LINKING ANTHOCYANIN DIVERSITY, HUE, AND GENETICS IN PURPLE CORN

Laura A. Chatham¹ and John A. Juvik¹

¹Department of Crop Sciences, University of Illinois at Urbana Champaign

Abstract

While maize with anthocyanin-rich pericarp (purple corn) is rising in popularity as a source of natural colorant for foods and beverages, information on color range and stability—factors associated with anthocyanin decorations and compositional profiles—are currently limited. Using the natural anthocyanin diversity present in a purple corn landrace, Apache Red, we generated a population with variable flavonoid profiles—flavonol-anthocyanin condensed forms (0-83%), acylated anthocyanins (2-72%), pelargonidin-derived anthocyanins (5-99%), and C-glycosyl flavone co-pigments up to 1904 µg/g—all of which contributed in part to the absorbance profile, used here as a proxy for hue. This variability offers targets of selection for breeders looking to expand both stability and the available range of colors that can be sourced from purple corn. With genotyping-by-sequencing of this population we mapped these anthocyanin profile traits. Major QTL for anthocyanin type were found near loci previously identified only in aleurone-pigmented maize varieties [*Purple aleurone1 (Pr1)* and *Anthocyanin acyltransferase1 (Aat1)*]. A QTL near *P1 (Pericarp color1)* was found for both flavone content and flavanol-anthocyanin condensed forms. A significant QTL associated with peonidin-derived anthocyanins near a candidate S-adenosylmethionine-dependent methyltransferase was also identified, warranting further investigation. This population represents the most anthocyanin diverse pericarp-pigmented maize variety characterized to date. Moreover, the candidates identified here will serve as branching points for future research studying the genetic and molecular processes determining anthocyanin profile in pericarp.

Keywords: Anthocyanins, flavonoids, maize pericarp, purple corn, genotype-by-sequencing, GWAS

Introduction

Purple corn could provide an economical and natural source of colorants, but as demand rises and the desired application ranges broaden, new challenges are likely to emerge. Matching the range of possible colors, stability, and price that had been attained previously with artificial colorants will likely require additional development through breeding. A major breeding objective for purple corn is to expand its food and beverage applicability by increasing the range of available colors and shelf life. However, these goals must be accomplished while maintaining grain yield and other agronomic traits.

Understanding the genetics underlying anthocyanin biosynthesis in purple corn could expedite the breeding of new cultivars. Kernel anthocyanin content has long been studied in maize due to its easily observable phenotype, but most research to date has focused on kernels with pigmented aleurone, the outermost, single cell layer of the triploid endosperm. However, purple corn contains anthocyanin in the pericarp, the maternal, outermost layer of the kernel. Compared to aleurone, pericarp pigmented lines comprise a relatively small portion of the available germplasm and thus have limited variability. Despite this disadvantage, lines with anthocyanin-pigmented pericarp often have concentrations an order of magnitude greater than aleurone lines¹. Anthocyanin extraction may also be more efficient from pericarp lines². Moreover, established milling processes separating pericarp could provide valuable pigment-rich pericarp fractions and non-pigmented grain and germ fractions to be used for the production of food, feed, or fuel, creating a value-added product³.

Anthocyanins in purple corn consist primarily of the glucosides, and malonyl and dimalonyl glucosides of the anthocyanidins cyanidin, pelargonidin, and peonidin (Supplementary Figure 1). Color is largely influenced by anthocyanidin type, shifting from orange/red to pink/red as the number of hydroxyl and methoxy groups on the B ring increases⁴. However, anthocyanidin type alone cannot account for the diversity in colors observed. Color stability and extract shelf life have been associated with the proportion of acylated anthocyanins⁵. Another category of anthocyanin types is the flavonol-anthocyanin condensed forms, heterodimers consisting of a flavan-3-ol covalently linked to an anthocyanin, but these have only been found in significant quantities in pericarp pigmented varieties¹. Condensed forms have also been associated with both altered hues and stability^{6,7}.

The anthocyanin biosynthesis pathway in maize is well studied and most structural genes involved, with the exception of those controlling the formation of peonidin-based anthocyanins and flavonol-anthocyanin condensed forms, have been characterized^{8,9}. However, these structural genes were primarily identified in aleurone pigmented varieties. While the pathway is well-conserved among plants and most structural genes are assumed to operate similarly in aleurone and pericarp, empirical evidence of this is lacking. The genes and mechanisms controlling anthocyanidin type (e.g. pelargonidin, cyanidin, peonidin) and anthocyanin decorations such as acylation and condensation with flavanols are of interest in purple corn because of their potential to influence the color and stability of extracts. *Pr1* (*Red aleurone1*) is known to control the ratio of cyanidin and pelargonidin in aleurone⁹, and more recently, an anthocyanin acyltransferase (*Aat1*) was identified in an aleurone pigmented population with reduced proportions of acylated anthocyanins in some individuals¹⁰. This work seeks to confirm whether these same genes function in pericarp and to identify loci associated with the biosynthesis of peonidin and condensed forms.

Another factor controlling purple corn extract color and stability is copigmentation. The addition of maize C-glycosyl flavone-rich extract to maize anthocyanin-rich extract can significantly alter hue and stability¹¹. Flavone and anthocyanin co-occurrence in the same lines and its effect on extract hue has not been studied or reported previously. *Pericarp color1* (*PI*) is known to regulate the biosynthesis of flavones¹² but factors controlling the relative ratios of anthocyanin to flavone within a variety are not known.

Color plays a significant role in a consumer's perception of food and beverage identity, flavor, and quality¹³, making hue an important consideration in the formulation of a natural colorant. Even subtle differences in product color could have a large impact on consumer acceptance, and thus understanding how various anthocyanin and flavonoid chemistries lead to differences in hue will be important for optimizing color in food and beverage applications. Moreover, linking chemistries and hue to their underlying genetics may aid in breeding purple corn lines for increased color range. Here we seek to explore the most prominent factors contributing to purple corn anthocyanin extract color and to identify the candidate loci responsible.

Materials and Methods

Plant materials

Seeds of a purple corn landrace, Apache Red (AR), were purchased from Siskiyou Seeds (Williams, OR)(S₀ population) and grown out to produce S₁ seeds. S₁ seeds were analyzed for anthocyanin content using HPLC and seven lines containing anthocyanins were advanced while non-anthocyanin containing progeny were discarded. Selected anthocyanin-rich S₁ ears along with a sample of S₀ seeds from the original AR source were grown and self-pollinated, producing 181 S₂s and 8 new S₁s, which were also analyzed using HPLC. The available S₂ lines were genotyped as described below. A selection (9 lines) of highly pigmented S₂ seeds were grown in a winter nursery to produce 86 S₃s. The following season, a selection of anthocyanin containing S₁, S₂, and S₃ lines spanning the available anthocyanin diversity in our AR lines were planted and S₂, S₃, and S₄ seed was recovered. Of these, anthocyanin-rich lines were selected from each S₁, S₂, and S₃ parent for genotyping and phenotyping (1148 lines, 103 S₂s, 730 S₃s, and 315 S₄s). To this, 69 samples containing no or minimal anthocyanin content and representing the original 7 S₁ families were added and genotyped and phenotyped to supplement the data set. A schematic showing how the population was created can be found in Supplementary Figure 2.

Extractions

The anthocyanin extraction method was based on previously published protocols and findings regarding the efficiency of anthocyanin extraction protocols^{1,14}. Briefly, a subset of kernels from each line was ground to a fine powder using a coffee grinder and 1 g of powder was weighed out for extraction. Samples were extracted in 5 ml of aqueous 2% formic acid for 1 h at 50 °C with constant agitation. Extracts were centrifuged and filtered through a 25 mm, 0.45 µm Millex LCR PTFE syringe filter (Millipore, Billerica, MA) prior to HPLC analysis and optical measurement with UV-visible spectroscopy.

HPLC analysis

Anthocyanin and flavone content was measured using an Agilent 1100 series high-performance liquid chromatography (HPLC) system and diode array detector (DAD) (Santa Clara, CA) in a similar manner to previous reports¹⁴. The stationary phase consisted of a Poroshell 120 SB-C18 100 mm x 4.6 mm, 2.7 µm column (Agilent), and the mobile phase was a two-solvent mixture, with 2% formic acid in water as solvent A, and acetonitrile as solvent B.

The protocol began at 93% solvent A and 7% solvent B, and solvent B increased linearly to 18% over 30 min. Column temperature was kept at 30 °C, an injection volume of 20 µl was used, and absorbance was measured at 520 for anthocyanin content and 340 nm for flavone content. ChemStation (rev A10.02) software (Agilent) was used for integration and anthocyanins were quantified using standard curves of cyanidin, pelargonidin, and peonidin 3-glucosides, while flavones were quantified using C-hexosyl apigenin compounds (vitexin and isovitexin) (Extrasynthese, Genay, France). Identities were confirmed by mass spectrometry as described previously¹⁴. To ensure consistency between batches of samples and retention times, an AR mixture was created by combining ground powder from multiple AR lines, a sample of which was extracted and run on HPLC with each batch of samples.

UV-Visible spectroscopy

Absorbance was recorded between 380 and 780nm on a Synergy 2 multi-well plate reader (Biotek, Winooski, VT) using a 96-well plate. Each extract was run in triplicate and the maximum absorbance (Abs_{max}) and the wavelength at Abs_{max} (λ_{max}) were recorded.

Statistical analysis

Anthocyanin content, flavone content, and all other factors used in analyzing the relationship between various anthocyanin species and hue were normalized before plotting, and all statistics were calculated using R.

GBS library construction and sequencing

Fourteen S₂, S₃, and S₄ seeds of each line were sown in 11 x 21 cell propagation trays (Proptek, Watsonville, CA) filled with a soilless seedling mix (Sunshine Redi-Earth Plug & Seedling; SunGro, Agawam, MA) and grown in a greenhouse at the University of Illinois Plant Care Facility. Once seedlings had emerged, tissue of a similar size and proportion was harvested from at least seven seedlings, depending on germination rate. Samples were frozen to -20 °C, lyophilized, and ground in 15ml conical tubes with stainless steel ball bearings using an automated tissue homogenizer (SPEX SamplePrep, Metuchen, NJ). A 96-well optimized CTAB DNA extraction protocol was used for DNA isolation, and concentrations were determined using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA). GBS

libraries were constructed as described previously using a double restriction enzyme digest with *HinPII* and *PstI*^{15,16}. All lines were multiplexed into three lanes and sequenced using the Illumina HiSeq 4000 system with single-end 100 nucleotide reads at the Roy J. Carver Biotechnology Center at the University of Illinois in Urbana, IL.

SNP discovery and association mapping

SNP discovery was carried out using the TASSEL 5.0 GBS v2 pipeline¹⁷. Reads were aligned to the *Zea Mays* B73 reference genome version 4¹⁸ using Bowtie 2¹⁹, and minor allele frequency was set to 0.05. This generated 149,342 SNP sites with an average depth of 12 and average proportion covered of 0.65. With imputation, performed using Beagle 5.0²⁰, and filtering, a total of 41,986 nonredundant SNPs per sample were generated. Association mapping was performed using GAPIT (Genome Association and Prediction Integrated Tool) in R²¹ for several phenotypic traits. Seven principle components were used to adequately account for population structure in addition to the kinship matrix provided by GAPIT.

