs. dent kernels, in which sweet corn kernels have been bred to have increased sugar content and less starch content at maturity (Tracy, 2010). Popcorn had the highest amount of Ankom crude fiber, crude fiber, and fiber as is and differed statistically from the dent germplasm, that is, stiff stalk, nonstiff stalk, and Iodent (Table 2). The majority of the fiber in corn kernels is found in the pericarp, which is the outermost layer that encases the entire kernel except the tip cap and is comprised largely of hemicelluloses including xylans and arabinoxylans (Chateigner-Boutin et al., 2016; García-Lara et al., 2019). In a study comparing anatomical percentages of kernel anatomy between dent, flint, and popcorn, popcorn had the highest proportion of pericarp at 7% followed by flint at 6.5% then dent at 6.0% (Serna-Saldivar, 1996). Stiff stalk and Iodent types differed statistically for crude fat from tropical and for fat as is from tropical and sweet corn types, with the lowest values compared with the other maize types. In the kernel, the majority (~85%) of fatty acids are found in the embryo (Shen & Roesler, 2017). This finding suggests that stiff stalk and Iodent types may have been selected and bred over time to have a smaller proportion of germ (i.e., embryo) and less protein in the endosperm, though not significantly different, which would increase starch content at the cost of protein within the endosperm compared with other maize types. No statistically significant differences between BLUP adjusted means for ash, nitrogen combustion, nitrogen

Kjeltec, and protein as is were observed across maize types (Table 2).

3.2 | Genetic and environmental factors account for most of the variation in compositional traits

Given the high degree of variation observed for all of the compositional traits, we next sought to determine the sources of the variation. The proportion of variance explained by each term from the random effects model was analyzed, and the environment term, which combines year and location, was statistically significant (p value $< .001$) for all traits and accounted for 0.02–84.65% of the total observed variation, with a median of 39.59% (Figure 2; Supplemental Table S5). The environment term explained the majority of the phenotypic variance for ash, crude fat, crude fiber, fiber as is, fructose, moisture, starch as is, sucrose, and total sugars. To investigate the role of a location or year effect on compositional traits, the environment term was partitioned into location and year effects within the random effects model. The location effect ranged from 0 to 72.93%, the year effect ranged from 0 to 4281%, and the location-within-year effect ranged from 0 to 77.04% (Supplemental Figure S6). The location effect was the predominant source of variation for Ankom crude fiber, crude fat, fat as is, fructose, and total sugars (Supplemental Figure S6). The year effect explained the most phenotypic

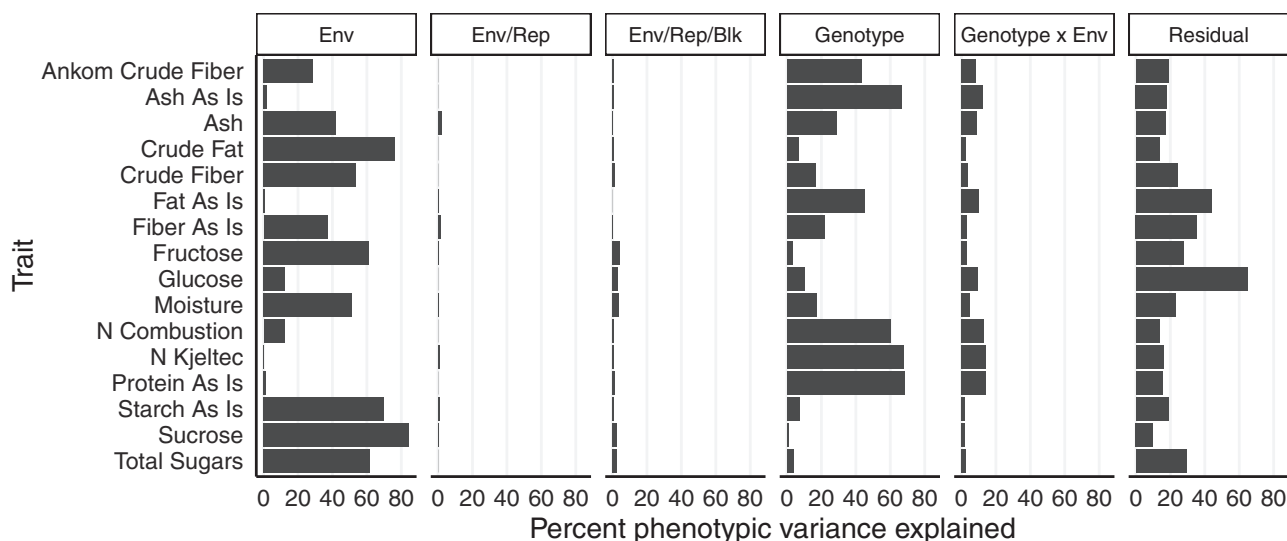


FIGURE 2 Percentage of phenotypic variance explained by each factor from the ANOVA of the random effects model for 16 compositional traits collected from the Wisconsin Diversity association panel. Env, environment; Rep, replication; Blk, block

variance for glucose. The location-within-year effect was the largest for crude fiber, fiber as is, starch as is, and sucrose. For many of the carbohydrate compounds analyzed, the location was the driving factor of the environmental effect, which can also be seen in the distribution of the raw phenotypic data (Supplemental Figure S6). Of the five unique environments, the MO 2017 environment, which is the most geographically and climatically distinct environment, was the lowest for all carbohydrate traits. Previous studies have also found a statistically significant effect of the environment on compositional traits. For instance, oil concentration in the grain has been shown to be greatly influenced by geographic location and planting year (Dunlap et al., 1995; Jellum & Marion, 1966). Likewise, a study of seven maize hybrids grown in Europe analyzing compositional traits found 1,986 statistically significant differences out of 4,935 comparisons between the four distinct environments included in the study, highlighting the importance of environment effects for these traits (Reynolds et al., 2005). These results highlight that environmental factors, such as location and year effects, need to be taken into consideration for breeders to meet breeding targets for compositional traits.

