

## GWAS AND GENOMIC SELECTION FOR INCREASED ANTHOCYANIN CONTENT IN PURPLE CORN

Laura A. Chatham<sup>1</sup> and John A. Juvik<sup>1</sup>

<sup>1</sup>Department of Crop Sciences, University of Illinois at Urbana Champaign

### Abstract

Purple corn offers an attractive source of economical natural anthocyanin-based colorant for use in foods and beverages. Yet to maximize the scalability and meet growing demands, both anthocyanin concentrations and agronomic performance must improve in purple corn varieties. We studied flux through the flavonoid biosynthesis pathway using GWAS data derived from a diverse purple corn landrace with anthocyanin-rich pericarp, Apache Red. Trends between flavonoid endpoints suggest that regulators of total flux into the pathway and regulators of partitioning within the pathway may both represent targets for maximizing anthocyanin content. A peak at the end of chromosome 1 near *Aat1* (*Anthocyanin acyltransferase1*) was highly significant in all approaches taken to map anthocyanin flux, suggesting the structural modification of malonylation is required for maximal anthocyanin production. We also identified several candidate MATEs and H<sup>+</sup> ATPases that could assist in the preferential transport of acylated anthocyanins into the vacuole. These and other candidates identified here suggest there is still much to learn about the regulation of anthocyanin biosynthesis in the pericarp of purple corn. The efficacy of genomic predictions in the population was also studied, yielding an accuracy of 0.71 with cross validation for total anthocyanin content with no improvement found when known anthocyanin regulators were added to the model. These data suggest that genomic selection could be employed effectively in a purple corn breeding program, and especially for a landrace improvement program.

Keywords: Anthocyanins, flavonoid biosynthesis pathway, flux, maize pericarp, purple corn, genotyping-by-sequencing, genomic selection

## Introduction

With continued trends in consumer preference for less processed foods, rising demand for naturally pigmented food and beverages necessitates the development of economical natural colorant sources. Purple corn is a good source of anthocyanins, the water soluble, red to purple pigments found in many types of fruits and vegetables. As a scalable commodity, corn could provide a natural colorant source capable of meeting growing consumer demands. However, widescale production will require breeding to create lines with maximum anthocyanin production while maintaining yield and other agronomic traits.

Various strategies can be taken in breeding for increasing the economic feasibility of using purple corn to replace artificial colorants. Ultimately the goal is to increase anthocyanin yield per unit input and ideally both per-kernel anthocyanin content and per-acre yield could be increased to compound the total anthocyanin yield per acre. Accomplishing this will require identifying the highest anthocyanin producing purple corn lines and backcrossing into current elite inbreds. With each round of backcrossing, selections must be made to minimize linkage drag while avoiding the loss of any factors associated with maximal anthocyanin production. Anthocyanin biosynthesis involves a number of different structural genes and regulatory mechanisms, all of which have the potential to become unfixated alleles upon backcrossing with non-pigmented lines. Loss of any of these factors could influence anthocyanin yield to varying degrees.

Nearly all of the anthocyanin structural genes have been identified with the exception of those associated with anthocyanin decorations post-synthesis<sup>1</sup>. While the major regulatory players have been identified, current models in maize are simplistic compared to those in other anthocyanin producing species<sup>2</sup>. The anthocyanin MBW regulatory complex consists of an R2R3-MYB protein, a bHLH (basic helix-loop-helix) protein, and WDR (WD-repeat) domain protein<sup>3</sup>. Typically, *Colored aleurone1* (*C1*, MYB), *Colored1* (*R1*, bHLH), and *Pale aleurone color1* (*Pac1*, WDR) together create the regulatory complex and activate structural genes in the kernel aleurone. A bHLH-like recessive intensifier *in1* (*intensifier1*) with similarity to *R1*, increases anthocyanin pigmentation in aleurone<sup>4</sup>. More recently an EMSY-like factor, *RIF1* (*R interacting factor1*) was found to act as a switch to help control activation of structural genes<sup>5</sup>. An analogous MBW complex likely forms to regulate anthocyanin biosynthesis in plant tissue and kernel pericarp. *Booster1* (*B1*) and *Plant color1* (*Pl1*) are the bHLH and MYB regulatory

factors, respectively, most often associated with regulation in plant tissues. However, there are exceptions in which certain alleles of *R1* (e.g. *R1-ch*) operate in pericarp and certain *B1* alleles (e.g. *B1-Peru*) operate in aleurone<sup>6</sup>. Whether the WDR, *RIF1*-like, or *INI*-like proteins function in pericarp has yet to be determined, and no homologous pericarp versions of these have been identified either. Furthermore, R2R3-MYB repressors of anthocyanin biosynthesis have been identified in a number of other species, yet none have been identified in maize<sup>7</sup>. MYB repressors associated with the phenylpropanoid pathway and downstream lignin biosynthesis have been identified<sup>8</sup>.

This regulation of the structural genes occurs in the nucleus and the structural gene products function together in a flavonoid metabolon at the cytoplasmic face of the endoplasmic reticulum<sup>9</sup>. Anthocyanins are then transported to the vacuole for storage. In maize, *Bz2* (*Bronze2*) encodes a glutathione S-transferase which assists in anthocyanin transport and is required to prevent anthocyanin oxidation, resulting in a bronze colored kernel<sup>10</sup>. *Mrp3* (*Multidrug resistance-associated protein3*) produces an ABC transporter that shuttles anthocyanins into the vacuole<sup>11</sup>. Both were identified and studied in aleurone-pigmented lines. In other anthocyanin-producing species two other mechanisms of vacuolar localization have been identified in addition to the GST/ABC transporter system – MATE-transporters and vesicle trafficking. MATE-transporters, either as H<sup>+</sup>/flavonoid antiporters or in combination with other vacuolar proton gradient maintaining mechanisms have been identified in *Arabidopsis*, *Medicago*, tomato (*S. lycopersicum*) and radish (*R. sativa*)<sup>12-17</sup>. Support for vesicle trafficking of anthocyanins and a subsequent autophagy-like mechanism of vacuolar transport has been found in *Arabidopsis*, *Brassica napus*, and grapevine (*V. vinifera*). Some evidence for this mechanism exists in maize. Anthocyanin-containing vesicles and anthocyanin vacuolar inclusions (AVIs) were shown to fuse in response to light<sup>18</sup>

In addition to anthocyanins, several other flavonoids are present in maize. Phlobaphenes, orange to brick red pigments accumulating in cob and pericarp, are polymers of the 3-deoxyflavonoids, luteoforol or apiforol, and share the early steps of their biosynthesis and thus carbon flux with anthocyanins. The structural genes responsible for phlobaphene biosynthesis are controlled by *Pericarp color1* (*PI*), an R2R3 MYB similar to *CI/PII*. In contrast to the anthocyanin regulating MYBs (*CI/PII*), *PI* functions independently, not requiring a bHLH partner for regulatory activity. *PI* also regulates C-glycosyl flavone biosynthesis, a group of

compounds that includes maysin, an insecticidal compound conferring resistance to corn earworm (*Helicoverpa zea*)<sup>19</sup>.

With multiple endpoints, branch points and shared precursors, flux through the flavonoid pathway in maize is complex. Breeding for maximal anthocyanin content is likely to benefit greatly from a better understanding of the mechanisms regulating this flux. Furthermore, the identification of additional factors associated with anthocyanin content may help prevent their loss during backcrossing and enable marker assisted selection to expedite the process. Here we explore flux through the flavonoid pathway in a landrace with variability in anthocyanin, flavone, and flavonol content and candidate genes underlying this variability. Given the complexity and quantitative nature of the trait, we also gauge the accuracy of genomic selection with and without major regulatory players.