Methyltransferase candidates

The amino acid sequence of *VvAOMT1*, an S-adenosyl-L-methionine-dependent methyltransferase, with 3-*O*-methyltransferase activity²² was blasted against the maize genome to identify homologous candidates in maize. These candidates along with known caffeoyl-co A *O*-methyltransferases (CCoAOMTs), anthocyanin *O*-methyltransferase (AOMTs) and caffeic acid *O*-methyltransferases (CaOMTs) gathered from NCBI. Candidate sequence selection was based in part on phylograms from Provenzano et al.²³ and Du et al.²⁴. Sequences were aligned in R using the DECIPHER²⁵ package, and a phylogram was created using the phangorn package²⁶. Alignment of leucoanthocyanidin reductases and anthocyanidin reductase was performed in a similar manner using the same R packages.

Results and Discussion

Anthocyanin diversity

Previously we showed that Apache Red (AR) contains a variety of anthocyanin profiles with some containing mostly pelargonidin-derived anthocyanins, a profile not previously seen in anthocyanin-rich pericarp lines^{1,14}. We also identified AR lines with high concentrations of C-

glycosyl flavones, which copigment with anthocyanins to alter color intensity, hue, and stability²⁷. Given the diversity and unique qualities present in this population, we self-pollinated lines of the original landrace to extract as much phenotypic diversity from the population as possible. The wavelength at maximum absorption (λ_{Max}) of anthocyanin lines ranged from 497nm to 521nm. As a proxy for hue, this range suggests good variability in extract colors among the samples. Anthocyanin and C-glycosyl flavones were also highly concentrated compared to previous reports (1598 $\mu\text{g/g}$ whole corn and 1904 $\mu\text{g/g}$, respectively, Table 1)^{1,14}. Many lines were abundant in both, with flavone:anthocyanin ratios as high as 11:1.

Compared to reports of other purple corn varieties, we observed higher percentages of acylated anthocyanins (72%), pelargonidin- (99%) and peonidin-derived anthocyanins (50%), and condensed forms (83%) (Table 1). Total pelargonidin-derived anthocyanins reached nearly 1400 $\mu\text{g/g}$, cyanidin-derived 872 $\mu\text{g/g}$, and peonidin 274 $\mu\text{g/g}$. The lower peonidin content is consistent with our previous findings¹, but peonidin content is of interest due to its potential to shift hue to a slightly more bluish-red color compared to cyanidin²⁸. While cyanidin content is lower than previously reported in other pericarp pigmented purple corn lines, the pelargonidin content is much higher. Currently, orange-red natural colorant sources are limited. Carotenoids can be used in applications requiring orange but are not water soluble and can be expensive and difficult to incorporate into some applications. Carmine, a red pigment derived from cochineal insects, also provides a true red/orange, but can be controversial due to its non-vegetarian source. A pelargonidin-dominant, high anthocyanin content purple corn line could provide a much-needed source of water-soluble, acid-stable plant-based red/orange colorant for the food and beverage industry¹³.

Figure 1A illustrates the overall chemical diversity in the AR population using principal component analysis (PCA) based on proportions of total anthocyanin content attributed to each anthocyanin species. An HPLC chromatogram with the peaks used for this analysis can be found in Figure 1B and PCA loadings and peak identities in Table 2. Each anthocyanin was confirmed using retention time and previous mass spectrometry results¹⁴.

The first principal component (PC1) explained 34.5% of the variability in the data set and represents a contrast of pelargonidin versus cyanidin-derived anthocyanins. This is enforced by the color gradient across Figure 1A, in which point color represents wavelength. The λ_{Max} for pelargonidin 3-glucoside (PgG) is ~496, the λ_{Max} for cyanidin 3-glucoside (CyG) is ~510²⁸, and

a gradient from low to high can be seen across the x-axis in the biplot. The second PC explained 18.5% of the phenotypic variability in anthocyanin species and was related to the presence of condensed forms and flavones (high PC2 value) or acylated forms (low PC2 value). Wavelength also appeared to increase with PC2, suggesting condensed forms and flavones may play a role in hue.

Relationship between anthocyanin profile and λ_{Max}

To further analyze the relationship between anthocyanin profiles and hue, λ_{Max} was modeled using PC1 and PC2. PC1, PC2, and their interaction were all significant, which is unsurprising given the color gradients that can be observed in the PCA in Figure 1. We also looked at the major factors associated with PC1 and PC2 based on loadings for each component (Supplementary Figure 3, Table 2). These factors, including proportions of condensed, acylated, and cyanidin/peonidin-derived forms (a measure of pelargonidin versus cyanidin), and flavone content, were plotted against λ_{Max} to identify correlations. These factors were also used to create a complex global model from which the most parsimonious model was selected based on the lowest AIC (Akaike Information Criterion) score, an estimate of model quality. Collinearity between variables and multi-way interactions between factors make interpretation of the model and gauging whether these interactions have a biochemical basis difficult. Nonetheless, cyanidin content especially ($p = 1.09e-235$), and the other main effects [flavone, $p = 7.66e-21$; condensed, $p = 9.65e-30$; acylated ($p = 3.67e-8$)] were significant, suggesting each can be considered to influence λ_{Max} and overall color in AR lines.

A strong relationship between cyanidin-derived anthocyanins and λ_{Max} was expected, however the correlations with condensed forms and with acylated forms were more surprising. Flavonol-anthocyanin condensed forms have been reported to contribute a bluish hue to wine,²⁹ and in purple corn extracts appeared to result in a darker red color but otherwise did not affect color⁷. Acylated and condensed forms are also breeding targets because of their association with anthocyanin stability. Acylated anthocyanins are known to positively impact stability⁵, but reports on the stability of condensed forms are mixed. Some found that lines containing condensed forms produced more stable extracts⁶, while others showed no contribution to stability⁷. Flavone content was also significant in modeling λ_{Max} . This is consistent with previous findings¹¹ but is our first evidence suggesting that when co-occurring in the same line, C-

glycosyl flavone and anthocyanin copigmentation is significant in the determination of extract color. This will likely have major implications on breeding objectives; i.e. when an orange extract is desired, it will be essential to minimize flavone content to avoid an increase in λ_{Max} and a shift from orange to red. Yet when a red-pink or more berry-colored extract is desired, flavone content should be maximized.

While all major anthocyanin factors were significant, interpretation of these results must consider correlations between factors. The proportion of condensed and acylated forms were negatively correlated ($r = -0.79$, $p < 2.2\text{e-}16$). Some of the condensed forms identified did include acylated groups, and while this could contribute to the correlations, acylated condensed forms were typically present in only very minor amounts. This correlation suggests that acylation and polymerization to make condensed forms are competing processes. Furthermore, if either influenced λ_{Max} , the correlation could result in both being significant factors. The proportion of condensed and cyanidin forms were also negatively correlated ($r = -0.26$, $p = 5.2\text{e-}15$). Likewise, this correlation could influence the significance of condensed forms in modeling λ_{Max} .

Flavones were positively correlated with condensed forms ($r = 0.39$, $p < 2.2\text{e-}16$) and negatively correlated with percent cyanidin-derived ($r = -0.50$, $p < 2.2\text{e-}16$) and acylated anthocyanins ($r = -0.29$, $p < 2.2\text{e-}16$). While it is possible that these correlations are artifacts of population structure or linkage in the data set, there are also biochemical-based hypotheses that could explain them (Figure 2). For example, it is not unreasonable to hypothesize that the flavonol component of condensed forms, and flavones could share similar regulatory mechanisms, thus explaining their co-occurrence in many of the lines. The relationship between flavones and pelargonidin derived anthocyanins could be explained by substrate specificity. If *Fns1* had a preference for naringenin, a correlation between flavone content and pelargonidin content would be expected, with low flavone concentrations in cyanidin dominant lines. The lack of abundant luteolin-based (eriodictyol-derived) flavones and the exclusivity of apigenin-based (naringenin-derived) flavones supports this hypothesis. Some luteolin compounds were detected with mass spectrometry but were not present in high enough concentrations to adequately measure. More research and intercrossing between AR lines could help determine whether these observed correlations are due to population structure or have biological explanations.

Analyzing the flavonoid chemistry in Apache Red reveals that extract color is determined by more than just cyanidin to pelargonidin ratio. To successfully breed for specific hues, all of

these factors should be considered in a breeding program. However, phenotyping to this extent is time consuming and expensive. Identifying the loci responsible for various anthocyanin decorations and flavones and the ratio between them would be beneficial for eventually creating markers for use in marker assisted selection (MAS).

Genotyping and population structure

Given that all the lines were derived from the same landrace, controlling for population structure was a major concern in order to reduce false positives, yet maintain the power to detect significant associations between genotypes and traits. A PCA of the genotype data for all lines showed that PC1 explained about 10% of the variance in the data set. A drop in the amount of variance explained by each PC (to 2%) was found between PC6 and PC7, and thus 7 PCs were used to help control for population structure in addition to the kinship matrix calculated by GAPIT²¹. To maximize signal detection, we also used a GWAS model that calculates relatedness using a marker subset comprised of all but those on the chromosome where the tested marker resides³⁰. Linkage began to decay significantly by 1.5 Mb when calculated by GAPIT, and when a 10 site or 50 site window was used to calculate LD in Tassel, significant decay was seen by 3 Mb and 20 Mb, respectively (Supplementary Figure 4). This is indicative of the number of historical recombination events captured in this study and thus its resolution for marker-trait associations.

Mapping pelargonidin, cyanidin, and peonidin content

All anthocyanins identified in the AR population were derived from pelargonidin, cyanidin, or peonidin. As discussed above, percentages of these anthocyanidins have a significant effect on extract color, warranting the identification of responsible genetic elements and eventually markers for use in marker assisted selection (MAS). In mapping the proportion of cyanidin-derived anthocyanins, we expected to find *Pr1*, which produces a flavonoid 3' hydroxylase (F3'H)⁹ in cyanidin containing aleurone lines. However, *Pr1* has not been associated with cyanidin/pelargonidin ratio in pericarp. F3'H catalyzes the conversion of naringenin to eriodictyol and dihydrokaempferol to dihydroquercetin, the steps necessary for producing cyanidin. In maize aleurone, recessive *pr1/pr1* results in the accumulation of almost exclusively pelargonidin-derived anthocyanins and therefore red or pink kernels instead of the

purple or blue kernels associated with the presence of cyanidin in most anthocyanin-containing aleurone varieties^{8,9,31}. The abundance of pelargonidin-exclusive AR lines and the distribution seen in Supplementary Figure 3D suggests a single gene encoding a F3'H is responsible in pericarp as well.

Indeed, mapping cyanidin-derived anthocyanins resulted in a highly significant peak on Chromosome 5 (Chr 5) near *Pr1* (Figure 3A, Table 4). In aleurone, three nonfunctional *pr1* alleles have been identified, including a 24 nucleotide TA repeat insertion in the 5' upstream region, a 17 nucleotide deletion near the TATA box, and a Ds insertion in the first exon⁹. Cloning and sequencing of *Pr1* and *pr1* alleles in AR suggests that the same 17 nucleotide deletion observed in aleurone is responsible for nonfunctional *pr1* activity in AR (data not shown). In addition to the highly significant peak near *Pr1*, we also identified several other signals. Most had relatively low minor allele frequencies except for one at the beginning of Chr 8 (<1 Mb). No promising candidates were found for any of these additional loci.