The genotype and G×E interaction terms were also statistically significant for all traits ($p < .001$) and ranged from 1.07 to 68.06% (median 19.49%) and 1.77 to 14.19% (median 6.48%), respectively (Figure 2; Supplemental Table S5). In contrast, a previous study analyzing a different maize association panel found kernel moisture and starch content had a nonsignificant G×E interaction, while oil and protein content did have a statistically significant G×E interaction (Flint-Garcia et al., 2005). However, in that study the authors noted that many environments had only a single replication, which caused the G×E interaction to be confounded with the resid-

ual term in the mixed model, which would affect significance testing. While the G×E interaction was not the main source of phenotypic variation, it was still a statistically significant source of variation for all traits. Breeders have options for dealing with G×E interactions: choose to ignore these interactions and instead choose varieties that are superior when averaged across environments, try to reduce these effects by testing varieties in similar environments, or try to exploit these interactions by selecting varieties that excel in a particular environment (Bernardo, 2010). For Ankom crude fiber, ash as is, fat as is, nitrogen combustion, nitrogen Kjeltec, and protein as is, the genotype term in the model explained the most phenotypic variation, with 43.35–68.06% of the variation explained. While for ash, crude fiber, fiber as is, and moisture intermediate values of variation explained by genotype were observed ranging from 16.43 to 28.82% or low values for the carbohydrate traits ranging from 1.07 to 10.00% (Figure 2). The approach for dealing with G×E interactions will likely need to be different for these different classes of traits.

Finally, the residual error ranged from 9.90 to 64.69% of variance explained and accounted for the highest variance explained for glucose, which had only 10.00% of the variation explained by genotype (Figure 2). The high residual variance for some traits may reflect the lower quality of the local prediction equations for those traits (Supplemental Figure 3) and also the detection limits of analytical techniques specifically in the case of total sugars, glucose, fructose, and sucrose (<0.15%; Supplemental Table S2). Still, a sizable proportion of variance explained was due to residual variation for the rest of the compositional traits. Some of these factors could be epistatic effects or micro-environmental factors that were not taken into consideration in our random effects model.

3.3 | Heritabilities of compositional traits and implications for single-trait breeding targets

Trait heritability is an important consideration in determining the ability to make gains from selection for a particular trait by a breeder. Traits that have a high heritability have better response to selection (Falconer & Mackay, 1996), and understanding the heritability for compositional traits in the breeding population helps breeders plan where to invest or divest resources to make genetic gains for the desired breeding target. Heritabilities on a line-means for the diversity panel ranged from 0.44 (sucrose) to 0.94 (ash as is, protein as is, nitrogen combustion, and nitrogen Kjeltect) (Table 1). With the exceptions of fructose, glucose, sucrose, and total sugars, the compositional traits measured in this study all had relatively high heritabilities (>0.75) (Table 1), indicating a larger genotypic variance than G×E and residual variances that were measured in multiple environments. A previous study evaluating heritability in a panel of diverse inbred lines (Flint-Garcia et al., 2005) also found high heritabilities for protein, oil, and starch from NIR predictions that ranged from 0.83 to 0.91 (Cook et al., 2012). The lower heritabilities seen in the sugar-related traits in our study are likely a reflection of a lower quality prediction model ($r = 0.260\text{--}0.649$; Supplemental Figure S3) than the global predictions equations but could also, to some extent, reflect true biological differences in the degree of environmental control of these traits.

The observed variation for compositional traits within the WiDiv association panel (Figure 1) and the relatively high heritabilities (Table 1) are critical for breeders to make gains from selection for quality traits within their breeding programs. The high heritabilities for these NIR predictions indicate that NIR is a suitable, high-throughput, and affordable method, compared with wet lab techniques, to quickly evaluate germplasm within a breeding program and obtain robust data for making breeding decisions. Ten of the 16 traits showed that the environment was the predominant source of phenotypic variation in the random-effects model. For these traits, potential improvement could be made by breeding for target environments or growing and sourcing from certain environments that are more favorable for specific traits of interest. Ankom crude fiber, ash as is, nitrogen combustion, nitrogen Kjeltect, and protein as is had the majority of the phenotypic variation partitioned to the genotype term in the random effects model and had very high heritabilities with an average of 0.94 (Table 1). For these traits, breeders can make genetic gains through selection that can be deployed across more diverse environments.

3.4 | Correlations between compositional traits among types of maize and implications for multiple-trait breeding

Phenotypic correlations were detected among the WiDiv association panel as a whole (Supplemental Table S6). The strongest correlations were between the various protein and nitrogen measurements, including nitrogen Kjeltect, nitrogen combustion, and protein as is ($r = 0.99\text{--}1.00$). Nitrogen combustion is collected via the Dumas method (Dumas, 1831), which determines total nitrogen content (organic and inorganic nitrogen) by combusting a sample at a high temperature in an oxygen-rich environment to convert the nitrogen in the sample into nitrogen gas. Nitrogen Kjeltect is measured with the Kjeldahl method (Kjeldahl, 1883), which determines the total organic nitrogen content of a sample through a sulfuric acid digestion. Protein as is is then calculated by a protein conversion factor of 6.25 for maize to estimate the total protein content in a sample. The fact that these traits are highly correlated is not particularly surprising, as they are measuring the same compound in different ways, but gives confidence to the quality of the data.

By investigating phenotypic correlations of distinct compounds, we found that protein as is was negatively correlated with starch as is ($r = -0.23$) (Supplemental Table S6). The negative trade-off between protein and starch throughout the history of maize breeding is well known. Breeding for increased yield in maize has driven an increase in starch content over time with an overall decrease in protein content (Duvick & Cassman, 1999). This is because a lower energy threshold is required of the plant to make starch since it is a primary product of photosynthesis, while protein requires more energy and relies on external factors such as nitrogen availability in the soil (Jenner et al., 1991). In a previous study by Cook et al. (2012) a negative correlation was found between NIR-predicted starch and protein ($r = -0.56$) in a panel of 282 diverse inbred lines. We also found a negative phenotypic correlation between starch and fat ($r = -0.46$) and a positive correlation between protein and fat ($r = 0.54$), both of which were consistent with Cook et al. (2012). In the maize kernel, fat content is primarily found in the germ, and starch and protein are found in the endosperm. The endosperm is roughly divided into the floury endosperm, which contains mostly starch and is located in the interior of the kernel, and the vitreous endosperm surrounding it, which contains much of the protein found in the kernel (Holmes et al., 2019). For the purpose of food-grade maize, a shift in the proportion of floury endosperm to vitreous endosperm, or a larger germ relative to the endosperm, will be needed to increase the relative amount of protein and fat in the kernel.