## Methods

### *Plant materials*

Apache Red seeds (S<sub>0</sub>) were purchased from Siskiyou Seeds (Williams, OR) and lines were created as described and illustrated in this article's companion article (Chatham and Juvik, submitted to G3). Briefly, each round of selection involved phenotyping (HPLC) and selecting lines with anthocyanin content only and with the most variability in anthocyanin composition. This produced 181 S<sub>2</sub> lines that were used for a pilot round of genotyping and phenotyping. After further breeding and selection, 1148 S<sub>2</sub>, S<sub>3</sub>, and S<sub>4</sub> lines were produced that were used for genotyping and phenotyping. Selected S<sub>2</sub>, S<sub>3</sub>, and S<sub>4</sub> seeds were sown the following season and bulked by plot to produce seed for use in other collaborative projects. Kernels from S<sub>3</sub>, S<sub>4</sub>, and S<sub>5</sub> plots were phenotyped as described below.

### *Extractions, HPLC analysis, and UV-visible spectroscopy*

All anthocyanin/flavone phenotyping, including extraction, HPLC analysis, and UV-visible spectroscopy were performed as described previously (in this article's companion). This reference also describes seedling germination, growth, and sampling for DNA extraction. Variables were transformed as needed for comparisons.

### *Genotyping and SNP discovery*



GBS libraries were prepared as described previously and sequencing was performed by the Roy J. Carver Biotechnology Center at the University of Illinois in Urbana-Champaign (Chatham and Juvik, submitted to G3). The 1148 lines (S<sub>2</sub>, S<sub>3</sub>, and S<sub>4</sub> selfed progeny) were sequenced on three lanes of Illumina HiSeq 4000 with single-end 100 nucleotide reads and multiplexed to obtain approximately equal numbers of reads on each lane. SNP discovery was carried out using the TASSEL 5.0 GBS v2 pipeline<sup>20</sup> as described previously (Chatham and Juvik, submitted to G3)

#### *Association mapping and genomic selection*

The R GAPIT (Genome Association and Prediction Integrated Tool) package<sup>21</sup> was used for association mapping, and the RRBlup package<sup>22</sup> was used for genomic selection. GWAS was also run individually for each chromosome using a leave-one-out method of calculating kinship to retain maximal power for detecting significant associations<sup>23</sup>. For genomic selection, prediction accuracy averages were measured from 25 replicates of five-fold cross validation.

## **Results and Discussion**

### *Apache red population*

Phenotypes were generated at each stage of the Apache Red (AR) population development and used to calculate heritability from parent offspring correlations<sup>24</sup>. Anthocyanin content heritability ranged from 0.44 to 0.57, suggesting improvement can be made through selection. Figure 1 shows a summary of kernel colors obtained from the AR population. AR S<sub>0</sub> seeds were all dark purple, indicative of anthocyanins in pericarp, but S<sub>1</sub>s varied in color. Most were dark purple, anthocyanin-containing, but we observed a range of water insoluble phlobaphene-containing kernels colored from orange to purple-red and some kernels that were colorless or had minimal anthocyanin or phlobaphene pigmentation. While selection for highly anthocyanin-pigmented kernels appears to have been successful given the decrease in the proportion of colorless or phlobaphene containing kernels across generations of self-pollination, the presence of colorless and phlobaphene containing kernels by the S<sub>4</sub> (S<sub>3</sub> pericarp) generation suggests that anthocyanin content is controlled by multiple segregating loci in this population. An alternative explanation is that some of the segregating alleles could be paramutable, spontaneously converting to paramutagenic alleles, which then heritably silence other

paramutable alleles<sup>25</sup>. Even in the most advanced lines from the last generations grown (S<sub>3s</sub>, S<sub>4s</sub>, and S<sub>5s</sub>), which were intended to bulk seed for other collaborative projects on anthocyanin-rich maize, colorless and phlobaphene containing ears were still present and had to be discarded at the time of harvest.

### *Phenotyping flux*

To understand the factors influencing total anthocyanin content we compared the relationship between proportions of flavonoid end products and total anthocyanin content. Several simplistic models of flux can be considered. Under a finite flux, zero-sum hypothesis, one would expect anthocyanin content to decline as proportions of other flavonoids increase. Conversely if pathways are regulated independently or are regulated together by some universal regulator, one would expect to see no correlation or positive correlations between anthocyanin content and proportions of other flavonoids, respectively.

Flavone and anthocyanin content were positively correlated ( $r = 0.28$ ,  $p < 2.2e-16$ ) (Figure 2A), suggesting a major regulatory factor functioning at a step prior to the branch point between anthocyanins and flavones. Conversely, the proportion of anthocyanins present as flavonol-anthocyanin condensed forms was negatively correlated with total anthocyanin content ( $r = -0.27$ ,  $p = 6.24e-16$ ) (Figure 2B), lending support to a competitive zero-sum-like model of flux for these forms. Correlations with specific anthocyanin types were also examined. Proportion of acylated anthocyanins was positively correlated with anthocyanin content ( $r = 0.23$ ,  $p = 1.51e-14$ ) (Figure 2C), and the proportion of cyanidin-derived anthocyanins was negatively associated with anthocyanin content ( $r = -0.31$ ,  $p < 2.2e-16$ ). These correlations could be an indication of substrate specificity in downstream processes or could be due to linkage or population structure. All lines here were derived from the AR landrace, making population structure an important consideration.

Flavonoid contents were also compared with maximal absorbance ( $Abs_{Max}$ ) in the ~495-525nm range, a value indicative of darkness and often used in a method for determining anthocyanin content (Supplementary Figure 1)<sup>26</sup>. Unsurprisingly, anthocyanin content was highly correlated with  $Abs_{Max}$  ( $r = 0.92$ ,  $p = 2.2e-16$ ). As with anthocyanin content, flavones were positively correlated with  $Abs_{Max}$ , but the magnitude of the correlation was lower ( $r = 0.43$ ,  $p < 2.2e-16$ ). This may be due to the absorbance increasing effect of flavone-anthocyanin

copigmentation<sup>27</sup>. In contrast, neither the proportion of condensed forms nor the proportion of acylated forms was significantly correlated with Abs<sub>Max</sub> ( $p = 0.15$ ,  $p = 0.07$ ). For condensed forms this loss of a negative correlation (compared to total anthocyanin content) could be attributed to color enhancing properties of condensed forms as has been seen previously in purple corn<sup>28</sup> (companion paper). The lack of a correlation between proportion of acylated anthocyanins and Abs<sub>Max</sub> despite the significant relationship seen with anthocyanin content is more difficult to explain. Acylation has not been associated with a significant increase in absorbance. However, this could be an indirect effect of the correlation that exists between acylated anthocyanins and flavones in this population (Chatham and Juvik, submitted to G3). Generally, high flavone concentration was observed in lines with smaller proportions of acylated anthocyanins. These extracts could have lower acylation and lower overall anthocyanin content, but not necessarily have a correspondingly low Abs<sub>Max</sub> due to the presence of flavones and the effects of copigmentation.