Mapping the proportion of peonidin-derived anthocyanins produced similar results to mapping the cyanidin proportion - a signal on Chr 5 near *Pr1* (Figure 3B). This is expected since the synthesis of the peonidin precursor, believed to be homoeriodictyol³², depends on the conversion of naringenin to eriodictyol catalyzed by the *Pr1* gene product. *Pr1* is likely epistatic with a gene encoding a product with anthocyanin *S*-adenosyl-L-methionine-dependent *O*-methyltransferase (AOMT) activity in the formation of homoeriodictyol and thus peonidin.

To isolate AOMT action, we created a new phenotype variable describing the proportion of all eriodictyol-based compounds that were converted to homoeriodictyol-based compounds. This was done by dividing the total peonidin content by the sum of cyanidin and peonidin content. When mapping this variable (Figure 3B), markers near *Pr1* were still significant, but a signal on Chr 4 around 200 Mb emerged. Several other significant SNPs were identified (Table 4), but most had low minor allele frequency, which suggests a higher probability of error. Individual chromosome Manhattan plots showing loci of interest can be found in Supplementary Figure 5.

Methyltransferase candidates

Many AOMTs have been identified in plants producing *O*-methoxy anthocyanidins, yet these have all been identified in dicots, and only Caffeoyl CoA *O*-methyltransferase

(CcoAOMT) homologs have been identified in monocot species²³. Recently a CcoAOMT was identified in wheat (*Triticum aestivum*) through transcriptome analysis of purple pericarp. *TaCcoAOMT* was upregulated in peonidin producing wheat pericarp, but the locus was not further characterized³³. No methyltransferase with AOMT activity responsible for peonidin biosynthesis in maize has been identified, but the presence of peonidin-derived anthocyanins suggests its existence. Amino acid sequences of AOMTs from *Vitis vinifera* and *Glycine max* were blasted against the maize genome¹⁸, and candidates are listed in Supplementary Table 1. To narrow down the likely candidates, amino acid sequences of AOMTs, CCoAOMTs, and caffeic acid OMTs from other species were aligned with the maize candidates and a maximum likelihood phylogeny was created (Supplementary Figure 6). A clear outgroup containing the caffeic acid OMTs is evident, helping to exclude Zm00001d052683 and Zm0001d052684. Zm00001d052843 and Zm00001d024596 were also in outgroups apart from the remaining AOMTs and CcoAOMTs, which clustered nicely into two groups. Zm00001d045206 and Zm00001d036293 clustered with the CcoAOMTs, but are located on Chr 9 and Chr 6, respectively. The remaining candidates, Zm00001d052841 and Zm00001d052842 were the next most closely related candidates to the AOMTs and CcoAOMTs, are identified as putative CcoAOMTs, and are located on Chr 4 around 202 Mb.

The methyltransferase phylogeny and the proximity to the signal mapped to Chr 4 for peonidin content suggests Zm00001d052841 and Zm00001d052842 are likely candidates. These candidates were aligned to several methyltransferases with known AOMT and CcoAOMT activity, as well as the CcoAOMT expressed in purple wheat pericarp (Supplementary Figure 7). Amino acid residues associated with enzyme function are highlighted and were conserved across all sequences in most cases. Variability was observed between sequences for Arg217 (CoA binding), Tyr219, and Tyr223. Modeling of substrate binding showed that the bulky tyrosine residues present in CcoAOMTs in this location clash with the anthocyanin 3-glucose moiety, i.e. the site cannot accommodate the larger anthocyanin 3-glucose moiety due to obstruction by a bulky tyrosine²³. In petunia AOMTs, these tyrosines have been replaced by the smaller amino acids, glycine and leucine. In *TaCcoAOMT*, these have been replaced by arginine and leucine, and by isoleucine, leucine and phenylalanine in the two maize candidates. Replacement with phenylalanine certainly would not reduce the size or bulkiness to the same degree that a glycine or leucine substitution would, but a tyrosine to phenylalanine conversion at one of these

locations (Tyr219 or Tyr223) was observed for VvAOMT1, VvAOMT3, DhAOMT, and GmAOMT, all of which have anthocyanin OMT activity. Moreover, the relatively low proportions of peonidin (max = 50%) compared to cyanidin (max=86%) observed here and in purple wheat³⁴, as well as in other purple corn studies¹ suggest that the responsible enzyme has relatively low efficiency. If CcoAOMTs were co-opted for anthocyanin use, it is reasonable that they would have suboptimal performance compared to the AOMTs specialized for peonidin production. Here we found higher proportions of peonidin-derived anthocyanins when anthocyanin content was low ($r=-0.24$, $p<2.2e-16$), lending support for this hypothesis.

While these residue changes suggest that Zm00001d052841 and Zm00001d052842 could better accommodate the anthocyanin-3-glucoside moiety than the canonical maize CcoAOMTs, it is believed that the methyltransferase in maize acts prior to glycosylation³². Evidence for this comes from a study that identified an S-adenosylmethionine-flavonoid 3'-O-methyltransferase capable of utilizing eriodictyol, luteolin, and quercetin as substrates, but not quercetin 3-glucoside in crude protein extracts from maize seedlings, leaf sheath, and aleurone³². Comparing Zm00001d052841 and Zm00001d052842 shows that the two are highly similar except for several amino acids and an extra string of 33 amino acids at the beginning of the Zm00001d052841 sequence. Further studies will be required to determine which is responsible for AOMT-like activity, what substrates can be accommodated by the product, and the residues most important for enzyme efficiency.

Mapping anthocyanin decorations

Acylated anthocyanins

Wide ranges of acylated anthocyanins, flavonol anthocyanin condensed forms, and flavones were observed in the phenotypic dataset. Acylated and condensed forms had somewhat bimodal distributions (Supplementary Figure 2), suggesting they are simply inherited traits. Mapping acylated anthocyanin proportions (Figure 4A) produced a peak on Chr 1 near 306 Mb, which aligns with *Aat1*, the anthocyanin acyltransferase recently identified in aleurone pigmented lines and located at 305.7 Mb (Supplementary Figure 5). This helps confirm that *Aat1* is associated with acylated anthocyanin content in pericarp as well as in aleurone. While this was by far the most significant SNP for acylated anthocyanins, several other significant loci were also found and are listed in Table 4.

Flavonol-anthocyanin condensed forms

For flavonol-anthocyanin condensed forms, mapping produced a wide signal surrounding *PI*, a known MYB regulator of phlobaphene and flavone biosynthesis³⁵ (Figure 4B). We also observed a signal on Chr 9 (127.6-147.8 Mb) near *FnsI* (130.6 Mb), which encodes a flavanone 2-hydroxylase required for flavone biosynthesis. Given the correlation between flavones and condensed forms, finding these loci was not surprising. Several other significant SNPs were also identified (Table 4) but determining whether these correspond to flavone or condensed form biosynthesis is muddled by the correlation between these two phenotypes.

To separate the flavone and condensed form phenotypes, we mapped condensed forms using flavone content as a covariate (Supplementary Figure 8) and looked for both new SNPs and SNPs that remained significant in both analyses. Though drastically reduced in size, the signal on Chr 1 near *PI* was still present. However, the most significant SNP was found at 40 Mb, about 8.5 Mb away from *PI*. The peak on Chr 9 near *FnsI* also remained, but the range of significant SNPs expanded. Along with the phenotypic correlations described, the stability of these signals with the flavone covariate suggests they play an indirect role in the amount of condensed forms accumulated, perhaps by regulating flux through different branches of the flavonoid pathway.

Another signal on Chr 1 was found for both analyses near *ChiI* (*Chalcone isomerase* 298.6 Mb), a flavonoid biosynthesis structural gene shared by both condensed forms and flavones. However the range of the peak changed, with the most significant SNP near 297.1 Mb without the covariate and near 293.8 Mb with the covariate. This moves the signal away from *ChiI* and closer to a potential ANR or LAR (anthocyanidin reductase / leucoanthocyanidin reductase) ortholog (Zm00001d034357) located at 290.5 Mb (Supplementary Figure 9A). Mapping flavonol-anthocyanin condensed forms was anticipated to produce a signal near a candidate LAR or ANR, the enzymes required for the biosynthesis of the flavan-3-ol component of condensed forms (Figure 2). To identify candidates prior to mapping, known LAR and ANR proteins were blasted against the maize genome¹⁸ and a maximum likelihood phylogeny was created with the hits and orthologs from other species (Supplementary Figure 10). The suggested candidate on Chr 1 is more closely related to the LAR group from this phylogeny. LAR has been associated with the extension of proanthocyanidin polymers of flavan-3-ols³⁶, leading to the

hypothesis that this same enzyme could be associated with the addition of flavan-3-ols to anthocyanins to make condensed forms.

On Chr 2, significant SNPs were found around 226.78 Mb. The closest candidate we identified is a MYB transcription factor (*myb110*, Zm00001d007191, 224.25 Mb), which was identified by blasting a transcriptional repressor of anthocyanin and proanthocyanidin genes from *Medicago truncatula* (MtMYB2) against the maize genome^{18,37}. A phylogeny containing MYB activators and repressors associated with anthocyanin biosynthesis and potential orthologous maize candidates can be found in the supplementary materials of this article's companion (Chatham and Juvik, submitted to G3). We also observed a consistent signal at the beginning of Chr 3 near 1.4 Mb, 1.3 Mb from a blast hit for *Arabidopsis TCP3* (*tcptf33*, Zm00001d039371, 2.7 Mb; Supplementary Figure 9E). TCP3 is a transcription factor that interacts with the MYB/bHLH complex to assist in activating flavonoid biosynthesis³⁸. A signal near *Pl1* (Chr 6 at 113 Mb), the MYB required for anthocyanin biosynthesis in maize plant tissue, was also found. In addition to the peak on Chr 9 near *Fns1*, we also identified a range of significant SNPs at the beginning of the chromosome (Supplementary Figure 9C). The most significant SNP in this range is less than 1 Mb from *Bz1* (*Bronze1*) a UDP-glucose flavonol glycosyltransferase responsible for glycosylating anthocyanins³⁹. Given the wide range created when the flavone covariate was added, other candidates may include a flavonoid 3-*O*-glucosyltransferase (Zm00001d045254) and *ZmMrp3*, an ABC transporter associated with transport of anthocyanins to the vacuole. However both of these are over 4 Mb away from the most significant SNP. This signal could also be tagging *Bp1* (*Brown pericarp1* - located in bin 9.02 (11.8 – 23.3 Mb in B73 RefGen_v3)), a locus that has been suggested to play a role in the polymerization of flavan-4-ols to create phlobaphenes and is epistatic with *Pl1*¹². No gene model has been associated with the *Bp1* locus, but if polymerizing activity is confirmed, it could participate in the heterodimerization between flavan-3-ols and anthocyanins to make condensed forms. Conversely, its potential regulation of flavone biosynthesis could explain why it was found for condensed forms since flavones and condensed forms are correlated in this population. Other signals observed consistently between the analysis with and without the flavone covariate were found on Chr 7 around 142 Mb and 178 Mb, each with relatively large effect sizes. A potential candidate was found at 138.2 Mb (LAR/ANR ortholog - Zm00001d020970), but this is over 3 Mb away and the signal observed was comparatively weak. A glycosyltransferase

annotated as an anthocyanidin 3-O-glucosyltransferase (Zm00001d022467 at 178.22 Mb) was found close to the signal at 178. Other significant SNPs detected can be found in Table 4.