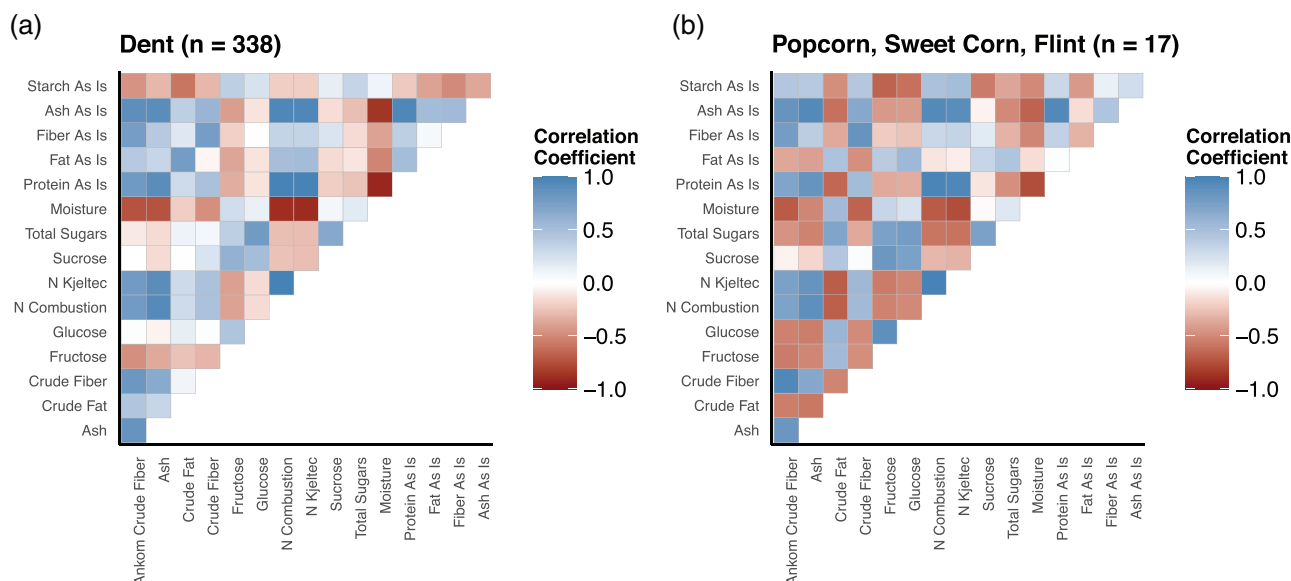


FIGURE 3 Heatmaps of Pearson correlation matrices for compositional traits partitioned by (a) dent, and (b) popcorn, sweet corn, and flint

Breeding efforts for different applications (e.g., food-grade corn, sweet corn, popcorn, etc.) are focused on different needs based on the product market, which have resulted in very different trait means (Table 2) and potentially different correlations among traits within these groups of germplasm. Pearson correlations were recalculated after separating the lines into dent (stiff stalk, nonstiff stalk, and Iodent) vs. popcorn, sweet corn, and flint since these two groups showed the most statistically significant differences in BLUP-adjusted means between traits (Table 2). Overall, more meaningful pairwise correlations were detected for popcorn, sweet corn, and flint than in the dent germplasm (Figure 3). The overall pattern of pairwise correlations were relatively similar, but the changes in magnitude and direction were substantial in some instances (Figure 3). Starch as is had weak positive correlations with glucose, fructose, and sucrose ($r = 0.23$, 0.37 , and 0.14 , respectively) in pooled dent germplasm, but in popcorn, sweet corn, and flint lines these correlations changed direction and were stronger negative correlations ($r = -0.61$, -0.66 , and -0.56 , respectively) (Figure 3). Similar changes in correlation patterns were also observed after separating lines into dent and nondent for correlations between crude fat and numerous other traits (Ankom crude fiber, ash, fructose, nitrogen Kjeltetec, and nitrogen combustion) as well as the correlation of crude fiber with both moisture and protein. The correlation between protein and crude fat was positive ($r = 0.28$) in the dent germplasm but was reversed to a negative relationship ($r = -0.65$) in the nondent germplasm.

Breeding for a single compositional trait can be a difficult task, given controlling genetic and environmental factors, but when improving multiple traits simultaneously in a breeding

program, the task becomes even more complex and difficult. Phenotypic correlations in the full WiDiv association panel and parsing the panel into dent and nondent types revealed that there are many phenotypic trade-offs that need to be assessed by a breeder in order to make desired genetic gains. For instance, if a breeder has a program goal of improving quality, certain phenotypic correlations may work for or against the breeding target. Protein in the dent germplasm pool was negatively correlated with starch ($r = -0.23$), but positively correlated with ash, fiber, and fat ($r = 0.96$, 0.52 , and 0.53 , respectively) (Figure 3). Selections for increasing protein content could then also translate to gains in ash, fiber, and fat but a decrease in starch content that would meet the breeding targets of food-grade maize. These compositional targets help with physiological aspects of the kernel, such as kernel damage, stress cracks, and higher protein content, that impact processing. There are a few methods for multiple-trait selection such as tandem selection, independent culling, and index selection (Bernardo, 2010). In tandem selection, a breeder selects for a single trait at a time until breeding targets have been met. This method can be time consuming but can be efficient if traits are correlated with each other in a way that meets the breeding targets. Independent culling, on the other hand, selects for multiple traits in a single breeding cycle with minimum threshold levels determined for each trait, and entries that are above these thresholds continue through the breeding pipeline. Index selection selects for multiple traits simultaneously using an index value calculated by the breeder. The index can weight traits based on economic value alone, genetic and phenotypic covariances with economic values, and performance (Elston, 1963; Hazel, 1943; Smith, 1936; Williams, 1962).