From a global model using all factors described above and in Supplementary Figure 1, the most parsimonious model was chosen. In this model anthocyanin content, condensed forms, flavones, and the interactions between them were all significant, but anthocyanins and condensed forms had the smallest  $p$  values and largest effect sizes by far. Variables were scaled so that effect sizes can be directly compared. The anthocyanin term had a scaled effect size of 1.5 standard deviations from the mean and a  $p$ -value too low to estimate in R. The condensed form factor had a scaled effect size of 0.5 ( $p = 2.7e-77$ ) and the interaction between anthocyanin and condensed forms had an effect size of 0.6 ( $p = 7.3e-59$ ). Taken together, this suggests a complex model of flux, in which each of these factors should be considered in the context of the others. Interactions between factors could directly influence Abs<sub>Max</sub> as is the case for copigmentation<sup>27</sup>, or the interactions could be indications of the underlying biology, for example, a regulator controlling both anthocyanins and flavones.

Figure 3 shows schematics that describe two approaches to the regulation of pathway flux. With Figure 3A as a starting point illustrating how total carbon flux into the pathway is portioned between phlobaphenes, flavones, anthocyanins, and condensed forms, 3B illustrates a finite flux approach in which losses of some products are realized as gains in others (zero-sum model). Figure 3C illustrates an approach in which flux is regulated universally, where gains in one class of compounds result in gains in another (win-win scenario). The ability to breed for

both scenarios simultaneously would be ideal (Figure 3D), but further manipulation of the pathway in this way will require a better understanding of genetic factors controlling flux. Flux likely has attributes of both models to varying degrees and at different places in the pathway, along with a number of other factors including environmental conditions.

### *Mapping anthocyanin concentration and Abs<sub>Max</sub>*

In mapping anthocyanin content and other flavonoid products and the subsequent selection of candidate genes, known structural and regulatory factors associated with the pathway were considered first (Table 1). A list of potential candidates based on anthocyanin and flavonoid regulatory and trafficking genes in other species was also prepared by blasting against the maize genome (Supplementary Table 1)<sup>29</sup>. Mapping anthocyanin content was performed using both Abs<sub>Max</sub> and total anthocyanin content (Figure 4). Given the differences observed between their correlations with different flavonoid profiles, we expected slightly different mapping results. Each produced highly significant signals on Chr 1 (Chromosome 1) near *Aat1*, with the most significant SNP less than 1 Mb away. This suggests that despite being a structural gene, acylation by *Aat1* is important for achieving highly pigmented extracts. In *V. vinifera*, MATE transporters were identified that selectively transported acylated anthocyanins into the vacuole compared to their un-acylated versions<sup>14</sup>. While this offers a plausible explanation for a highly significant peak near *Aat1*, no anthocyanin-transporting MATEs have been identified in maize and this phenomenon has not been associated with the GST/ABC transporter system (*Bz2/Mrp3* in maize).

Loci identified for Abs<sub>Max</sub> but not for total anthocyanin content are of interest because they likely represent factors associated with increasing the darkness of extracts that are unrelated to anthocyanin content. The most plausible reason for the occurrence of these loci is involvement in flavone production and thus copigmentation with anthocyanins<sup>27</sup>. For Abs<sub>Max</sub> a locus on Chr 1 spanning 32-40 Mb was detected and suspected to be *PI*, a major regulator of flavone biosynthesis. This signal was only observed for Abs<sub>Max</sub> and not total anthocyanin content, which could be explained by the color intensifying effect of flavones in the presence of anthocyanins<sup>27</sup>. As a proxy for color intensity, Abs<sub>Max</sub> would be expected to tag major loci associated with flavone content. Targets of *PI* were thus included as candidates in Supplementary Table 1<sup>30</sup>. However, the range observed is at least 8 Mb away from *PI*, located at 48 Mb. Another potential

candidate considered is a gene encoding the  $\alpha 3$  subunit of a vacuolar proton ATPase (Zm00001d028436, 34.96 Mb). If MATE transporters were found to play a role in the trafficking of anthocyanins to the vacuole, as has been observed in other anthocyanin producing species, a proton pump would likely be necessary to generate the gradient required for MATE function<sup>2</sup>. Individual chromosome plots showing signals and candidate gene locations can be found in Supplementary Figure 2.

Other loci observed for  $Ab_{S_{Max}}$  but not total anthocyanin content were found on Chr 3 around 14.8 Mb and Chr 9 at 11.8 Mb (Table 2). While no candidates were identified for the former, the latter is close to *Bz1* (*Bronze1*), a UDP-glucose flavonol glycosyltransferase at 11.2 Mb. Surprisingly, this signal was significant only for  $Ab_{S_{Max}}$  but not for total anthocyanin content.

Examining the percent of acylation in Figure 2B and Supplementary Figure 1D shows a bimodal distribution, with the lower acylation group having both fewer observations and a smaller spread of anthocyanin content and  $Ab_{S_{Max}}$ . This association could be due to linkage between *Aat1* and some other factor related to anthocyanin content, but LD in the surrounding region suggests this may be unlikely. Nonetheless, to account for potential linkage, mapping was performed using acylation as a covariate, and modeling *Aat1* as a fixed factor, but the signal was still present. To overcome this, the tails of the acylation proportion distribution—less than 0.35 and greater than 0.75—were removed from the data to isolate anthocyanin content.

Using the truncated data set for mapping (acyl proportion tails removed, leaving 929 lines) eliminated the signal near *Aat1* as well as several others. Examining these eliminated loci could reveal factors associated uniquely with acylation. In addition to Chr 1 at 306.5 Mb, signals on Chr 2 at 150.4 Mb and on Chr 5 at 8.6 Mb were lost when the data was truncated, and mapping was rerun. The closest candidates identified for these two signals were a potential MATE transporter located on Chr 2 at 154.1 Mb (Zm00001d005018) and a UDP-D-glucose dehydrogenase on Chr 5 at 6.95 Mb involved in sugar metabolism and expressed in pericarps of both *PI-rr* (phlobaphene rich) and *PI-ww* (no phlobaphene) lines<sup>30</sup>. Another potential candidate is *Chi3* (*Chalcone isomerase3*), located at 2.6 Mb. The proximity of these signals to candidate loci can be seen in Supplementary Figure 2.

Two signals were consistently significant for all analyses. The first, on Chr 1 at 219.8 Mb, did not have any immediately obvious candidates. A signal in this area was also identified

previously for several traits in this article's companion (Chatham and Juvik, submitted to G3). The significant SNP for this signal is about 4 Mb away from *myb83*, discussed previously, and is very close to a WRKY transcription factor located at 219.8 Mb (*wrky25*; Zm00001d032265). WRKY transcription factors are a more recent addition to the anthocyanin regulatory complex identified in both *Arabidopsis* (*TTG2*) and *Petunia* (*PH3*), and these likely interact with the WD repeat protein component of the MBW regulatory complex<sup>3</sup>. The candidate WRKY, *wrky25* is also highly expressed in pericarp tissue<sup>31</sup>; however, when *TTG2* and *PH3* were blasted against the maize genome, *wrky25* was a weak match. We also identified a *wrky* candidate (*wrky62*, Zm00001d035323) near a QTL on Chr 6 for anthocyanin content, but it too was a weak match for both *TTG2* and *PH3*.

With what is known regarding the regulation of anthocyanin content in pericarp, and previous reports breeding for purple corn<sup>32,33</sup>, we expected to find *BI* and *P11*. Instead, SNPs proximal to *RI* (less than 1 Mb away) were found in all four analyses. The *RI* locus is diverse, and while usually associated with aleurone color, some alleles of *RI* (e.g. *RI-ch*), and its neighboring (or allelic) loci *Sn1* (*Scutellar node1*)<sup>34</sup>, *Lc1* (*Leaf color1*)<sup>35</sup>, and *Hopi*<sup>36</sup> can produce anthocyanins in pericarp<sup>6</sup>. Additional research will be required to identify the specific bHLH allele in AR and whether it interacts with other proteins in the manner described previously for aleurone-pigmented lines.