Several new signals were identified when the flavone covariate was used. A peak on Chr 5 was found near 203.3 Mb, 1.5 Mb from *Pac1* (*Pale aleurone color1*) and 3 Mb from an ANR/LAR candidate (Zm00001d017771; Supplementary Figure 9D). *Pac1* is a WD40 repeat protein and member of the ternary MBW complex that regulates anthocyanin biosynthesis^{40,41}, but *Pac1* is only required for complete anthocyanin production in aleurone, not vegetative tissue⁴². Whether *Pac1* functions in pericarp has not yet been determined. The other candidate, Zm00001d017771, groups with orthologous LARs (Supplementary Figure 10). New signals on Chr 8 (164.4 Mb) and Chr 10 (104.6 Mb) were also found but no potential candidates were identified, and both had minor allele frequencies below 0.05.

In addition to these analyses we also mapped condensed forms as a proportion of the total anthocyanin content (Figure 4C). Several signals were redundant with mapping total condensed forms, including signals near *P1*, *Chi1*, *tcptf33*, *Fns1*, and at Chr7:142.26 Mb. *Aat1* was also previously identified for the proportion of acylated anthocyanins but given the inverse relationship between the proportion of acylated and condensed forms, its recurrence here is not surprising.

Though relatively weak, a signal on Chr 1 was found less than 1 Mb away from *myb83*, a transcription factor with similarity to *AtMYB60*, which functions as a repressor of anthocyanin biosynthesis⁴³. We also found peaks on Chr 1 and Chr 5, both near UDP-D-glucose dehydrogenases with potential roles in sugar metabolism and expression in pericarp tissue³⁵. On Chr 3, a peak near the *A1* (*Anthocyaninless1*) locus was identified. *A1* encodes a dihydroflavonol 4-reductase required for anthocyanin biosynthesis. Another candidate for this signal may be *mybr97*, a protein with similarity to MYB repressors from Petunia (*MYBx*) and *Arabidopsis* (*CPC*). CAPRICE (*CPC*) is an R3-MYB repressor that negatively regulates anthocyanin biosynthesis by competing for bHLH binding⁴⁴. Similar repressors were identified in petunia (*MYBx*) and tomato (*Atroviolacea* (*Atv*))⁴⁵. *Mybr97*, located at 221.6 Mb on Chr 3 is close to the *A3* locus (between 205.4 Mb and 215.9 Mb based on B73_refgen_V2) identified as a negative regulator of anthocyanin biosynthesis and for which no gene models have been associated.⁴⁶ On Chr 4, significant SNPs were found at 176.6 and 238.2. The best candidates identified for these QTN were a blast hit for an ATPase from petunia (*PH5*) involved in flower color⁴⁷ and a

vacuolar proton pump (vpp3) (Zm00001d052022 and Zm00001d053765, respectively). *PH5* influences the vacuolar pH and thus flower color and is required for the accumulation of proanthocyanidins. Since both proanthocyanidins and condensed forms require flavan-3-ols as building blocks, it is plausible that an ATPase in maize could be required for effective transport of flavonols and thus the synthesis of condensed forms⁴¹.

On Chr 6 we identified several significant SNPs that appear to be on either side of the centromeric region. Stronger LD in this region likely limits our ability to locate with much certainty the region containing a causative SNP. Nonetheless, we found a potential MYB candidate close to one of the QTNs. This candidate, *myb1* (Zm00001d035918) was a close match when *CPC* and *AtMYB60*, both repressors of anthocyanin biosynthesis in *Arabidopsis*, were blasted against the maize genome^{18,44}. Additionally, a signal on Chr 10 was found about 3 Mb away from *R1*, the bHLH component of the MBW ternary regulatory complex typically associated with activating anthocyanin biosynthesis in aleurone⁴¹. Originally, mapping condensed forms was expected to produce several large effect loci, especially given the current understanding of the pathway (Figure 2) and the somewhat bimodal distribution observed for condensed forms (Supplementary Figure 3b). However, the regulation of condensed forms appears to be a rather complex, quantitative trait, at least in this AR population.

Mapping flavone content

C-glycosyl flavones can largely influence anthocyanin extract hue and intensity and are therefore an important consideration in a breeding program aiming to optimize hue¹¹. They are also the precursors to maysin, a C-glycosyl flavone that provides resistance to corn earworm (*Helicoverpa zea*)⁴⁸. Surprisingly, we were not able to identify significant quantities of maysin in the AR population, suggesting fixation of a null *Sm1* (*Salmon silks1*) or *Sm2* (*Salmon silks2*) allele. *Sm1* and *Sm2* encode the rhamnose synthase and rhamnose transferase that convert C-hexosyl flavones into maysin or apimaysin⁴⁹ (Figure 2). As determined by HPLC and confirmed using mass spectrometry¹⁴ the major C-glycosyl flavones present in AR lines were di-C, C-hexosyl apigenin, C-hexosyl-C-pentosyl apigenin, C-pentosyl-C-hexosyl apigenin, and C-hexosyl apigenin.

Mapping flavone content was performed with and without a covariate for condensed forms, but many of the same loci were detected as for condensed forms (Table 4). As expected, a

large signal near *PI* was identified (Figure 5A). However, the signal was wide, ranging from 20 Mb to 53 Mb with the most significant SNP around 40 Mb, more than 8 Mb away from *PI*. Several loci involved in sugar metabolism are near *PI* and are regulated by it³⁵. While it's possible that this signal represents multiple signals, it is equally possible that linkage blocks in this area are contributing to the signal width (Supplementary Figure 11). In addition to this major locus, a signal was identified on Chr 1, 1.5 Mb from *Chi1*, which encodes a chalcone isomerase that functions prior to the branch point in the pathway for flavones or anthocyanins. We also detected a peak around 216 Mb found previously when mapping the proportion of condensed forms (Supplementary Figure 12).

While not detected for condensed forms, we also found an isoflavone reductase-like protein on Chr 3 (*Ir11*, Zm00001d040173, 30.6 Mb) which is a close blast match to both the LAR from *Medicago* and an LAR from *Desmodium uncinatum* (a member of the isoflavone reductase-like group of the reductase-epimerase-dehydrogenase protein family) (Supplementary Figure 9)^{50,51}. While this candidate may be more appropriate for the biosynthesis of condensed forms than flavones as a response variable, this locus should not be overlooked as a strong candidate for a maize orthologous LAR due to the correlation between flavones and condensed forms in this population. Also unexpected was a relatively weak signal on Chr 2, 2 Mb away from *BI*, a plant-specific MYB activator of anthocyanin biosynthesis⁵². If this locus contributed to flux toward the anthocyanin side of the pathway, it could influence flavone content as well given these two compound classes share common precursors.

Unsurprisingly, we also found a strong signal near *Fns1*, the flavone synthase required for flavone biosynthesis in maize⁵³. This confirms both that the same pathway is functioning in Apache Red, as well as the power of this dataset to detect loci underlying flavonoid biosynthesis despite the number of strong correlations between response variables. Also expected, was a signal near *Pr1*. The majority of flavones detected in AR were apigenin-derived, containing a single hydroxyl group on the B-ring. Rather than an inability to produce luteolin compounds (two hydroxyl groups on the B-ring) though, we suspect this is more likely due to trait linkage resulting from a fixed recessive *pr1* in AR families that produced large quantities of flavones. Intercrossing between *Pr1*/- lines (evident by cyanidin-dominant anthocyanins) and high flavone lines would likely produce luteolin-derived flavones; however, this should be tested to confirm. Production of different flavone types in AR may also be useful for investigating the

copigmentation properties of various flavone-anthocyanin combinations. Other signals identified when mapping flavones, but without promising candidates, are listed in Table 4.

With data quantifying different flavone species from HPLC analysis we were also able to map the proportion of each flavone type relative to total flavone content. We again identified a signal near *Fns1*, as well as one on Chr 6 near *Cgt1*, a dual C-/O-glycosyltransferase implicated in the formation of C-hexosyl flavones⁵⁴. *Cgt1* is capable of C-glycosylating at both the 6-C and 8-C positions, but this is the first indication of its role in creating di-C,C-hexosyl flavones, glycosylated at both positions on the same molecule. While it is possible that *Cgt1* is also capable of C-glycosylating pentose moieties, it is difficult to separate the functions in this data set as no C-pentosyl flavones were identified that did not also contain a hexose. In addition to these signals, peaks were found on Chrs 5, 7, and 8, which are listed in Table 4. No obvious candidates were identified for these, but several plausible loci were considered. For the peak on Chr 7 with the most significant SNP located at 144.83 Mb and spanning 138–146 Mb, two glycosyltransferases were found around 144 Mb (Zm00001d021167 and Zm00001d021168). This signal is also close to *myb152* (148.15 Mb) a positive regulator of the phenylpropanoid pathway⁵⁵. We also found a MYB transcription factor close to the signal on Chr 8 (*myb147*) as well as a flavonoid 3'-hydroxylase orthologous to Transparent Testa 7 (*TT7*), which functions similarly to *Pr1* in maize, determining the ratio of kaempferol to quercetin in Arabidopsis seed coats⁵⁶. A table and phylogram of candidate MYBs and known flavonoid MYBs from other species can be found in the supplementary materials of Chatham and Juvik, submitted to G3. Variations in flavone structure could influence copigmentation efficacy²⁷, however more information is needed before specific breeding objectives for optimal flavone content can be defined.

Applications

The development of separate flavone-rich lines and anthocyanin-rich lines is warranted to allow for the maximum range of applications. Ideally, a pelargonidin-rich line producing orange extract, a cyanidin-rich line producing red-pink extract, and a colorless line rich in C-glycosyl flavones with the greatest capacity for copigmentation would be bred. Doing this would allow a manufacturer to design a range of hues for use in food and beverage products by mixing different ratios of the pelargonidin extract, cyanidin extract, and flavone extract, as described previously¹¹.

Anthocyanin-rich lines should be further optimized to maximize stability and anthocyanin content. Acylated anthocyanins should be maximized, while condensed forms should be minimized. Reports of improved stability from condensed forms are sparse, and condensed forms were associated with flavone content in this population, something that should be minimized in breeding anthocyanin lines. Condensed forms could also represent a reduction in overall content, since two units (1 flavonol + 1 anthocyanin) are required to make one condensed form.

The development of markers for the candidates and known genes identified herein to be associated with anthocyanin type and decoration could improve selection efficiency. For example, pelargonidin-rich lines could be selected at the seedling stage by choosing *prl/prl* lines and could be tracked through backcrossing by differentiating between *Pr1/prl* and *Pr1/Pr1* individuals to avoid subsequent test crosses ensuring maintenance of the recessive allele). Combined with markers for *Pr1*, the identification of a methyltransferase candidate for peonidin content could enable the development of markers that aid in creating peonidin-dominant lines (requiring both *Pr1/-* and a functional *AOMT* due to their complimentary gene action). Based on results presented here, purple corn breeding programs should select for nonfunctional *pl* and nonfunctional ANR or LAR candidates to eliminate flavones and condensed forms while selecting for functional *Aat1* to maintain acylated anthocyanins.

Conclusions

Together, the diverse range of flavonoids present in AR provide an assortment of specialized metabolite profiles that can be modified through breeding. Variation in flavonoid profiles may lead to wider color ranges, improved stability, and intensified color for the use of purple corn in foods and beverages.