Index selection is a more effective method than both tandem selection and independent culling while having the ability to have multiple traits imputed into an index. Index selection can also be tailored to a breeding program with multiple methods from which a breeder can choose. For instance, a food-grade maize breeder with the goal of increased quality would put higher weights on protein, fiber, and fat content with a smaller weight on starch content. Knowledge of how compositional traits interact with each other at a phenotypic level is important for making genetic gains in a breeding program, and understanding the genetic control of these traits can help accelerate breeding goals for improved quality in maize grain.

3.5 | Compositional traits are mostly controlled by a large number of small-effect loci

To determine the genetic control of our compositional traits we used a GWAS approach, as this population was originally developed for this purpose (Hansey et al., 2011). For this analysis, we used a set of 2,386,666 SNPs that had been previously called from whole-genome resequencing data (O'Connor et al., 2020). Principal component analysis using these SNPs showed clear separation of the three major maize dent heterotic groups: stiff stalk, Iodent, and nonstiff stalk (Figure 4a). The patterns observed in the principal component analyses are similar to other published results (Romas et al., 2013; van Heerwaarden et al., 2012; White et al., 2020) describing the separation of maize types. The clear separation of maize types from the whole-genome resequencing data gave us confidence in the genotypic data and provided covariates to use in the GWAS to account for population structure.

For all 16 compositional traits in this study, the genetic control was mapped using 446 inbreds with resequencing data using the MLM in FarmCPU (Liu et al., 2016) with 2,386,666 SNPs. A total of 72 statistically significant SNPs across 11 of the 16 traits were detected based on a genome-wide corrected Bonferroni threshold of $-\log_{10}(p) = 7.68$ (Figure 4b; Supplemental Figure S7, Supplemental Table S7). The traits with statistically significant GWAS SNP associations included Ankom crude fiber (7), ash (7), ash as is (5), crude fiber (7), fat as is (3), fiber as is (5), fructose (5), nitrogen combustion (8), nitrogen Kjeltex (10), protein as is (9), and starch as is (6) (Figure 4b). In a previous study conducted by Cook et al. (2012) using NIR-predicted data for starch, protein, and oil, no significant associations were found using a MLM method for GWAS, which is contrary to our study in which starch, protein, and fat from the global equations had statistically significant SNPs. The differences between the two studies could be due to a few factors, such as having more individuals in the WiDiv association panel, a more balanced allele frequency profile because the WiDiv is less genetically diverse than the 282 panel, a much larger SNP dataset (43

times as many markers), and a more powerful GWAS model to detect significance for these complex compositional traits.

As we had multiple traits in this study that measured the macromolecules in different ways for protein, starch, and ash, we decided to first look at the overlap in hits within these sets of traits within 50 kb. There were eight instances of overlap between compositional traits (Figure 4b; Supplemental Table S7), including SNPs on chromosomes 1, 2, 4, and 5 between nitrogen combustion, nitrogen Kjeltex, and protein as is; SNPs on chromosome 2 between ash, ash as is, nitrogen combustion, and nitrogen Kjeltex; SNPs on chromosome 3 for Ankom crude fiber and ash; SNPs on chromosome 5 between ash as is, nitrogen combustion, nitrogen Kjeltex, and protein as is; and SNPs on chromosome 9 between nitrogen Kjeltex and protein as is (Figure 4b). While these overlaps are observed, this does not necessarily indicate that the same causative variant is underlying variation across each of these traits, and it is possible that this observation is the product of variants that are in close physical proximity to each other. However, for the shared statistically significant SNPs between nitrogen combustion, nitrogen Kjeltex, and protein as is, the same exact SNPs was identified. This is not unexpected as these traits are highly correlated ($r = 1.00, 0.99, 0.99$, respectively) with each other and are measuring nitrogen content in slightly different fashions as described above (Supplemental Table S6). While the nitrogen-related compositional traits shared many SNP hits, the fiber and carbohydrate-related traits all gave different statistically significant SNPs. This is not particularly surprising, as fiber can be measured in multiple ways reflecting different compounds within the fiber fraction (Mertens, 1992). This is exacerbated when comparing the local equations with the Perten global equations. The local equations were developed with a single method, while the global equations were developed using a mixture of methods that integrated data from different laboratories (D.E. Honigs, Perten Instruments, personal communication, 2020), which may reduce signal gained from using equations developed from a single method and measuring a specific aspect of fiber.

Across the 72 statistically significant SNPs, 48 were located within a gene in the B73v4 annotation, and the remaining 24 were within 18 kb of a gene model (Supplemental Table S7). Functional annotations for the genes nearest each statistically significant SNP were determined by using a series of databases (see Materials and Methods section). Only 13 statistically significant SNPs yielded uncharacterized protein annotations from the combined gene annotation methods (Supplemental Table S7). A notable association for Ankom crude fiber includes a WUSCHEL-related homeobox gene on chromosome 8, which has previously been shown to be associated with maize embryo development (Nardmann et al., 2007). For fat as is, a bzip transcription factor 111 previously found to be involved in the carbon–nitrogen balance regulation (Wu et al., 2019) was identified as a candidate gene. While these