*RI* is a bHLH transcription factor and part of the MBW ternary regulatory complex associated with anthocyanin biosynthesis across species<sup>3</sup>. Therefore, we expected to also find a MYB and a WD40 as well. While not as obvious as for *RI*, a peak can be seen near the plant color associated MYB, *P11* (Chr 6 at 113.0 Mb), in Figure 4, but no signal was found near *Pac1* (Chr 5 at 201.8 Mb), the WD repeat protein component of the MBW regulatory complex. *RI* has also been shown to interact with *Rif1* (*R-interacting factor1*), potentially to regulate control of early versus late biosynthetic steps in the anthocyanin pathway<sup>5</sup>. No signal was observed close to *Rif1* (Chr 7 at 76.6 Mb) either, though it is possible that some key components of the regulatory complex (e.g. *Rif1* and *Pac1*) are fixed in the Apache Red population, offering an explanation for these missing signals. Alternatively, pericarp-specific alleles of these could exist that have not yet been identified.

In addition to those described above, several signals were identified that were present in some but not all the described mapping approaches (Figure 4). On Chr 1, two significant SNPs

were found at 182 and 190 Mb, for which the closest candidate found was a set of potential ANR/LAR like proteins (Zm00001d031488, Zm00001d031489). On Chr 7 we identified a peak near *In1*, a recessive intensifier of anthocyanin biosynthesis in aleurone-pigmented lines (Supplementary Figure 2). To our knowledge, there have been no reports of *In1* functioning in pericarp. However given the capacity of *in1* to significantly increase anthocyanin content in aleurone<sup>37</sup>, this should be explored further as a potential strategy to maximize anthocyanin content in pericarp. Another peak on Chr 7 was found close to *myb162* (Supplementary Figure 2).

### Mapping molar flux

While the mapping variables used up to this point primarily correspond to the model of increased overall flux from Figure 3C, we also wanted to look for regulators that may fit the competitive flux model seen in Figure 3B. To do this, concentrations of individual anthocyanin and flavone compounds were converted from  $\mu\text{g/ml}$  to  $\text{mols/ml}$ , and the proportion of the total molar flux into the pathway was calculated and mapped for anthocyanins, flavones, and condensed forms (Figure 5). Given the relatively small proportions of condensed forms, anthocyanin and flavone proportion made up the majority of flux through the pathway for these calculations and thus mapping both anthocyanin and flavone proportions was redundant (e.g. flavone and anthocyanin proportions are tightly inversely correlated). Manhattan plots from flavone flux and anthocyanin flux were nearly identical and the same signals were observed for both. Nonetheless, these variables still represent partitioning between branches of the pathway, and candidates identified may play key roles in regulating this.

For all proportions mapped, signals close to *PI* and *Aat1* were identified. *PII*, the bHLH regulatory factor identified previously, was conspicuous both for proportions of flavone and anthocyanin flux. While *RI* showed up consistently when mapping anthocyanin content overall, it was barely detectable when looking at proportions of flux through the pathway. This suggests *PII* may play a more significant role in the competition for flux through the pathway, influencing the ratio of flavones and anthocyanins, while *RI* may contribute more to overall flux.

Similar to results of mapping absorbance, the peak at the beginning of Chr 1 was close to *PI* but not nearly as close as other candidates have been to their assumed loci (e.g. *Pr1*, *RI*). The most significant SNP was more than 10 Mb away from *PI*. This could be explained if large



linkage blocks in this area were detected. Pairwise recombination frequency was calculated for the region on Chr 1 between 33 and 50 Mb (Supplementary Figure 3). *PI* is located at 48 Mb, but the most significant SNP for anthocyanin/flavone flux was found at 35 Mb. Higher linkage in these areas could contribute to the gap between the most significant SNPs and *PI* (Supplementary Figure 4). The frequency with which we found *PI* for various traits, both here and in this article's companion, is somewhat puzzling given the degree of selective pressure for anthocyanin content during the development of the AR population (Figure 1). One explanation may be the presence of multiple loci located in a relatively small area, making separation difficult. When this region was searched, a candidate [Zm00001d028436 (vacuolar proton ATPase a3)], at 35 Mb was identified. This or other additional candidates could explain the frequent detection of the peak on Chr 1 in phenotypes not expected to be influenced by *PI*. However, it cannot be determined without further research whether the positive correlation between flavones and anthocyanins is a cause (leading to detection of *PI* in anthocyanin mapping) or an effect (correlation resulting from linked anthocyanin and flavone loci) of the significant peak observed on Chr 1.

We also identified a signal near *Pr1* when mapping the proportion of anthocyanin and flavone content (Figure 5). In general, we observed that *pr1/pr1* lines typically had higher flavone content compared to *Pr1/-* lines. Moreover, the only flavone compounds detected here in abundance were apigenin derived (monohydroxylated B-ring), consistent with their presence in *pr1/pr1* lines containing primarily pelargonidin (also monohydroxylated B-ring). Therefore, we suspect the occurrence of the signal on Chr 5 is an artifact of population structure. Yet interestingly, studying *pr1* in the context of C-glycosyl flavone production showed that recessive *pr1* increased apimaysin content (monohydroxylated) but *Pr1* had little to no effect on maysin content (dehydroxylated B-ring) suggesting the presence of divergent pathways and the expression of an additional F3'H<sup>38</sup>. This might explain the lack of luteolin-derived flavones (dehydroxylated B-ring), even in the presence of *Pr1* (cyanidin-dominant lines). Regarding anthocyanin content, a structural gene like *Pr1* seems unlikely to influence total anthocyanin content yet we have noticed that pelargonidin lines generally have lower ceilings of anthocyanin production in comparison with cyanidin-dominant lines in the purple corn germplasm<sup>39</sup>. Additional research would be required to determine whether there is a true biological explanation, such as substrate specificity, or if this is an artifact of population structure.

Several other significant SNPs or SNP ranges were observed (Table 2). On Chr 7, a peak was found at 156.4 Mb, which aligns closely with a potential ortholog of the *Arabidopsis TT12* (*Transparent Testa12*) MATE transporter of flavonoids (Zm00001d021629, 158.6 Mb)<sup>12</sup> and other flavonoid transporting MATEs in other species (Supplementary Figure 3)<sup>14,15,17</sup>. This is particularly interesting given the highly significant effect of SNPs near *Aat1* for both competitive flux and overall flux (Figures 4 and 5). *VvAM1*, *VvAM3*, and *MtMATE2* all preferentially transport acylated anthocyanins<sup>14,15</sup>, while *MtMATE1* is specific to epicatechin<sup>15,40</sup>. A candidate MATE with specificity for acylated anthocyanins could explain how *Aat1* appears to contribute so significantly to anthocyanin content. A phylogeny of candidate MATEs and MATEs associated with flavonoid transport in other species can be found in Supplementary Figure 5. Differences in specificity could also play a role in the relative amounts of condensed forms versus anthocyanins that are stored in the vacuole. We also found a single significant SNP on Chr 9 (23.9 Mb), for which the closest candidate we found is another MYB transcription factor (*myb148*, 26.7 Mb; Supplementary Figure 2). While this QTN was highly significant, minor allele frequency was low (0.03) raising concern that this could be a false positive result.