This survey provides a comprehensive analysis of anthocyanin species diversity in purple corn and highlights the importance of AR as breeding material in programs selecting for natural colorants. Through mapping, we have provided evidence to confirm the function of *Pr1* and *Aat1* in pericarp in addition to aleurone, a fact that has been assumed but not previously tested. Furthermore, several candidate loci involved in the partitioning of various anthocyanin species affecting the resulting extract hue and stability were identified. Among these, a candidate methyltransferase was identified and believed to be associated with peonidin content. If confirmed with additional studies, this would be the first characterized methyltransferase with

anthocyanin *O*-methyltransferase-like activity in a cereal. This study also represents the first investigation of genetic factors underlying the formation of flavanol-anthocyanin condensed forms. The candidate genes identified here will provide a foundation for future research on loci associated with anthocyanin species and decorations.

List of Abbreviations Used:

ABC: ATP-Binding Cassette, **AR:** Apache Red, **bHLH:** basic Helix-Loop-Helix, **GBS:** Genotyping-By-Sequencing, **HPLC:** High-Performance Liquid Chromatography, **MAF:** Minor Allele Frequency, **MATE:** Multidrug and toxic compound extrusion **MYB:** derived from “Myeloblastosis protein”, **PCA:** Principal Component Analysis, **SNP:** Single Nucleotide Polymorphism, **QTL:** Quantitative Trait Loci.

Declarations:

Ethics Approval

Not applicable

Consent for Publication

Not applicable

Availability of Data and Material

Lines from the Apache Red population are available by request. All genotype and phenotype datasets will be made available at the time of publication via Figshare under project name “Linking anthocyanin diversity, hue, and genetics in purple corn”.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from the Kraft Heinz Company of Glenview, Illinois and from DD Williamson, Inc. of Louisville, Kentucky. Fellowship support for LAC was provided by the Illinois Corn Grower's Association and PEO Scholar Award.

Author's Contributions

LAC developed, genotyped, and phenotyped the Apache Red population, analyzed the data, and prepared the manuscript. JAJ assisted in conceiving the study and revising the manuscript. All authors read and approved the manuscript.

Acknowledgements

We would like to thank Dr. Pat Brown for assisting with GBS library construction.

References

- (1) Paulsmeyer, M.; Chatham, L.; Becker, T.; West, M.; West, L.; Juvik, J. Survey of Anthocyanin Composition and Concentration in Diverse Maize Germplasms. *J. Agric. Food Chem.* **2017**, *65* (21), 4341–4350. <https://doi.org/10.1021/acs.jafc.7b00771>.
- (2) Li, Q.; Somavat, P.; Singh, V.; Chatham, L.; Gonzalez de Mejia, E. A Comparative Study of Anthocyanin Distribution in Purple and Blue Corn Coproducts from Three Conventional Fractionation Processes. *Food Chemistry* **2017**, *231*, 332–339. <https://doi.org/10.1016/j.foodchem.2017.03.146>.
- (3) Somavat, P.; Kumar, D.; Singh, V. Techno-Economic Feasibility Analysis of Blue and Purple Corn Processing for Anthocyanin Extraction and Ethanol Production Using Modified Dry Grind Process. *Industrial Crops and Products* **2018**, *115*, 78–87. <https://doi.org/10.1016/j.indcrop.2018.02.015>.
- (4) Ananga, A.; Georgiev, V.; Ochieng, J.; Tsoleva, B. P. and V. Production of Anthocyanins in Grape Cell Cultures: A Potential Source of Raw Material for Pharmaceutical, Food, and Cosmetic Industries. *The Mediterranean Genetic Code - Grapevine and Olive* **2013**. <https://doi.org/10.5772/54592>.
- (5) Zhao, C.-L.; Yu, Y.-Q.; Chen, Z.-J.; Wen, G.-S.; Wei, F.-G.; Zheng, Q.; Wang, C.-D.; Xiao, X.-L. Stability-Increasing Effects of Anthocyanin Glycosyl Acylation. *Food Chemistry* **2017**, *214*, 119–128. <https://doi.org/10.1016/j.foodchem.2016.07.073>.
- (6) Haggard, S.; Luna-Vital, D.; West, L.; Juvik, J. A.; Chatham, L.; Paulsmeyer, M.; Gonzalez de Mejia, E. Comparison of Chemical, Color Stability, and Phenolic Composition from Pericarp of Nine Colored Corn Unique Varieties in a Beverage Model. *Food Research International* **2018**, *105*, 286–297. <https://doi.org/10.1016/j.foodres.2017.11.038>.
- (7) Luna-Vital, D.; Li, Q.; West, L.; West, M.; Gonzalez de Mejia, E. Anthocyanin Condensed Forms Do Not Affect Color or Chemical Stability of Purple Corn Pericarp Extracts Stored under Different PHs. *Food Chemistry* **2017**, *232*, 639–647. <https://doi.org/10.1016/j.foodchem.2017.03.169>.
- (8) Cone, K. C. Anthocyanin Synthesis in Maize Aleurone Tissue. In *Endosperm*; Olsen, O.-A., Ed.; Plant Cell Monographs; Springer Berlin Heidelberg, 2007; pp 121–139. https://doi.org/10.1007/7089_2007_117.

- (9) Sharma, M.; Cortes-Cruz, M.; Ahern, K. R.; McMullen, M.; Brutnell, T. P.; Chopra, S. Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. *Genetics* **2011**, *188* (1), 69–79. <https://doi.org/10.1534/genetics.110.126136>.
- (10) Paulsmeyer, M. N.; Brown, P. J.; Juvik, J. A. Discovery of Anthocyanin Acyltransferase1 (AAT1) in Maize Using Genotyping-by-Sequencing (GBS). *G3: Genes, Genomes, Genetics* **2018**, *8* (11), 3669–3678. <https://doi.org/10.1534/g3.118.200630>.
- (11) Chatham, L. A.; Howard, J. E.; Juvik, J. A. A Natural Colorant System from Corn: Flavone-Anthocyanin Copigmentation for Altered Hues and Improved Shelf Life. *Food Chemistry* **2019**, 125734. <https://doi.org/10.1016/j.foodchem.2019.125734>.
- (12) Grotewold, E.; Chamberlin, M.; Snook, M.; Siame, B.; Butler, L.; Swenson, J.; Maddock, S.; Clair, G. S.; Bowen, B. Engineering Secondary Metabolism in Maize Cells by Ectopic Expression of Transcription Factors. *The Plant Cell* **1998**, *10* (5), 721–740. <https://doi.org/10.1105/tpc.10.5.721>.
- (13) Galaffu, N.; Bortlik, K.; Michel, M. 5 - An Industry Perspective on Natural Food Colour Stability. In *Colour Additives for Foods and Beverages*; Scotter, M. J., Ed.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Woodhead Publishing: Oxford, 2015; pp 91–130. <https://doi.org/10.1016/B978-1-78242-011-8.00005-2>.
- (14) Chatham, L. A.; West, L.; Berhow, M. A.; Vermillion, K. E.; Juvik, J. A. Unique Flavanol-Anthocyanin Condensed Forms in Apache Red Purple Corn. *J. Agric. Food Chem.* **2018**. <https://doi.org/10.1021/acs.jafc.8b04723>.
- (15) Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S.; Mitchell, S. E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE* **2011**, *6* (5), e19379. <https://doi.org/10.1371/journal.pone.0019379>.
- (16) Poland, J. A.; Brown, P. J.; Sorrells, M. E.; Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* **2012**, *7* (2), e32253. <https://doi.org/10.1371/journal.pone.0032253>.
- (17) Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics* **2007**, *23* (19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>.
- (18) Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M. C.; Wang, B.; Campbell, M. S.; Stein, J. C.; Wei, X.; Chin, C.-S.; et al. Improved Maize Reference Genome with Single-Molecule Technologies. *Nature* **2017**. <https://doi.org/10.1038/nature22971>.
- (19) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat Methods* **2012**, *9* (4), 357–359. <https://doi.org/10.1038/nmeth.1923>.
- (20) Browning, B. L.; Zhou, Y.; Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* **2018**, *103* (3), 338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>.
- (21) Lipka, A. E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P. J.; Gore, M. A.; Buckler, E. S.; Zhang, Z. GAPIT: Genome Association and Prediction Integrated Tool. *Bioinformatics* **2012**, *28* (18), 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>.
- (22) Huguene, P.; Provenzano, S.; Verriès, C.; Ferrandino, A.; Meudec, E.; Batelli, G.; Merdinoglu, D.; Cheynier, V.; Schubert, A.; Ageorges, A. A Novel Cation-Dependent O-Methyltransferase Involved in Anthocyanin Methylation in Grapevine. *Plant Physiology* **2009**, *150* (4), 2057–2070. <https://doi.org/10.1104/pp.109.140376>.

- (23) Provenzano, S.; Spelt, C.; Hosokawa, S.; Nakamura, N.; Brugliera, F.; Demelis, L.; Geerke, D. P.; Schubert, A.; Tanaka, Y.; Quattrocchio, F.; et al. Genetic Control and Evolution of Anthocyanin Methylation. *PLANT PHYSIOLOGY* **2014**, *165* (3), 962–977. <https://doi.org/10.1104/pp.113.234526>.
- (24) Du, H.; Wu, J.; Ji, K.-X.; Zeng, Q.-Y.; Bhuiya, M.-W.; Su, S.; Shu, Q.-Y.; Ren, H.-X.; Liu, Z.-A.; Wang, L.-S. Methylation Mediated by an Anthocyanin, O-Methyltransferase, Is Involved in Purple Flower Coloration in Paeonia. *J Exp Bot* **2015**, *66* (21), 6563–6577. <https://doi.org/10.1093/jxb/erv365>.
- (25) Wright, E., S. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* **2016**, *8* (1), 352. <https://doi.org/10.32614/RJ-2016-025>.
- (26) Schliep, K. P. Phangorn: Phylogenetic Analysis in R. *Bioinformatics* **2011**, *27* (4), 592–593. <https://doi.org/10.1093/bioinformatics/btq706>.
- (27) Trouillas, P.; Sancho-García, J. C.; De Freitas, V.; Gierschner, J.; Otyepka, M.; Dangles, O. Stabilizing and Modulating Color by Copigmentation: Insights from Theory and Experiment. *Chem. Rev.* **2016**, *116* (9), 4937–4982. <https://doi.org/10.1021/acs.chemrev.5b00507>.
- (28) Giusti, M. M.; Wrolstad, R. E. Characterization and Measurement of Anthocyanins by UV-Visible Spectroscopy. *Current Protocols in Food Analytical Chemistry* **2001**, *00* (1), F1.2.1-F1.2.13. <https://doi.org/10.1002/0471142913.faf0102s00>.
- (29) Dipalmo, T.; Crupi, P.; Pati, S.; Clodoveo, M. L.; Di Luccia, A. Studying the Evolution of Anthocyanin-Derived Pigments in a Typical Red Wine of Southern Italy to Assess Its Resistance to Aging. *LWT - Food Science and Technology* **2016**, *71*, 1–9. <https://doi.org/10.1016/j.lwt.2016.03.012>.
- (30) Chen, A. H.; Lipka, A. E. The Use of Targeted Marker Subsets to Account for Population Structure and Relatedness in Genome-Wide Association Studies of Maize (*Zea Mays* L.). *G3: Genes, Genomes, Genetics* **2016**, *6* (8), 2365–2374. <https://doi.org/10.1534/g3.116.029090>.
- (31) Coe, E. H., Jr.; Neuffer, M. G.; Hoisington, D. A. The Genetics of Corn. **1988**, 81–258.
- (32) Larson, R. L. Flavonoid 3'-O-Methylation by a *Zea Mays* L. Preparation1. *Biochemie und Physiologie der Pflanzen* **1989**, *184* (5), 453–460. [https://doi.org/10.1016/S0015-3796\(89\)80046-0](https://doi.org/10.1016/S0015-3796(89)80046-0).
- (33) Liu, D.; Li, S.; Chen, W.; Zhang, B.; Liu, D.; Liu, B.; Zhang, H. Transcriptome Analysis of Purple Pericarps in Common Wheat (*Triticum Aestivum* L.). *PLOS ONE* **2016**, *11* (5), e0155428. <https://doi.org/10.1371/journal.pone.0155428>.
- (34) Abdel-Aal, E.-S. M.; Young, J. C.; Rabalski, I. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J. Agric. Food Chem.* **2006**, *54* (13), 4696–4704. <https://doi.org/10.1021/jf0606609>.
- (35) Morohashi, K.; Casas, M. I.; Ferreyra, M. L. F.; Mejía-Guerra, M. K.; Pourcel, L.; Yilmaz, A.; Feller, A.; Carvalho, B.; Emiliani, J.; Rodriguez, E.; et al. A Genome-Wide Regulatory Framework Identifies Maize Pericarp Color1 Controlled Genes. *The Plant Cell* **2012**, *24* (7), 2745–2764. <https://doi.org/10.1105/tpc.112.098004>.
- (36) Liu, C.; Wang, X.; Shulaev, V.; Dixon, R. A. A Role for Leucoanthocyanidin Reductase in the Extension of Proanthocyanidins. *Nature Plants* **2016**, *2* (12), 16182. <https://doi.org/10.1038/nplants.2016.182>.
- (37) Jun, J. H.; Liu, C.; Xiao, X.; Dixon, R. A. The Transcriptional Repressor MYB2 Regulates Both Spatial and Temporal Patterns of Proanthocyanidin and Anthocyanin Pigmentation in