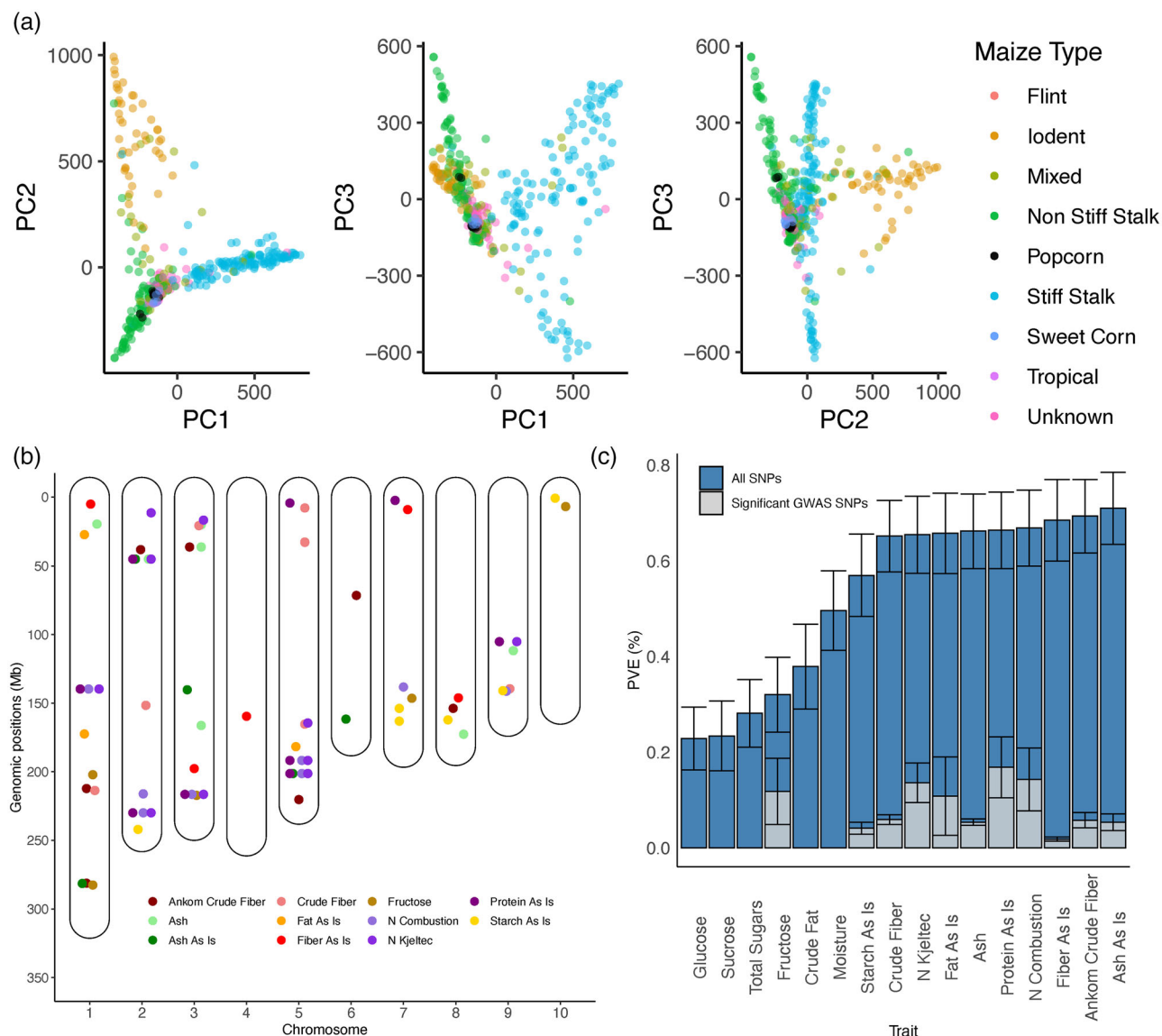


FIGURE 4 Genetic control of compositional traits in a maize diversity panel. (a) Pairwise plots of the principal components (PCs) of population structure using 2,386,666 single nucleotide polymorphisms (SNPs) for the 446 inbred lines used for genome-wide association studies (GWAS). Each point is colored based on maize type classification. (b) Statistically significant GWAS SNP associations of compositional traits. Each point represents the physical location of a statistically significant SNP on the chromosome. Points are colored corresponding to a particular compositional trait. (c) Percentage variation explained (PVE) by the full set of 2,386,666 SNPs used in the GWAS are colored in blue, while the gray corresponds to significant SNPs that were specific for each trait based on GWAS results. Error bars represent the standard error

results provide an exciting resource to breeders to make quality gains in maize, and this study identified more statistically significant SNPs than previous GWAS studies of compositional traits in maize (Cook et al., 2012), the SNPs identified in this study do not explain the full breadth of phenotypic variation in this population.

To gain more insight into the underlying genetic control of these compositional traits, we estimated the PVE by the full set of 2,386,666 SNPs generated from the whole-genome resequencing. The percentage of PVE ranged from 22.88 to 70.98%, with glucose having the smallest amount and ash as is having the most (Figure 4c). The PVE was also calculated

for just the significant SNPs identified for each trait using only those SNPs that were significant in the GWAS for the trait. The PVE for these significant GWAS SNPs was substantially lower and ranged from 1.84 to 16.86%, with fiber as is having the smallest PVE from the significant SNPs and protein as is having the most (Figure 4c). The carbohydrate traits tended to have the lowest amount of variation explained, ranging from 22.88 to 57.00% (Figure 4c) vs. the other traits. This could be due to the accuracy of the NIR prediction equations used to generate values for the sugars or reflects the large role that the environment contributes to these traits (Figure 2). Nine of the 16 traits had relatively large (>60%) PVE attributed

to the full set of SNPs, indicating that many loci of small effect contribute to variation of compositional traits and did not meet the threshold of significance rather than a few large effect SNPs identified in the GWAS. A previous GWAS study for kernel compositional traits found no significant SNPs in their association panel Cook et al. (2012) but postulated if the population size of the diversity panel increased as well as the SNPs density it would be easier to detect small-effect loci. This result is consistent with results found in the Illinois long-term selection experiment in a QTL study for oil concentration, where it was estimated that more than 50 loci may be involved in controlling the trait (Laurie et al., 2004). The upper limits for protein and oil have not been reached in the Illinois long-term selection experiment even after over 100 generations of selection (Lucas et al., 2013), giving reason to believe many of these small-effect loci can be used to make gains for improved quality grain for these and other compositional traits. This result brings more complexity to the challenge for breeding to improve for quality traits and the cost of marker data across many small-effect loci. Multiple strategies are available to breeders today to try and make genetic gain for compositional traits including the production of doubled-haploid populations, marker-assisted selection, genomic prediction, and recurrent selection programs in order to harness these small-effect loci.