While in this article's companion, total condensed forms and the proportion of condensed anthocyanins were mapped, here we mapped the molar proportion of condensed forms synthesized compared to total pathway flux. The signal on Chr 9 (130.67 Mb) is suspected to correspond with *Fns1* (Chr 9 at 30.7 Mb). While *Fns1* is responsible for the formation of flavones, it could participate in the competition of flux toward flavones versus anthocyanins and flavan-3-ols, thereby influencing condensed forms indirectly. *Fns1* is regulated by both *PI* and the MYB (*P11/C1*) and bHLH (*BI/RI*) system, which could contribute to the frequent co-occurrence of anthocyanins, flavones, and condensed forms in Apache Red extracts<sup>30,41</sup>.

A signal on Chr 3 at 174.4 Mb was also identified 1.3 Mb from a candidate ANR (anthocyanidin reductase) (Zm00001d042594). In this article's companion paper, several potential ANR/ LAR candidates were proposed, but no signals were produced close to this particular candidate. Candidates identified previously clustered with LARs while this candidate groups with ANRs. While LARs have only been shown to produce 2,3-*trans* flavan-3-ols (e.g. catechin) and ANR shown primarily to produce 2,3-*cis* flavan-3-ols (e.g. epicatechin), epimerase activity converting 2,3-*cis* flavan-3-ols to 2,3-*trans* flavan-3-ols has been found in several species, suggesting an ANR may be sufficient for the production of both flavan-3-ol types<sup>42-44</sup>.

However, LARs have also been implicated in proanthocyanidin biosynthesis by regulating the number of epicatechin starter and extension units used to create polymers<sup>45</sup>. This has not been established for the formation of condensed forms, but a similarly functioning LAR could play a role in the dimerization of flavan-3-ols and anthocyanins to make condensed forms<sup>2</sup>.

Mapping flux toward condensed forms produced several signals that had not been identified in previous analyses, including a hit on Chr 4 at 176.6 Mb. A UDP-D-glucuronic acid 4-epimerase (Zm00001d052053) regulated by *PI*<sup>30</sup> was found 1.2 Mb away, and putative MATE transporter (Zm00001d051886) with similarity to *Vv*AMT anthocyanin-transporting MATEs was found 3 Mb away from the peak on Chr 4. However the closest candidate (<0.2 Mb) is an ATPase (Zm00001d052022) similar to TT13 from *Arabidopsis*, a P-type H<sup>+</sup> ATPase required for epicatechin transport into the vacuole and proanthocyanin accumulation in seeds. TT13 provides the H<sup>+</sup> gradient required to fuel TT12, a MATE transporter of epicatechin in *Arabidopsis*<sup>46</sup>. This candidate also has similarity to *PH5* from petunia, an H<sup>+</sup> P-ATPase that controls vacuolar pH and flower color<sup>47</sup>.

We also observed several signals that had been previously detected when mapping other traits. This includes a peak at the beginning of Chr 3 near *tcptf33*, a blast hit for *TCP3* from *Arabidopsis* which interacts with the MBW regulatory complex in the control of flavonoid biosynthesis<sup>48</sup>. This same signal was observed when mapping total condensed forms in this article's companion paper (Chatham and Juvik, submitted to G3). Similarly, signals on Chr 3 around 52 Mb and on Chr 7 at 142 Mb were again detected. Potential candidates for these include *myb40* (Chr 3, 54.36 Mb) and a potential LAR/ANR ortholog (Zm00001d020970, Chr 7, 138.21 Mb) or glucosyltransferases (Zm00001d021167, Zm00001d021168, Chr 7, 144.7 Mb). Several potential MYBs have been identified here and in this article's companion. A phylogeny of known anthocyanin activators and repressors from other species and maize candidates can be found in Supplementary Figure 6.

While a single significant SNP at the beginning of Chr 9 was found for anthocyanin proportions, a large peak with a string of significant SNPs was found here for proportions of condensed forms, with several trailing significant SNPs extending toward the centromeric region (Supplementary Figure 3). While the most significant SNP for this peak was located at 16.6 Mb, the width of the region makes identifying candidates difficult. Previously, this same region at the beginning of Chr 9 has been identified in association with flavonoid content. *Bp1* (*Brown*

*pericarp1*) is located in bin 9.02 (11.8 – 23.3 Mb in B73 RefGen\_v3)) and is a locus associated with modifying red pericarp to brown when recessive and in the presence of *PI*<sup>49</sup>. It was therefore suggested to play a role in the formation of phlobaphenes<sup>50</sup>. This same area was later associated with maysin biosynthesis and corn earworm antibiosis (*recessive enhancer of maysin1* (*rem1*)) and was mapped between *Bz1* (11.2 Mb) and *Wx1* (23.1 Mb)<sup>38</sup>. In addition to phlobaphenes and C-glycosyl flavones, we have now found this region to be important for the proportion of flavonol-anthocyanin condensed forms produced. Interestingly, this QTL region was not nearly as evident for the proportions of anthocyanin / flavone content, which based on previous reports would have been expected. Nor did we identify this same region for total C-glycosyl flavone content in this article's companion. Several scenarios could account for the significance of this region in multiple traits. Proximally located genes with different functions and their resulting linkage could help explain the correlation between flavones and condensed forms. Alternatively, the *rem1* / *bp1* region might contain a single gene whose product acts at a point in the phenylpropanoid pathway prior to differentiation between phlobaphene, flavone, and flavanol branches. Similarly a single gene in this region could act as a regulator at multiple points in the pathway.

Given this region's involvement in phlobaphene, flavone, and condensed form content, several potential candidates were considered. A plasma membrane ATPase (Zm00001d045122, 13.45 Mb), a glucosyltransferase labeled as an anthocyanidin 5,3-O glucosyltransferase (*ufgt3*, Zm00001d045254, 17.2 Mb) and *ZmMrp3* (17.51 Mb) the ABC transporter involved in anthocyanin transport<sup>11</sup>. Locations of these are shown in Supplementary Figure 3. Condensed forms and C-glycosyl flavones are sequestered in the vacuole, while phlobaphene polymers are destined for the cell wall. If a MATE transporter were required to transport flavones and flavan-3-ols into the vacuole (as in *Arabidopsis*), and to move phlobaphene precursors across the cell membrane, a generalized ATPase providing the necessary proton gradient could play a role in the transport of all three classes of compounds. Failure to transport the necessary precursors could result in compound degradation leading to brown kernels lacking phlobaphenes and a decrease in maysin content. Similarly, *Mrp3*, while associated with anthocyanin transport into the vacuole, could function similarly for other compounds. In this case, *mrp3* could result in degraded phlobaphene precursors, flavones, and flavan-3-ols.

The glucosyltransferase candidate could also be versatile enough to play a role in the accumulation of flavones, flavan-3-ols, and flavan-4-ols. Anthocyanidins must be glycosylated for stabilization and transfer to the vacuole, and thus failure of the anthocyanin glycosyltransferase *Bz1* in aleurone results in the degradation of anthocyanidins and bronze kernels instead of blue, purple, or pink kernels (depending on the anthocyanin makeup)<sup>1</sup>. While little is known about the transport and storage of C-glycosyl flavones or the transport, polymerization and storage of phlobaphenes, an analogous scenario is plausible in which flavan-4-ols (phlobaphene precursors), flavan-3-ols, and flavones are degraded before reaching their storage destination. More evidence for this hypothesis comes from the phenotype described for the *Bp1* locus (red kernels become brown in the presence of *P1* and *bp1/bp1*), which is akin to the *bz1* mutant phenotype (blue/pink/purple kernels to bronze in the presence of *bz1/bz1*). However, flavones are C-glycosylated, flavan-3-ol components of condensed forms are not known to be glycosylated at all before condensation with an anthocyanin, and the pathway for phlobaphenes is unknown after the formation of flavan-4-ols. Furthermore, a bifunctional C-/O-glycosyltransferase associated with C-glycosyl flavone biosynthesis and utilizing both apigenin and luteolin as substrates has already been identified on Chr 6<sup>51</sup>. However, this candidate is supported by our observations of condensed form species. Condensed forms from AR lines consist of an (epi)catechin or (epi)afzelechin linked to an anthocyanidin 3,5-diglucoside. Interestingly we have not found free (non-condensed, no flavonol unit) anthocyanidin 3,5-diglucosides nor have we found condensed forms with anthocyanins containing a single glucose moiety<sup>52</sup>. This supports the *ufgt3* candidate, especially given the rise in significance seen when mapping the proportion of condensed forms<sup>53</sup>.