- Medicago Truncatula. *The Plant Cell* **2015**, 27 (10), 2860–2879. <https://doi.org/10.1105/tpc.15.00476>.
- (38) Li, S.; Zachgo, S. TCP3 Interacts with R2R3-MYB Proteins, Promotes Flavonoid Biosynthesis and Negatively Regulates the Auxin Response in Arabidopsis Thaliana. *The Plant Journal* **2013**, 76 (6), 901–913. <https://doi.org/10.1111/tpj.12348>.
 - (39) Marrs, K. A.; Alfenito, M. R.; Lloyd, A. M.; Walbot, V. A Glutathione S-Transferase Involved in Vacuolar Transfer Encoded by the Maize Gene Bronze-2. *Nature* **1995**, 375 (6530), 397–400. <https://doi.org/10.1038/375397a0>.
 - (40) Carey, C. C.; Strahle, J. T.; Selinger, D. A.; Chandler, V. L. Mutations in the Pale Aleurone Color1 Regulatory Gene of the Zea Mays Anthocyanin Pathway Have Distinct Phenotypes Relative to the Functionally Similar TRANSPARENT TESTA GLABRA1 Gene in Arabidopsis Thaliana. *Plant Cell* **2004**, 16 (2), 450–464. <https://doi.org/10.1105/tpc.018796>.
 - (41) Chatham, L. A.; Paulsmeyer, M.; Juvik, J. A. Prospects for Economical Natural Colorants: Insights from Maize. *Theor Appl Genet* **2019**, 132 (11), 2927–2946. <https://doi.org/10.1007/s00122-019-03414-0>.
 - (42) Selinger, D. A.; Chandler, V. L. A Mutation in the Pale Aleurone Color1 Gene Identifies a Novel Regulator of the Maize Anthocyanin Pathway. *The Plant Cell* **1999**, 11 (1), 5–14. <https://doi.org/10.1105/tpc.11.1.5>.
 - (43) Liu, J.; Osbourn, A.; Ma, P. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Molecular Plant* **2015**, 8 (5), 689–708. <https://doi.org/10.1016/j.molp.2015.03.012>.
 - (44) Zhu, H.-F.; Fitzsimmons, K.; Khandelwal, A.; Kranz, R. G. CPC, a Single-Repeat R3 MYB, Is a Negative Regulator of Anthocyanin Biosynthesis in Arabidopsis. *Molecular Plant* **2009**, 2 (4), 790–802. <https://doi.org/10.1093/mp/ssp030>.
 - (45) Albert, N. W.; Lewis, D. H.; Zhang, H.; Schwinn, K. E.; Jameson, P. E.; Davies, K. M. Members of an R2R3-MYB Transcription Factor Family in Petunia Are Developmentally and Environmentally Regulated to Control Complex Floral and Vegetative Pigmentation Patterning. *The Plant Journal* **2011**, 65 (5), 771–784. <https://doi.org/10.1111/j.1365-3113.2010.04465.x>.
 - (46) Styles, E. D.; Coe, E. H. Unstable Expression of an R Allele with A3 in Maize A Recessive Intensifier of Plant Color. *J Hered* **1986**, 77 (6), 389–393.
 - (47) Verweij, W.; Spelt, C.; Sansebastiano, G.-P. D.; Vermeer, J.; Reale, L.; Ferranti, F.; Koes, R.; Quattrocchio, F. An H⁺ P-ATPase on the Tonoplast Determines Vacuolar PH and Flower Colour. *Nature Cell Biology* **2008**, 10 (12), 1456–1462. <https://doi.org/10.1038/ncb1805>.
 - (48) Byrne, P. F.; McMullen, M. D.; Snook, M. E.; Musket, T. A.; Theuri, J. M.; Widstrom, N. W.; Wiseman, B. R.; Coe, E. H. Quantitative Trait Loci and Metabolic Pathways: Genetic Control of the Concentration of Maysin, a Corn Earworm Resistance Factor, in Maize Silks. *PNAS* **1996**, 93 (17), 8820–8825. <https://doi.org/10.1073/pnas.93.17.8820>.
 - (49) Casas, M. I.; Falcone-Ferryra, M. L.; Jiang, N.; Guerra, M. K. M.; Rodriguez, E. J.; Wilson, T.; Engelmeier, J.; Casati, P.; Grotewold, E. Identification and Characterization of Maize Salmon Silks Genes Involved in Insecticidal Maysin Biosynthesis. *The Plant Cell* **2016**, tpc.00003.2016. <https://doi.org/10.1105/tpc.16.00003>.

- (50) Dixon, R. A.; Xie, D.-Y.; Sharma, S. B. Proanthocyanidins – a Final Frontier in Flavonoid Research? *New Phytologist* **2005**, *165* (1), 9–28. <https://doi.org/10.1111/j.1469-8137.2004.01217.x>.
- (51) Tanner, G. J.; Francki, K. T.; Abrahams, S.; Watson, J. M.; Larkin, P. J.; Ashton, A. R. Proanthocyanidin Biosynthesis in Plants PURIFICATION OF LEGUME LEUCOANTHOCYANIDIN REDUCTASE AND MOLECULAR CLONING OF ITS CDNA. *J. Biol. Chem.* **2003**, *278* (34), 31647–31656. <https://doi.org/10.1074/jbc.M302783200>.
- (52) Lloyd, A.; Brockman, A.; Aguirre, L.; Campbell, A.; Bean, A.; Cantero, A.; Gonzalez, A. Advances in the MYB–BHLH–WD Repeat (MBW) Pigment Regulatory Model: Addition of a WRKY Factor and Co-Option of an Anthocyanin MYB for Betalain Regulation. *Plant Cell Physiol* **2017**, *58* (9), 1431–1441. <https://doi.org/10.1093/pcp/pcx075>.
- (53) Ferreyra, M. L. F.; Emiliani, J.; Rodriguez, E. J.; Campos-Bermudez, V. A.; Grotewold, E.; Casati, P. The Identification of Maize and Arabidopsis Type I FLAVONE SYNTHASEs Links Flavones with Hormones and Biotic Interactions. *Plant Physiology* **2015**, *169* (2), 1090–1107. <https://doi.org/10.1104/pp.15.00515>.
- (54) Ferreyra, M. L. F.; Rodriguez, E.; Casas, M. I.; Labadie, G.; Grotewold, E.; Casati, P. Identification of a Bifunctional Maize C- and O-Glucosyltransferase. *J. Biol. Chem.* **2013**, *288* (44), 31678–31688. <https://doi.org/10.1074/jbc.M113.510040>.
- (55) Agarwal, T.; Grotewold, E.; Doseff, A. I.; Gray, J. MYB31/MYB42 Syntelogs Exhibit Divergent Regulation of Phenylpropanoid Genes in Maize, Sorghum and Rice. *Scientific Reports* **2016**, *6* (1). <https://doi.org/10.1038/srep28502>.
- (56) Schoenbohm, C.; Martens, S.; Eder, C.; Forkmann, G.; Weisshaar, B. Identification of the Arabidopsis Thaliana Flavonoid 3'-Hydroxylase Gene and Functional Expression of the Encoded P450 Enzyme. *Biological Chemistry* **2000**, *381* (8).

Tables and Figures

Table 1: Diversity in anthocyanin content of 1152 Apache Red lines

Category	Range
Acylated anthocyanins	2% - 72%
Malonylated	0% - 43%
Dimalonylated	0% - 57%
Pelargonidin-derived anthocyanins	5% - 99%
Cyanidin-derived anthocyanins	0% - 86%
Peonidin-derived anthocyanins	0% - 50%
Flavanol-anthocyanin condensed forms	0% - 83%
Total flavone content	0-1904 $\mu\text{g/g}$
Total anthocyanin content	0-1598 $\mu\text{g/g}$

Table 2: Summary of variables used for PCA and loadings for PC1 and PC2. Peaks correspond to Figure 2B.

Peak	Anthocyanin	PC1	PC2
1	CaCydG catechin-(4,8)-cyanidin-3,5-diglucoside	0.144	0.404
2	CaPgG catechin-(4,8)-pelargonidin-3,5-diglucoside		0.401
3	CatPndG catechin-(4,8)-peonidin-3,5-diglucoside	0.102	0.37
4	AfCydG afzelechin-(4,8)-cyanidin-3,5-diglucoside	0.111	0.372
5	CaCyMdG catechin-(4,8)-cyanidin-3-malonylglucoside-5-glucoside	0.172	0.284
6	CaPgMdG catechin-(4,8)-pelargonidin-3-malonylglucoside-5-glucoside		0.158
7	AfPgG afzelechin-(4,8)-pelargonidin-3,5-diglucoside	-0.146	0.185
8	AfPndG afzelechin-(4,8)-peonidin-3,5-diglucoside		0.297
9	CyG cyanidin 3-glucoside	0.35	
10	PgG pelargonidin 3-glucoside	-0.372	
11	AfPgMdG afzelechin-(4,8)-cyanidin-3-malonylglucoside-5-glucoside	0.269	
12	PnG peonidin 3-glucoside	0.277	-0.113
13	CyMGs cyanidin malonylglucosides	0.338	-0.117
14	PgMGs pelargonidin malonylglucosides	-0.331	-0.145
15	PnMG peonidin 3-malonylglucoside	0.275	-0.152
16	CydiMG cyanidin 3-5-dimalonylglucoside	0.172	
17	PgMdiGs pelargonidin dimalonylglucosides	-0.321	
18	PndiMG peonidin 3-5-dimalonylglucoside		
	Flavone sum of di-C,C-hexosyl apigenin, C-hexosyl-C-pentosyl apigenin, C-pentosyl-C-hexosyl apigenin, and C-hexosyl apigenin	-0.181	0.3

Table 3: Anthocyanin structural genes corresponding to abbreviated genes and gene products labeled in Figure 2.