4 | CONCLUSIONS

In summary, we collected NIR compositional traits on a set of 501 inbred lines adapted to the upper Midwest grown in five unique environments. The WiDiv association panel showed substantial variation for compositional traits, and it was shown that genetic and environmental factors account for most of the phenotypic variation observed. Statistically significant differences in means among types of maize and meaningful correlations between compositional traits were found to reflect the diversity of maize types and the underlying anatomical and structural makeup of the kernels. A total of 2,386,666 SNPs derived from whole-genome resequencing data was used to conduct a GWAS on 446 individuals from the WiDiv association panel, which resulted in 72 statistically significant SNPs detected. The proportion of variance explained by all of the SNPs showed that many compositional traits are controlled by small-effect loci. Insights into the underlying genetic control of these traits and how they are correlated will help breeders in implementing breeding strategies for quality enhancement breeding targets that have been largely neglected to date.

CONFLICT OF INTEREST

NA, AJW, and DE are employed by PepsiCo, Inc., a goods and beverage company that sources food grade corn. The views

expressed in this manuscript are those of the authors and do not necessarily reflect the position or policy of PepsiCo, Inc.

AUTHOR CONTRIBUTIONS


CNH, MYN, SFG, DE, AW, NA conceived this experiment. JSR, AMG, TJH conducted the experiments, JSR, CHO, PJM analyzed the data, JSR visualized the data, JSR, CNH wrote the original draft, all co-authors edited and approved the final manuscript.

ACKNOWLEDGMENTS

This work was funded in part by NSF IOS-1546272 to CNH and MDY-N, PepsiCo, Inc. to CNH, the Iowa Agriculture and Home Economics Research Station Project IOW03649 to MDY-N, and USDA-ARS base funds to SF-G. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported in this paper.

ORCID

Sherry A. Flint-Garcia  <https://orcid.org/0000-0003-4156-5318>

Candice N. Hirsch  <https://orcid.org/0000-0002-8833-3023>

REFERENCES

- Andorf, C. M., Cannon, E. K., Portwood, J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Braun, B. L., Campbell, D. A., Vinnakota, A. G., Sribalasu, V. V., Huerta, M., Cho, K. T., Wimalanathan, K., Richter, J. D., Mauch, E. D., Rao, B. S., Birkett, S. M., Sen, T. Z., & Lawrence-Dill, C. J. (2016). MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Research*, 44, D1195–D1201. <https://doi.org/10.1093/nar/gkv1007>
- AOAC. (2019a). *Ash of animal feed* (AOAC 942.05-1943). AOAC International.
- AOAC. (2019b). *Fat (crude) or ether extract in animal feed* (AOAC 920.39-192). AOAC International.
- AOAC. (2019c). *Fiber (crude) in animal feed and pet food – Ceramic fiber filter method* (AOAC 962.09-1971). AOAC International.
- AOAC. (2019d). *Glucose, fructose, sucrose, and maltose in presweetened cereals – Liquid chromatographic method* (AOAC 982.14-1983). AOAC International.
- AOAC. (2019e). *Crude protein in meat and meat products including pet foods – Combustion method* (AOAC 992.15-1992). AOAC International.
- AOAC. (2019f). *Protein (crude) in animal feed, Forage (plant tissue), grain, and oilseeds. Block digestion method using copper catalyst and steam distillation into boric acid* (AOAC 2001.11-2005). AOAC International.
- AOAC. (2019g). *Protein (crude) in animal feed. Combustion method* (AOAC 990.03-2002). AOAC International.
- AOCS. (2020a). *Crude fiber in feed by filter bag technique* (AOCS Standard Procedure Ba 6a-05). American Oil Chemists' Society.
- AOCS. (2020b). *Crude fiber in oilseed by-products* (AOCS Official Method Ba 6-84). American Oil Chemists' Society.