Given the strong signal observed near *Aat1* for both of the above mapped traits, lines with proportions of acylated anthocyanins in the tails of the distribution were removed from the data, as described above. This effectively eliminated the signal near *Aat1*. Signals near *P1* on Chr 1 were still observed, and for anthocyanin proportions the most significant SNP was much closer to *P1* than observed previously (47.14 Mb). Several new signals also emerged for both analyses.

For anthocyanins (and flavones), the most obvious differences between the full and truncated data set were new signals that emerged on Chrs 7 and 9, however several other new signals were also present (Table 2). For Chr 7, signals at 24.1 Mb and 121.2 Mb were found. For the former, the closest candidates were *In1* (20.1 Mb) and *UFGT2* (UDP-flavonol-

glycosyltransferase2), a glycosyltransferase (Zm00001d019256, 24.2 Mb), while for the latter the closest candidates were 4-5 Mb away (predicted flavonoid 3'-monooxygenase (Zm00001d020628) at 125.5 Mb and *myb162* at 115.7 Mb. No candidates were found for the new signals on Chr 9. The peak on Chr 5 near *pr1* was still evident, but several significant SNPs were found around 205 Mb. The nearest candidates are a LAR/ANR candidate (Zm00001d017771) at 206.3 Mb and *Pac1* at 201.8 Mb.

When mapping condensed forms with the truncated data set a peak at the beginning of Chr 2 was found less than 1 Mb from *Fht1* (3.56 Mb), the flavanone 3-hydroxylase associated with the conversion of flavanones to dihydroflavanols. This step is required for the formation of both flavan-3-ols and anthocyanins; but surprisingly, this peak was only observed when mapping condensed forms. Signals on Chr 9 were also observed across a wide range as discussed previously for the un-truncated data set.

### *Genomic selection*

While understanding flux through the anthocyanin pathway may be beneficial for identifying large effect loci to be used in a marker-assisted selection (MAS) type approach, we also wanted to test the efficacy of a genomic selection approach. While MAS relies on a small number of large effect loci, the type easily detected using association mapping studies, genomic selection uses many small effect loci to predict a genomic estimated breeding value (GEBV). Therefore, complex traits with many small effect loci may be better suited to a genomic selection approach. The results for mapping proportions of flux through the pathway were fairly straightforward, with only a handful of significant loci detected. However, mapping the magnitude of flux was complex and many significant loci were identified. Genomic selection may be a better strategy for breeding this type of complex quantitative trait<sup>54</sup>.

Using five-fold cross validation we tested the effectiveness of genomic selection for several different traits (Figure 6). Focusing first on visual qualities, absorbance ( $Abs_{Max}$ ) and wavelength ( $\lambda_{Max}$ ), genomic selection was highly effective, with prediction accuracy averages of 0.76 ( $Abs_{Max}$ ) and 0.86 ( $\lambda_{Max}$ ). Accuracy was the lowest for mass (g/ear; 0.60) unsurprisingly. However, given that the ears were self-pollinations made by hand, there is likely variability in mass due to differences in pollination efficiency in addition to biological variation. Mass ranged from 4-136 g per ear. Nonetheless this prediction accuracy is relatively high for a trait as

complex as grain yield. This is likely attributable to the relatedness between samples, all having been generated from a single landrace<sup>55</sup>. The prediction accuracy average for anthocyanin content was 0.71, but we also calculated anthocyanin content on a per-plant basis, using anthocyanin concentration in  $\mu\text{g/g}$  of whole corn kernels and multiplying by grain yield. Ultimately, the feasibility of using corn for a natural colorant depends on the economical production of anthocyanins. This is influenced by both grain yield and grain anthocyanin content. When five-fold cross validation was performed for this trait, we again saw high prediction accuracies, with a mean of 0.70. This suggests a genomic selection approach may be highly effective in breeding for anthocyanin yield in the AR population.

Considering the models illustrated in Figure 3, we hypothesize that a dual approach selecting for both increased overall flux and increased partitioning toward anthocyanin biosynthesis could have a dramatic effect on anthocyanin content. Given the complexity observed with overall flux and the relatively few number of loci associated with partitioning of flux, we attempted to predict breeding value by adding known factors associated with flux to the genomic selection model as fixed effects. This approach has been suggested to improve prediction accuracy<sup>56</sup>. We added *RI*, *PII*, *PI*, and *Aat1* each individually to the GS model, but prediction accuracy did not improve from the 0.71 value observed for the GS model alone. We also added all of these to the model simultaneously which significantly reduced accuracy. These findings are similar to those recently published in a study simulating the addition of fixed effects in this manner<sup>57</sup>.

Together this data suggests that if inputs (cost, time, etc.) are equal, the GS approach is likely to outperform a MAS approach in breeding for anthocyanin content within the AR landrace. However, this accuracy is likely to decrease when parent offspring validations are calculated outside of the training population and if GS is used for making selections once new germplasm is introduced. In this scenario, a training set would need to be defined that more accurately represents the remainder of the population to be tested. While GS may be a faster way to breed for maximal anthocyanin content, it can underestimate the effect sizes of large effect loci. Furthermore, understanding the players involved in regulation and transport of anthocyanins may provide insight into the general biological processes surrounding metabolite trafficking. Identifying and understanding loci with large effects on the phenotype may also be useful if gene editing were to be used for increasing anthocyanin content.



## *Conclusion*

Highly concentrated purple pericarp is only found in a select portion of maize germplasm, most of which are not adapted to a Midwest environment and do not have yields comparable to today's hybrids. Apache Red, a pericarp-pigmented purple corn variety, offers high anthocyanin concentration and variability in extract colors<sup>52</sup>. However, due to linkage drag from donor purple corn parents, many generations of backcrossing to elite, adapted inbreds (as the recurrent parent) may be required to recover desired agronomic traits while maintaining anthocyanin concentrations. Furthermore, as different desirable traits are identified from the donor, (i.e. different anthocyanin types / decorations), additional introgression pipelines may be needed.

Pre-breeding of landraces or other un-adapted germplasm however, could make desirable traits from these sources more accessible by helping to eliminate deleterious alleles that may later be “dragged” via linkage during backcrossing<sup>58</sup>. Figure 7 illustrates the benefits of a landrace pre-breeding scheme in which the number of required backcross generations is reduced compared to an approach using a direct landrace by elite parent cross (Figure 7A). The high prediction accuracies observed here for genomic selection show that significant improvement could be made within landrace populations. Moreover, knowledge of the key factors associated with maximizing anthocyanin content could help ensure their fixation in the development of superior donor parents. This could improve the efficiency of introgression, maximizing the transfer of desirable traits while reducing movement of those that are undesirable. Lastly, superior donor parents could streamline the introgression process as breeding goals shift or change, saving the breeder from having to return to un-adapted sources.