	Locus	Gene product	Gene model identifier	Chromosome	Mb
<u>Structural</u>					
<i>c2</i>	colorless2	CHS (chalcone synthase)	Zm00001d052673	Chr 4	196.9
<i>whp1</i>	white pollen1		Zm00001d007403	Chr 2	231.0
<i>chi1</i>	chalcone flavanone isomerase1	CHI (chalcone isomerase)	Zm00001d034635	Chr 1	298.6
<i>chi3</i>	chalcone flavanone isomerase3		Zm00001d012972	Chr 5	2.6
<i>chi4</i>	chalcone flavanone isomerase4		Zm00001d018278	Chr 5	217.8
<i>chi5</i>	chalcone flavanone isomerase5		Zm00001d016144	Chr 5	146.5
<i>pr1</i>	red aleurone1	F3'H (flavonoid 3'-hydroxylase)	Zm00001d017077	Chr 5	184.5
<i>Fht1</i>	flavanone 3-hydroxylase1	F3H (flavonone 3-hydroxylase)	Zm00001d001960	Chr 2	3.6
<i>Fht2</i>	flavanone 3-hydroxylase2		Zm00001d029218	Chr 1	62.7
<i>A1</i>	anthocyaninless1	DFR / F4R (bifunctional	Zm00001d044122	Chr 3	219.9
<i>A4</i>	anthocyaninless4	dihydroflavonol 4-reductase / flavanone 4-reductase)	Zm00001d011438	Chr 8	150.5
<i>A2</i>	anthocyaninless2	ANS (Anthocyanidin Synthase / Leucoanthocyanidin Dioxygenase)	Zm00001d014914	Chr 5	68.0
<i>Fns1</i>	flavone synthase1	F2H (flavonone 2-hydroxylase)	Zm00001d047452	Chr 9	130.7
<i>Fnsii1</i>	flavone synthase typeII1	FNS (flavone synthase)	Zm00001d024946	Chr 10	96.3
			Zm00001d024943	Chr 10	96.1
<i>Cgt1</i>	c-glucosyl transferase1	CGT (c-glucosyl transferase)	Zm00001d037382	Chr 6	123.8
<i>Sm1</i>	salmon silk1	RHS (rhamnose synthase)	Zm00001d037784	Chr 6	137.7
<i>Sm2</i>	salmon silk2	RHT (rhamnosyl transferase)	Zm00001d006446	Chr 2	208.5
<i>Bz1</i>	bronze1	UFGT (UDP-glucose flavonol	Zm00001d045055	Chr 9	11.2
<i>Ufgt2</i>	UDP-flavonol-glycosyltransferase2	glycosyltransferase)	Zm00001d019256	Chr 7	24.2
<i>Aat1</i>	anthocyanin acyltransferase1	AAT (anthocyanin acyltransferase)	Zm00001d034925	Chr 1	305.7

Table 4: Significant SNPs* for each anthocyanin variable mapped

Variable	Range of significant SNPs			SNP with minimum p-value					
	Chr	Min	Max	Position (Mb)	P value	Effect	MAF		
		position (Mb)	position (Mb)						
Anthocyanidin Type	Proportion cyanidin	4	152.22	152.22	152.22	1.54E-06	-0.09	0.15	
		** 5	168.75	211.21	184.86	6.96E-105	0.31	0.48	
		** 6	71.57	71.57	71.57	4.10E-08	-0.13	0.03	
		** 6	161.00	161.00	161.00	6.09E-08	0.12	0.03	
		8	0.82	0.82	0.82	2.94E-06	0.61	0.50	
		10	104.61	104.61	104.61	3.51E-06	0.12	0.02	
	Proportion peonidin / (cyanidin + peonidin)	** 1	64.04	64.04	64.04	2.72E-07	-0.15	0.17	
		** 2	142.74	142.74	142.74	5.20E-08	-0.09	0.06	
		** 3	20.69	20.69	20.69	4.00E-07	0.10	0.02	
		** 4	157.82	240.70	195.50	2.13E-17	0.07	0.31	
		** 5	172.18	196.29	184.02	5.01E-30	-0.08	0.50	
		6	88.67	88.67	88.67	1.21E-06	0.07	0.04	
		7	131.25	131.58	131.58	1.77E-06	-0.06	0.05	
		9	23.85	23.85	23.85	1.35E-06	-0.15	0.03	
	Anthocyanin Decorations - Acylation	Proportion of acylated anthocyanins	** 1	17.95	45.12	35.49	1.28E-08	0.08	0.10
			** 1	190.18	190.18	190.18	4.92E-07	-0.09	0.03
			** 1	209.72	209.72	209.72	4.73E-07	-0.58	0.50
			** 1	233.55	233.55	233.55	1.59E-07	-0.11	0.03
** 1			249.31	249.31	249.31	1.60E-08	0.14	0.10	
** 1			298.41	306.42	306.42	2.47E-58	-0.17	0.19	
** 2			8.13	8.13	8.13	5.96E-07	0.07	0.10	
** 3			143.43	174.37	143.43	3.98E-08	0.06	0.09	
** 3			221.63	221.63	221.63	8.73E-08	-0.08	0.07	
** 4			2.54	5.97	5.97	1.02E-09	-0.15	0.02	
** 4			160.96	160.96	160.96	2.60E-08	-0.08	0.04	
** 5			36.21	36.21	36.21	3.74E-07	-0.07	0.08	
** 5			213.51	213.51	213.51	6.84E-07	0.14	0.01	
** 6			107.49	121.60	107.49	2.65E-08	-0.11	0.03	
9			23.85	23.85	23.85	8.87E-07	0.17	0.03	
** 10			45.38	45.38	45.38	2.17E-09	-0.05	0.12	
** 10			115.12	115.12	115.12	4.13E-12	0.11	0.04	
Condensed d Forms			Total condensed forms	** 1	19.59	53.46	39.96	3.11E-37	-5.90
	** 1	290.23		297.07	297.07	4.50E-18	4.07	0.05	
	** 2	130.57		130.57	130.57	5.25E-07	3.75	0.05	
	2	204.61		226.78	226.78	1.98E-06	-2.24	0.07	
	** 3	1.40		9.15	1.40	1.55E-08	-4.97	0.10	
	** 4	78.68		84.78	78.68	4.63E-07	2.39	0.07	
	** 4	134.03		140.38	134.03	3.05E-07	-1.94	0.12	
	4	192.69		192.69	192.69	1.54E-06	2.95	0.05	
	6	22.36		22.36	22.36	1.86E-06	2.67	0.05	
	** 6	113.93		117.74	117.74	5.37E-07	-1.88	0.05	
	7	142.26		142.26	142.26	1.15E-06	-4.27	0.10	
	** 7	177.59		177.72	177.72	1.25E-09	2.85	0.08	
	** 8	20.26		20.26	20.26	4.76E-09	2.38	0.14	
	** 9	11.75		11.75	11.75	1.87E-07	1.51	0.34	
	** 9	127.65		147.80	138.71	1.54E-13	3.42	0.07	
	** 1	17.95		51.84	39.96	1.77E-19	-3.47	0.23	
	1	293.75		293.75	293.75	2.06E-06	1.71	0.07	
	** 2	95.50		95.50	95.50	4.10E-07	1.58	0.10	

Anthocyanin Decorations - Flavanol Anthocyanin Condense	<i>Total condensed forms with flavone covariate</i>	**	2	225.98	226.78	226.78	5.93E-07	-1.96	0.07
		**	3	1.40	1.40	1.40	1.69E-07	-3.41	0.10
		**	3	52.09	52.09	52.09	2.40E-07	2.53	0.03
		**	5	2.34	2.34	2.34	3.04E-06	-1.97	0.07
		**	5	197.14	213.51	203.30	9.14E-10	-1.25	0.44
		**	6	96.57	96.57	96.57	2.19E-06	1.93	0.07
		**	6	109.57	116.70	113.98	9.79E-10	1.63	0.15
		**	7	142.26	142.26	142.26	1.70E-06	-3.08	0.10
		**	7	177.72	177.72	177.72	2.25E-10	2.56	0.08
		**	8	20.26	20.26	20.26	1.17E-06	1.61	0.14
	<i>Percent condensed forms</i>	**	8	164.36	164.36	164.36	3.83E-07	-2.38	0.03
		**	9	11.75	39.86	11.75	8.98E-09	1.39	0.34
		**	9	129.91	154.35	154.35	4.93E-09	-2.73	0.03
		**	10	104.61	115.12	104.61	1.23E-08	-3.08	0.02
		**	1	14.36	53.46	35.49	9.75E-22	-0.10	0.10
		**	1	216.31	216.31	216.31	2.25E-06	-0.06	0.07
		**	1	257.23	257.23	257.23	5.96E-07	0.03	0.07
		**	1	275.71	284.54	275.71	3.07E-14	0.09	0.07
		**	1	295.64	299.61	299.20	7.22E-11	-0.04	0.30
		**	1	300.19	306.42	306.42	2.13E-46	0.10	0.19
	<i>Total flavone content</i>	**	3	1.40	1.40	1.40	8.24E-08	-0.12	0.10
		**	3	18.64	18.64	18.64	1.49E-06	0.04	0.10
		**	3	121.83	121.83	121.83	5.78E-09	-0.05	0.09
		**	3	146.59	174.37	146.59	2.09E-11	0.06	0.06
		**	3	220.49	220.49	220.49	3.20E-07	0.05	0.20
		**	4	176.64	176.64	176.64	1.40E-07	0.04	0.07
		**	4	238.18	238.18	238.18	6.38E-08	0.04	0.14
		**	5	7.62	7.70	7.70	1.35E-11	0.05	0.08
		**	6	25.56	36.65	36.65	1.35E-06	0.03	0.14
		**	6	61.21	61.21	61.21	7.85E-10	0.08	0.06
		**	6	85.21	85.21	85.21	2.06E-08	0.07	0.07
		**	6	123.04	126.85	123.04	2.50E-08	0.07	0.07
		**	7	142.26	142.26	142.26	1.17E-07	-0.11	0.10
		**	9	130.35	130.35	130.35	3.63E-07	-0.04	0.14
		**	10	136.88	136.88	136.88	2.02E-07	0.05	0.06
-glycosyl Flavones	<i>Total flavone content</i>	**	1	24.59	53.46	32.15	1.32E-22	-25.70	0.16
		**	1	216.31	219.83	216.31	1.10E-12	-28.80	0.07
		**	1	290.23	297.07	297.07	2.22E-18	26.90	0.05
		**	2	21.07	21.07	21.07	9.12E-10	23.00	0.05
		**	3	31.08	34.73	31.08	6.59E-08	14.50	0.11
		**	3	160.28	160.28	160.28	1.42E-06	15.40	0.03
		**	4	52.19	93.54	84.78	4.72E-08	-18.50	0.07
		**	4	165.06	165.06	165.06	1.90E-06	-11.20	0.08
		**	4	192.69	194.32	192.69	3.23E-08	23.30	0.05
		**	5	181.88	184.86	184.86	5.84E-07	-8.69	0.48
		**	6	6.13	16.65	6.27	1.63E-07	17.00	0.05
		**	6	16.65	41.68	22.36	1.23E-12	27.20	0.05
		**	6	80.48	80.71	80.48	2.12E-08	-18.30	0.08
		**	7	121.75	121.75	121.75	3.07E-07	-11.70	0.17
		**	8	20.26	20.26	20.26	1.15E-07	14.00	0.14
		**	9	124.89	154.17	128.06	1.50E-13	20.70	0.09
	<i>Percent condensed forms</i>	**	1	19.59	53.46	39.96	1.07E-30	-30.77	0.23
		**	1	89.36	89.36	89.36	2.65E-06	-7.86	0.08
		**	1	216.31	219.83	219.83	1.29E-08	19.35	0.06
		**	1	293.75	296.60	295.59	2.75E-11	-20.28	0.05
		**	2	21.07	21.07	21.07	3.11E-07	18.13	0.05