- AOCS. (2020c). *Generic combustion method for crude protein* (AOCS Official Method Ba 4e-93). American Oil Chemists' Society.
- Babu, R., Nair, S. K., Kumar, A., Venkatesh, S., Sekhar, J. C., Singh, N. N., Srinivasan, G., & Gupta, H. S. (2005). Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). *Theoretical and Applied Genetics*, *111*, 888–897. <https://doi.org/10.1007/s00122-005-0011-6>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823*.
- Bernardo, R. (2010). *Breeding for quantitative traits in plants*. Stemma Press.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Chateigner-Boutin, A. L., Ordaz-Ortiz, J. J., Alvarado, C., Bouchet, B., Durand, S., Verherbruggen, Y., Barrière, Y., & Saulnier, L. (2016). Developing pericarp of maize: A model to study arabinoxylan synthesis and feruloylation. *Frontiers in Plant Science*, *7*, 1476. <https://doi.org/10.3389/fpls.2016.01476>
- Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E. S., & Flint-Garcia, S. A. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology*, *158*, 824–834. <https://doi.org/10.1104/pp.111.185033>
- Cummings, M. P. (2014). FigTree. In J. M. Hancock, & M. J. Zvelebil (Eds.), *Concise Encyclopaedia of Bioinformatics and Computational Biology* (1st ed., pp. 271–272). Wiley–Blackwell.
- Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., Zhang, X., Jin, M., Zhao, M., & Yan, J. (2017). The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnology Journal*, *15*, 1250–1263. <https://doi.org/10.1111/pbi.12712>
- Dudley, J. W., & Lambert, R. J. (2004). 100 generations of selection for oil and protein in corn. *Plant Breeding Reviews*, *24*, 79–110.
- Dumas, J. B. A. (1831). Procédes de l'analyse organique. *Annales de Chimie et de Physique*, *47*, 198–205.
- Dunlap, F. G., White, P. J., Pollak, L. M., & Brumm, T. J. (1995). Fatty acid composition of oil from adapted, elite corn breeding materials. *Journal of the American Oil Chemists' Society*, *72*, 981–987. <https://doi.org/10.1007/BF02660709>
- Duvick, D. N., & Cassman, K. G. (1999). Post-Green Revolution trends in yield potential of temperate maize in the north-central United States. *Crop Science*, *39*, 1622–1630. <https://doi.org/10.2135/cropsci1999.3961622x>
- Elston, R. C. (1963). A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics*, *19*, 85–97. <https://doi.org/10.2307/2527573>
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. Longman Group.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*, D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection: High-resolution maize association population. *The Plant Journal*, *44*, 1054–1064. <https://doi.org/10.1111/j.1365-313X.2005.02591.x>
- Gage, J. L., White, M. R., Edwards, J. W., Kaeppler, S., & De Leon, N. (2018). Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. *Genetics*, *210*, 1125–1138. <https://doi.org/10.1534/genetics.118.301487>
- García-Lara, S., Chuck-Hernandez, C., & Serna-Saldivar, S. O. (2019). Development and structure of the corn kernel. In S. O. Serna-Saldivar (Ed.) *Corn* (3rd ed., pp. 147–163). Elsevier. <https://doi.org/10.1016/B978-0-12-811971-6.00006-1>
- Garrison, E., & Marth, G. (2012). *FreeBayes*. *ArXiv preprint 1207.3907* bib>
- Hansey, C. N., Johnson, J. M., Sekhon, R. S., Kaeppler, S. M., & De Leon, N. (2011). Genetic diversity of a maize association population with restricted phenology. *Crop Science*, *51*, 704–715. <https://doi.org/10.2135/cropsci2010.03.0178>
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, *28*, 476–490. <https://doi.org/10.1093/genetics/28.6.476>
- Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M. A., Barry, K., De Leon, N., Kaeppler, S. M., & Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, *26*, 121–135. <https://doi.org/10.1105/tpc.113.119982>
- Holding, D. R., Otegui, M. S., Li, B., Meeley, R. B., Dam, T., Hunter, B. G., Jung, R., & Larkins, B. A. (2007). The maize *floury1* gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *The Plant Cell*, *19*, 2569–2582. <https://doi.org/10.1105/tpc.107.053538>
- Holmes, M., Renk, J. S., Coaldrake, P., Kalambur, S., Schmitz, C., Anderson, N., Gusmini, G., Annor, G., & Hirsch, C. N. (2019). Food-grade maize composition, evaluation, and genetics for masa-based products. *Crop Science*, *59*, 1392–1405. <https://doi.org/10.2135/cropsci2018.10.0605>
- Honigs, D. E., Hieftje, G. M., Mark, H. L., & Hirschfeld, T. B. (1985). Unique-sample selection via near-infrared spectral subtraction. *Analytical Chemistry*, *57*, 2299–2303. <https://doi.org/10.1021/ac00289a029>
- Jackson, D. S., Choto-Owen, C., Waniska, R. D., & Rooney, L. W. (1988). Characterization of starch cooked in alkali by aqueous high-performance size-exclusion chromatography. *Cereal Chemistry*, *65*, 493–496.
- James, M. G., Robertson, D. S., & Myers, A. M. (1995). Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. *The Plant Cell*, *7*, 417–429. <https://doi.org/10.1105/tpc.7.4.417>
- Jellum, M. D., & Marion, J. E. (1966). Factors affecting oil content and oil composition of corn (*Zea mays* L.) grain. *Crop Science*, *6*, 41–42. <https://doi.org/10.2135/cropsci1966.0011183X000600010012x>
- Jenner, C. F., Ugalde, T. D., & Aspinall, D. (1991). The physiology of starch and protein deposition in the endosperm of wheat. *Functional Plant Biology*, *18*, 211–226. <https://doi.org/10.1071/PP9910211>
- Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C. S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., ... Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*, 524–527. <https://doi.org/10.1038/nature22971>
- Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for