## **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***

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”. (Private links for review: <https://figshare.com/s/e116278ff976d2e19c6e> and <https://figshare.com/s/18d7f81d7d2e9b1c47b1>)

### ***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) 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](https://doi.org/10.1007/7089_2007_117).
- (2) Chatham, L. A.; Paulsmeyer, M.; Juvik, J. A. Prospects for Economical Natural Colorants: Insights from Maize. *Theor Appl Genet* **2019**. <https://doi.org/10.1007/s00122-019-03414-0>.
- (3) 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>.
- (4) Burr, F. A.; Burr, B.; Scheffler, B. E.; Blewitt, M.; Wienand, U.; Matz, E. C. The Maize Repressor-like Gene Intensifier1 Shares Homology with the R1/B1 Multigene Family of Transcription Factors and Exhibits Missplicing. *Plant Cell* **1996**, *8* (8), 1249–1259. <https://doi.org/10.1105/tpc.8.8.1249>.
- (5) Kong, Q.; Pattanaik, S.; Feller, A.; Werkman, J. R.; Chai, C.; Wang, Y.; Grotewold, E.; Yuan, L. Regulatory Switch Enforced by Basic Helix-Loop-Helix and ACT-Domain Mediated Dimerizations of the Maize Transcription Factor R. *Proc Natl Acad Sci USA* **2012**, *109* (30), E2091–E2097. <https://doi.org/10.1073/pnas.1205513109>.
- (6) Portwood, J. L.; Woodhouse, M. R.; Cannon, E. K.; Gardiner, J. M.; Harper, L. C.; Schaeffer, M. L.; Walsh, J. R.; Sen, T. Z.; Cho, K. T.; Schott, D. A.; et al. MaizeGDB 2018: The Maize Multi-Genome Genetics and Genomics Database. *Nucleic Acids Research* **2019**, *47* (D1), D1146–D1154. <https://doi.org/10.1093/nar/gky1046>.
- (7) Albert, N. W.; Davies, K. M.; Schwinn, K. E. REPRESSION - THE DARK SIDE OF ANTHOCYANIN REGULATION? *Acta Horticulturae* **2014**, No. 1048, 129–136. <https://doi.org/10.17660/ActaHortic.2014.1048.15>.
- (8) 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>.
- (9) Grotewold, E.; Davies, K. Trafficking and Sequestration of Anthocyanins. *NPC* **2008**, *3* (8), 1251–1258.
- (10) 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>.
- (11) Goodman, C. D.; Casati, P.; Walbot, V. A Multidrug Resistance–Associated Protein Involved in Anthocyanin Transport in Zea Mays. *The Plant Cell* **2004**, *16* (7), 1812–1826. <https://doi.org/10.1105/tpc.022574>.
- (12) Debeaujon, I.; Peeters, A. J. M.; Léon-Kloosterziel, K. M.; Koornneef, M. The TRANSPARENT TESTA12 Gene of Arabidopsis Encodes a Multidrug Secondary Transporter-like Protein Required for Flavonoid Sequestration in Vacuoles of the Seed Coat Endothelium. *The Plant Cell* **2001**, *13* (4), 853–871. <https://doi.org/10.1105/tpc.13.4.853>.
- (13) Kitamura, S.; Oono, Y.; Narumi, I. Arabidopsis Pab1, a Mutant with Reduced Anthocyanins in Immature Seeds from Banyuls, Harbors a Mutation in the MATE Transporter FFT. *Plant Mol Biol* **2016**, *90* (1), 7–18. <https://doi.org/10.1007/s11103-015-0389-8>.
- (14) Gomez, C.; Terrier, N.; Torregrosa, L.; Vialet, S.; Fournier-Level, A.; Verriès, C.; Souquet, J.-M.; Mazauric, J.-P.; Klein, M.; Cheynier, V.; et al. Grapevine MATE-Type Proteins Act as Vacuolar H<sup>+</sup>-Dependent Acylated Anthocyanin Transporters. *Plant Physiology* **2009**, *150* (1), 402–415. <https://doi.org/10.1104/pp.109.135624>.

- (15) Zhao, J.; Huhman, D.; Shadle, G.; He, X.-Z.; Sumner, L. W.; Tang, Y.; Dixon, R. A. MATE2 Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in *Medicago Truncatula*. *The Plant Cell* **2011**, *23* (4), 1536–1555. <https://doi.org/10.1105/tpc.110.080804>.
- (16) Mathews, H.; Clendennen, S. K.; Caldwell, C. G.; Liu, X. L.; Connors, K.; Matheis, N.; Schuster, D. K.; Menasco, D. J.; Wagoner, W.; Lightner, J.; et al. Activation Tagging in Tomato Identifies a Transcriptional Regulator of Anthocyanin Biosynthesis, Modification, and Transport. *The Plant Cell* **2003**, *15* (8), 1689–1703. <https://doi.org/10.1105/tpc.012963>.
- (17) M'mbone, M. E.; Cheng, W.; Xu, L.; Wang, Y.; Karanja, B. K.; Zhu, X.; Cao, Y.; Liu, L. Identification and Transcript Analysis of MATE Genes Involved in Anthocyanin Transport in Radish (*Raphanus Sativus* L.). *Scientia Horticulturae* **2018**, *238*, 195–203. <https://doi.org/10.1016/j.scienta.2018.04.029>.
- (18) Chanoca, A.; Kovinich, N.; Burkel, B.; Stecha, S.; Bohorquez-Restrepo, A.; Ueda, T.; Eliceiri, K. W.; Grotewold, E.; Otegui, M. S. Anthocyanin Vacuolar Inclusions Form by a Microautophagy Mechanism. *The Plant Cell* **2015**, *27* (9), 2545–2559. <https://doi.org/10.1105/tpc.15.00589>.
- (19) 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>.
- (20) 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>.
- (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) Endelman, J. *RrBLUP: Ridge Regression and Other Kernels for Genomic Selection*; 2018.
- (23) 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>.
- (24) Foolad, M. R.; Subbiah, P.; Ghangas, G. S. Parent-Offspring Correlation Estimate of Heritability for Early Blight Resistance in Tomato, *Lycopersicon Esculentum* Mill. **2002**, No. 126, 291–297.
- (25) Hollick, J. B. Paramutation and Related Phenomena in Diverse Species. *Nat Rev Genet* **2017**, *18* (1), 5–23. <https://doi.org/10.1038/nrg.2016.115>.
- (26) 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>.
- (27) 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>.
- (28) 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>.
- (29) 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>.
- (30) 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>.