C	<i>Flavone content with condensed forms as a covariate</i>	**	2	130.57	130.57	130.57	6.30E-09	26.33	0.05
		**	4	78.68	93.54	84.78	3.03E-10	-19.60	0.07
		**	4	117.60	144.58	134.03	5.72E-07	-11.74	0.12
		**	4	192.69	211.44	192.69	6.09E-09	23.44	0.05
		**	5	44.33	50.00	44.33	2.97E-07	-15.66	0.08
		**	5	178.38	189.15	183.59	3.21E-11	9.40	0.50
		**	6	22.36	22.36	22.36	7.79E-11	23.20	0.06
		**	7	121.75	121.75	121.75	3.38E-08	-11.53	0.17
		**	8	20.26	20.26	20.26	6.12E-10	15.51	0.13
		**	9	115.42	154.01	143.53	1.58E-15	20.74	0.07
	<i>Proportion C-hexosyl, C-pentosyl apigenin</i>		5	210.19	210.19	210.19	1.48E-06	0.04	0.20
		**	6	115.85	135.53	124.82	1.38E-12	-0.05	0.47
		**	7	102.41	102.41	102.41	6.73E-07	0.04	0.25
		**	7	138.23	146.65	144.83	7.91E-11	-0.07	0.10
			8	122.21	134.71	122.21	3.80E-06	0.05	0.08
		**	9	129.91	129.91	129.91	1.73E-07	0.06	0.11

* SNPs are significant at $\alpha_E=0.01$ except for those indicated by ** ($\alpha_E=0.05$), and signals with MAF <0.05 are colored in gray

Figure 3: Manhattan plots for percent cyanidin-derived anthocyanins (A) and the percent peonidin of all cyanidin and peonidin-derived anthocyanins (i.e. peonidin / (peonidin + cyanidin)) (B). Dotted lines indicate the experiment-wide significance thresholds using the Bonferroni correction. ($\alpha_E=0.05$ and $\alpha_E=0.01$)

Figure 4: Manhattan plots for anthocyanin decorations: the proportion of acylated anthocyanins relative to the total anthocyanin content (A) total condensed forms (B) and the proportion of condensed forms relative to the total anthocyanin content (C). Dotted lines indicate the experiment-wide significance thresholds using the Bonferroni correction. ($\alpha_E=0.05$ and $\alpha_E=0.01$)

Figure 5: Manhattan plots for total C-glycosyl flavones (A) and proportions of C-hexosyl,-C-pentosyl apigenin (B). Dotted lines indicate the experiment-wide significance thresholds using the Bonferroni correction. ($\alpha_E=0.05$ and $\alpha_E=0.01$)

Supplementary Table 1: Description of proteins used for methyltransferase phylogram (Supplementary Figure 6) and alignment (Supplementary Figure 7).

Species	Identifier	Genbank ID / MaizeGDB ID	Gene product action
<i>Arabidopsis thaliana</i>	At1g67980	NP_564916.2	Caffeoyl-CoA 3-O-methyltransferase
<i>Arabidopsis thaliana</i>	At4g26220	AAM64800.1	Caffeoyl-CoA O-methyltransferase-like
<i>Arabidopsis thaliana</i>	At4g34050	NP_195131.1	Caffeoyl coenzyme A O-methyltransferase
<i>Chrysplenium americanum</i>	CaFOMT	AAA80579.1	3' flavonoid O-methyltransferase
<i>Cyclamen persicum</i> <i>x Cyclamen purpurascens</i>	CkmOMT2	BAK74804.1	Anthocyanin O-methyltransferase
<i>Cyclamen persicum</i> <i>x Cyclamen purpurascens</i>	CkmOMT1	BAK74803.1	Caffeoyl CoA O-methyltransferase-like
<i>Doroceras hygrometricum</i>	DhAOMT	KZV17033.1	Anthocyanin O-methyltransferase
<i>Glycine max</i>	GmAOMT	NP_001242455.2	Anthocyanin 3'-O-methyltransferase
<i>Medicago sativa</i>	MsCCoAOMT	Q40313.1	trans-caffeoyl-CoA 3-O-methyltransferase
<i>Medicago sativa</i>	MsCAOMT	AAB46623.1	Caffeic acid 3-O-methyltransferase
<i>Mesembryanthemum crystallinum</i>	McPFOMT	AAN61072.1	O-methyltransferase (Caffeoyl CoA, Flavonols, Flavones, Caffeic acid)
<i>Nemophila menziesii</i>	NmAMT	BBA68558.1	anthocyanin methyltransferase
<i>Nicotiana tabacum</i>	NtCcoAOMT	AAC49913.1	Caffeoyl-coenzymeA O-methyltransferase
<i>Nicotiana tabacum</i>	NtCCoAOMT5	AAB80931.1	Caffeoyl-CoA 3-O-methyltransferase
<i>Nicotiana tabacum</i>	NtCAOMT	AAL91506.1	Caffeic acid O-methyltransferase II
<i>Oryza sativa</i>	Os06g016800	NP_001056910.1	Caffeoyl-CoA 3-O-methyltransferase (similar to)
<i>Oryza sativa</i>	OsROMT	ABB90678.1	Flavonoid 3'-O-methyltransferase
<i>Petroselinum crispum</i>	PcCoAMT	AAA33851.1	Caffeoyl-CoA 3-O-methyltransferase
<i>Petunia integrifolia</i> <i>subsp inflata</i>	PiAMT	AIE77050.1	Anthocyanin methyltransferase
<i>Petunia x hybrida</i>	PxhCCoMT	ALP75648.1	Caffeoyl CoA O-methyltransferase
<i>Petunia x hybrida</i>	PxhAMT	AIE77047.1	Anthocyanin methyltransferase
<i>Populus trichocarpa</i>	PtCCoAOMT	CAA12200.1	Caffeoyl-CoA 3-O-methyltransferase
<i>Populus trichocarpa</i>	PtCCoAOMT1	ACC63876.1	trans-Caffeoyl- CoA 3-O-methyltransferase
<i>Solanum lycopersicum</i>	SIAOMT	AIS23587.1	Anthocyanin O-methyltransferase
<i>Solanum tuberosum</i>	StAOMT	AGC31679.1	Anthocyanin O-methyltransferase
<i>Solanum tuberosum</i>	StCCoAOMT	BAC23054.1	Caffeoyl-CoA O-methyltransferase
<i>Stellaria longipes</i>	SICCoAMT	AAB61680.1	Caffeoyl CoA, Flavonol, Flavone, Caffeic esters
<i>Triticum aestivum</i>	TaCCoOMT	AXM42901.1	Caffeoyl CoA O-methyltransferase
<i>Vitis vinifera</i>	VvCCoAMT	CAA90969.1	Caffeoyl CoA O-methyltransferase
<i>Vitis vinifera</i>	VvAOMT1	ACO52469.1	Anthocyanin O-methyltransferase
<i>Vitis vinifera</i>	VvAOMT3	CBI27641.3	Anthocyanin O-methyltransferase
<i>Vitis vinifera</i>	VvCAOMT	AAF44672.1	Caffeic acid O-methyltransferase
<i>Zea mays</i>		Zm00001d052683	

Supp. Table 1 (cont.)

<i>Zea mays</i>		Zm00001d052684	
<i>Zea mays</i>		Zm00001d052842	
<i>Zea mays</i>		Zm00001d052843	
<i>Zea mays</i>	ZmOMT1	Zm00001d036293	Caffeoyl CoA O-methyltransferase
<i>Zea mays</i>	ZmOMT2	Zm00001d045206	Caffeoyl CoA O-methyltransferase
<i>Zea mays</i>	ZmOMT3	Zm00001d024596	Caffeoyl CoA O-methyltransferase
<i>Zea mays</i>	ZmOMT4	Zm00001d052841	Caffeoyl CoA O-methyltransferase
<i>Zinnia violaceae</i>	ZvCcoAMT	AAA59389.1	trans-Caffeoyl- CoA 3-O-methyltransferase

Supplementary Figure 1: Anthocyanin glycoside structures.

Supplementary Figure 2: Schematic for the creation of Apache Red lines.

Supplementary Figure 3: Correlations between anthocyanin composition factors and wavelength at maximum absorbance (λ_{Max}), a proxy for hue.

Supplementary Figure 4: LD decay using r^2 with either a 10 site window (A) or 50 site window (B)

Supplementary Figure 5: Individual chromosome plots for major signals identified for the proportion of cyanidin (A), peonidin (B), and acylated anthocyanins (C). Chromosome numbers are listed in the x-axis label for each plot.

Supplementary Figure 6: Phylogram of plant O-methyltransferases. Amino acid sequences were aligned and maximum likelihood tree was created. Branch numbers represent bootstrap values based on 100 replicates. Sequences correspond to table 4, and Zea mays sequences are highlighted in blue.

Supplementary Figure 8: Manhattan plots for condensed forms using a flavone covariate (A) and flavones using a condensed forms covariate (B). Dotted lines indicate the experiment-wide significance thresholds using the Bonferroni correction. ($\alpha_E=0.05$ and $\alpha_E=0.01$)

Supplementary Figure 9: Individual chromosome plots for major signals identified for condensed forms (A-C), condensed forms with a flavone covariate (D-E), and the proportion of condensed forms compared to all anthocyanins (F-J). Chromosome numbers are listed in the x-axis label for each plot.

Supplementary Figure 10: Maximum likelihood phylogeny based on peptide sequences of LAR and ANR sequences with possible LAR and ANR candidates. Branch numbers represent bootstrap values based on 100 replicates. GenBank proteins are listed with each LAR/ANR – *At*, *Arabidopsis thaliana*; *Mt*, *Medicago truncatula*; *Md*, *Malus domestica*; *Vv*, *Vitis vinifera*; *Mn*, *Morus notabilis*; *Lc*, *Lotus corniculatus*; *Tc*, *Theobroma cacao*; *Hv*, *Hordeum vulgare*.

Supplementary Figure 11: Pairwise LD for chromosome 1 from 33 to 50 Mb. SNPs labeled in green were highly significant SNPs from AR GWAS and the SNP labeled in dark purple is the SNP closest to *PI*.

Supplementary Figure 12: Individual chromosome plots for major signals identified for flavone content (A-D), and the proportion of C-hexosyl- C-pentosyl apigenin (E-G). Chromosome numbers are listed in the x-axis label for each plot.