- association mapping in soybean and maize. *Frontiers in Plant Science*, 10, 1794. <https://doi.org/10.3389/fpls.2019.01794>
- Kjeldahl, J. (1883). Neue methode zur bestimmung des stickstoffs in organischen körpern. *Zeitschrift Für Analytische Chemie*, 22, 366–382. <https://doi.org/10.1007/BF01338151>
- Kozak, M., Krzanowski, W., & Tartanus, M. (2012). Use of the correlation coefficient in agricultural sciences: Problems, pitfalls and how to deal with them. *Anais da Academia Brasileira de Ciências*, 84, 1147–1156. <https://doi.org/10.1590/S0001-37652012000400029>]
- Laurie, C. C., Chasalow, S. D., Ledeaux, J. R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T. R., & Dudley, J. W. (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics*, 168, 2141–2155. <https://doi.org/10.1534/genetics.104.029686>
- Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M. L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., & Yan, J. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics*, 45, 43–50. <https://doi.org/10.1038/ng.2484>
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28, 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, 12, e1005767. <https://doi.org/10.1371/journal.pgen.1005767>
- Lucas, C. J., Zhao, H., Schneerman, M., & Moose, S. P. (2013). Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize. In P. W. Bercraft (Ed.), *Seed Genomics* (pp. 217–236). <https://doi.org/10.1002/9781118525524.ch12>
- Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI Handbook*. 2nd edition. National Center for Biotechnology Information.
- Mazaheri, M., Heckwolf, M., Vaillancourt, B., Gage, J. L., Burdo, B., Heckwolf, S., Barry, K., Lipzen, A., Ribeiro, C. B., Kono, T. J. Y., Kaeppler, H. F., Spalding, E. P., Hirsch, C. N., Robin Buell, C., De Leon, N., & Kaeppler, S. M. (2019). Genome-wide association analysis of stalk biomass and anatomical traits in maize. *BMC Plant Biology*, 19, 45. <https://doi.org/10.1186/s12870-019-1653-x>
- McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S. E., Peterson, B., ... Buckler, E. S. (2009). Genetic properties of the maize nested association mapping population. *Science*, 325, 737–740. <https://doi.org/10.1126/science.1174320>
- Mertens, D. R. (1992). Critical conditions in determining detergent fibers. *Proceedings of the National Forage Testing Association Forage Analysis Workshop, Sept. 16–17, Denver, CO* (pp. C1–C8). National Forage Testing Association.
- Mertz, E. T., Bates, L. S., & Nelson, O. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, 145, 279–280. <https://doi.org/10.1126/science.145.3629.279>
- Nardmann, J., Zimmermann, R., Durantini, D., Kranz, E., & Werr, W. (2007). WOX gene phylogeny in Poaceae: A comparative approach addressing leaf and embryo development. *Molecular Biology and Evolution*, 24, 2474–2484. <https://doi.org/10.1093/molbev/msm182>
- Nelson, O. E., Mertz, E. T., & Bates, L. S. (1965). Second mutant gene affecting the amino acid pattern of maize endosperm proteins. *Science*, 150, 1469–1470. <https://doi.org/10.1126/science.150.3702.1469>
- O'Connor, C. H., Qiu, Y., Coletta, R. D., Renk, J. S., Monnahan, P. J., Noshay, J. M., Liang, Z., Gilbert, A., Anderson, S. N., McGaugh, S. E., Springer, N. M., & Hirsch, C. N. (2020). Population level variation of transposable elements in a maize diversity panel. *bioRxiv* 2020.09.25.314401. <https://doi.org/10.1101/2020.09.25.314401>
- Pflugfelder, R. L., Rooney, L. W., & Waniska, R. D. (1988). Fractionation and composition of commercial corn masa. *Cereal Chemistry*, 65, 262–266.
- Prasanna, B. M., Vasal, S. K., Kassahun, B., & Singh, N. N. (2001). Quality protein maize. *Current Science*, 1308–1319.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Foundation for Statistical Computing.
- Reynolds, T. L., Nemeth, M. A., Glenn, K. C., Ridley, W. P., & Astwood, J. D. (2005). Natural variability of metabolites in maize grain: Differences due to genetic background. *Journal of Agricultural and Food Chemistry*, 53, 10061–10067. <https://doi.org/10.1021/jf051635q>
- Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., Elshire, R. J., Acharya, C. B., Mitchell, S. E., Flint-Garcia, S. A., McMullen, M. D., Holland, J. B., Buckler, E. S., & Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, 14, R55. <https://doi.org/10.1186/gb-2013-14-6-r55>
- Scott, P., Pratt, R. C., Hoffman, N., & Montgomery, R. (2019). Specialty corns. In S. O. Serna-Saldivar (Ed.), *Corn* (pp. 289–303). Elsevier. <https://doi.org/10.1016/B978-0-12-811971-6.00010-3>
- Serna-Saldivar, S. O. (1996). *Química, almacenamiento e industrialización de los cereales*. AGT Editor.
- Serna-Saldivar, S. O., & Rooney, L. W. (2015). Industrial production of maize tortillas and snacks. In L. W. Rooney, & S. O. Serna-Saldivar (Eds.), *Tortillas* (pp. 247–281). AACCC International Press.
- Shen, B., & Roesler, K. (2017). Maize kernel oil content. In *The Book Maize Kernel Development* (pp. 160–174). University of Nebraska–Lincoln.
- Smith, H. F. (1936). A discriminant function for plant selection. *Annals of Eugenics*, 7, 240–250. <https://doi.org/10.1111/j.1469-1809.1936.tb02143.x>
- Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, 128, 851–864. <https://doi.org/10.1007/s00122-015-2475-3>
- Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6, 35–39. <https://doi.org/10.1177/875647939000600106>
- Tracy, W. F. (2010). History, genetics, and breeding of supersweet (*shrunk2*) sweet corn. In J. Janick (Ed.), *Plant breeding reviews* (pp. 189–236). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470650073.ch7>
- Tsai, C. Y. (1974). The function of the waxy locus in starch synthesis in maize endosperm. *Biochemical Genetics*, 11, 83–96. <https://doi.org/10.1007/BF00485766>

- USDA–National Agricultural Statistical Service. (2019). *Crop production report*. USDA–National Agricultural Statistical Service. https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan19.pdf
- Van Heerwaarden, J., Hufford, M. B., & Ross-Ibarra, J. (2012). Historical genomics of North American maize. *Proceedings of the National Academy of Sciences*, 109, 12420–12425. <https://doi.org/10.1073/pnas.1209275109>
- Wang, H., Huang, Y., Xiao, Q., Huang, X., Li, C., Gao, X., Wang, Q., Xiang, X., Zhu, Y., Wang, J., Wang, W., Larkins, B. A., & Wu, Y. (2020). Carotenoids modulate kernel texture in maize by influencing amyloplast envelope integrity. *Nature Communications*, 11, 5346. <https://doi.org/10.1038/s41467-020-19196-9>
- White, M. R., Mikel, M. A., Leon, N., & Kaeppler, S. M. (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. *Crop Science*, 60, 100–114. <https://doi.org/10.1002/csc2.20050>
- Williams, J. S. (1962). The evaluation of a selection index. *Biometrics*, 18, 375–393. <https://doi.org/10.2307/2527479>
- Wu, J., Lawit, S. J., Weers, B., Sun, J., Mongar, N., Van Hemert, J., Melo, R., Meng, X., Rupe, M., Clapp, J., Haug Collet, K., Trecker, L., Roesler, K., Peddicord, L., Thomas, J., Hunt, J., Zhou, W., Hou, Z., Wimmer, M., ... Habben, J. E. (2019). Overexpression of *zmm28* increases maize grain yield in the field. *Proceedings of the National Academy of Sciences*, 116, 23850–23858. <https://doi.org/10.1073/pnas.1902593116>
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565–569. <https://doi.org/10.1038/ng.608>
- Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B. M., Li, L., Li, J., & Yan, J. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Molecular Breeding*, 28, 511–526. <https://doi.org/10.1007/s11032-010-9500-7>
- Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178, 539–551. <https://doi.org/10.1534/genetics.107.074245>
- Zhang, X., Colleoni, C., Ratushna, V., Sirghie-Colleoni, M., James, M., & Myers, A. (2004). Molecular characterization demonstrates that the *Zea mays* gene *sugary2* codes for the starch synthase isoform SSIIa. *Plant Molecular Biology*, 54, 865–879. <https://doi.org/10.1007/s11103-004-0312-1>
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, 1. <https://doi.org/10.3835/plantgenome2008.02.0089>

SUPPORTING INFORMATION

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

How to cite this article: Renk, J. S., Gilbert, A. M., Hattery, T. J., O'Connor, C. H., Monnahan, P. J., Anderson, N., Waters, A. J., Eickholt, D. P., Flint-Garcia, S. A., Yandea-Nelson, M. D., & Hirsch, C. N. (2021). Genetic control of kernel compositional variation in a maize diversity panel. *Plant Genome*, 14;e20115. <https://doi.org/10.1002/tpg2.20115>