- (31) Hoopes, G. M.; Hamilton, J. P.; Wood, J. C.; Esteban, E.; Pasha, A.; Vaillancourt, B.; Provar, N. J.; Buell, C. R. An Updated Gene Atlas for Maize Reveals Organ-Specific and Stress-Induced Genes. *The Plant Journal* **2019**, *97* (6), 1154–1167. <https://doi.org/10.1111/tpj.14184>.
- (32) Lago, C.; Cassani, E.; Zanzi, C.; Landoni, M.; Trovato, R.; Pilu, R. Development and Study of a Maize Cultivar Rich in Anthocyanins: Coloured Polenta, a New Functional Food. *Plant Breeding* **2014**, *133* (2), 210–217. <https://doi.org/10.1111/pbr.12153>.
- (33) Lago, C.; Landoni, M.; Cassani, E.; Atanassiu, S.; Canta, E. Development and Characterization of a Coloured Sweet Corn Line as a New Functional Food. **2014**, *10*.
- (34) Tonelli, C.; Consonni, G.; Dolfini, S. F.; Dellaporta, S. L.; Viotti, A.; Gavazzi, G. Genetic and Molecular Analysis of Sn, a Light-Inducible, Tissue Specific Regulatory Gene in Maize. *Molec. Gen. Genet.* **1991**, *225* (3), 401–410. <https://doi.org/10.1007/BF00261680>.
- (35) Dooner, H. K.; Kermicle, J. L. Displaced and Tandem Duplications in the Long Arm of Chromosome 10 in Maize. *Genetics* **1976**, *82* (2), 309–322.
- (36) Petroni, K.; Cominelli, E.; Consonni, G.; Gusmaroli, G.; Gavazzi, G.; Tonelli, C. The Developmental Expression of the Maize Regulatory Gene Hopi Determines Germination-Dependent Anthocyanin Accumulation. *Genetics* **2000**, *155* (1), 323–336.
- (37) Reddy, A. R.; Peterson, P. A. The Action of the Intensifier (in) Gene in Flavonoid Production in Aleurone Tissue of Maize. *Can. J. Genet. Cytol.* **1978**, *20* (3), 337–347. <https://doi.org/10.1139/g78-038>.
- (38) Lee, E. A.; Byrne, P. F.; McMullen, M. D.; Snook, M. E.; Wiseman, B. R.; Widstrom, N. W.; Coe, E. H. Genetic Mechanisms Underlying Apimaysin and Maysin Synthesis and Corn Earworm Antibiosis in Maize (*Zea Mays* L.). *Genetics* **1998**, *149* (4), 1997–2006.
- (39) 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>.
- (40) Zhao, J.; Dixon, R. A. MATE Transporters Facilitate Vacuolar Uptake of Epicatechin 3'-O-Glucoside for Proanthocyanidin Biosynthesis in *Medicago truncatula* and *Arabidopsis*. *The Plant Cell* **2009**, *21* (8), 2323–2340. <https://doi.org/10.1105/tpc.109.067819>.
- (41) 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>.
- (42) Gargouri, M.; Manigand, C.; Maugé, C.; Granier, T.; Langlois d'Estaintot, B.; Cala, O.; Pianet, I.; Bathany, K.; Chaudière, J.; Gallois, B. Structure and Epimerase Activity of Anthocyanidin Reductase from *Vitis Vinifera*. *Acta Crystallographica Section D Biological Crystallography* **2009**, *65* (9), 989–1000. <https://doi.org/10.1107/S0907444909025013>.
- (43) Pang, Y.; Abeyasinghe, I. S. B.; He, J.; He, X.; Huhman, D.; Mewan, K. M.; Sumner, L. W.; Yun, J.; Dixon, R. A. Functional Characterization of Proanthocyanidin Pathway Enzymes from Tea and Their Application for Metabolic Engineering. *Plant Physiol.* **2013**, *161* (3), 1103–1116. <https://doi.org/10.1104/pp.112.212050>.
- (44) Zhu, Y.; Wang, H.; Peng, Q.; Tang, Y.; Xia, G.; Wu, J.; Xie, D.-Y. Functional Characterization of an Anthocyanidin Reductase Gene from the Fibers of Upland Cotton (*Gossypium hirsutum*). *Planta* **2015**, *241* (5), 1075–1089. <https://doi.org/10.1007/s00425-014-2238-4>.
- (45) 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>.
- (46) Appelhagen, I.; Nordholt, N.; Seidel, T.; Spelt, K.; Koes, R.; Quattrocchio, F.; Sagasser, M.; Weisshaar, B. TRANSPARENT TESTA 13 Is a Tonoplast P3A-ATPase Required for Vacuolar Deposition of Proanthocyanidins in *Arabidopsis thaliana* Seeds. *The Plant Journal* **2015**, *82* (5), 840–849. <https://doi.org/10.1111/tpj.12854>.



- (47) Verweij, W.; Spelt, C.; Sansebastiano, G.-P. D.; Vermeer, J.; Reale, L.; Ferranti, F.; Koes, R.; Quattrocchio, F. An H<sup>+</sup> 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) 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>.
- (49) Meyers, M. T. A Second Recessive Factor for Brown Pericarp in Maize. **1927**.
- (50) 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>.
- (51) 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>.
- (52) 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>.
- (53) González-Manzano, S.; Pérez-Alonso, J. J.; Salinas-Moreno, Y.; Mateus, N.; Silva, A. M. S.; de Freitas, V.; Santos-Buelga, C. Flavanol–Anthocyanin Pigments in Corn: NMR Characterisation and Presence in Different Purple Corn Varieties. *Journal of Food Composition and Analysis* **2008**, *21* (7), 521–526. <https://doi.org/10.1016/j.jfca.2008.05.009>.
- (54) Jannink, J.-L.; Lorenz, A. J.; Iwata, H. Genomic Selection in Plant Breeding: From Theory to Practice. *Brief Funct Genomics* **2010**, *9* (2), 166–177. <https://doi.org/10.1093/bfgp/elq001>.
- (55) Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J. M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* **2017**, *22* (11), 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>.
- (56) Bernardo, R. Genomewide Selection When Major Genes Are Known. *Crop Science* **2014**, *54* (1), 68–75. <https://doi.org/10.2135/cropsci2013.05.0315>.
- (57) Rice, B.; Lipka, A. E. Evaluation of RR-BLUP Genomic Selection Models That Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum. *The Plant Genome* **2019**, *12* (1). <https://doi.org/10.3835/plantgenome2018.07.0052>.
- (58) Gorjanc, G.; Jenko, J.; Hearne, S. J.; Hickey, J. M. Initiating Maize Pre-Breeding Programs Using Genomic Selection to Harness Polygenic Variation from Landrace Populations. *BMC Genomics* **2016**, *17* (1). <https://doi.org/10.1186/s12864-015-2345-z>.

## Tables and Figures

**Table 1:** Anthocyanin regulatory and transport / storage factors. See Chatham et al. (2019) for a complete overview of these factors.

	<b>Locus</b>	<b>Gene product</b>	<b>Gene model identifier</b>	<b>Chromosome</b>	<b>Mb</b>	
<b><u>Regulatory</u></b>						
	<i>pl1</i>	purple plant1	R2R3-MYB	Zm00001d037118	Chr 6	113.0
	<i>c1</i>	colored aleurone1		Zm00001d044975	Chr 9	9.0
	<i>b1</i>	colored plant1	bHLH (basic helix-loop-helix)	Zm00001d000236	Chr 2	19.0
	<i>r1</i>	colored1		Zm00001d026147	Chr 10	139.8
	<i>hopi1</i>	hopi r1/b1 family member1		Zm00001d026147		
	<i>lc1</i>	red leaf color1		Zm00001d026147		
	<i>sn1</i>	scutellar node color1		Zm00001d026147		
	<i>pac1</i>	pale aleurone color1	WDR (WD repeat protein)	Zm00001d017617	Chr 5	201.8
	<i>a3</i>	anthocyaninless3	unknown	no gene model		
	<i>p1</i>	pericarp color1	MYB	Zm00001d028842	Chr 1	48.4
	<i>in1</i>	intensifier1	bHLH-like inhibitor	Zm00001d019170	Chr 7	20.1
	<i>rif1</i>	r-interacting factor1	EMSY-like / ENT domain containing protein	Zm00001d019922	Chr 7	76.6
<b><u>Transport / Storage</u></b>						
	<i>bz2</i>	bronze2	GST (glutathione S-transferase)	Zm00001d032969	Chr 1	245.0
	<i>mrp3</i>	multidrug resistance-associated protein3	MRP (multidrug resistance-associated protein, ABC transporter)	Zm00001d045269	Chr 9	17.5



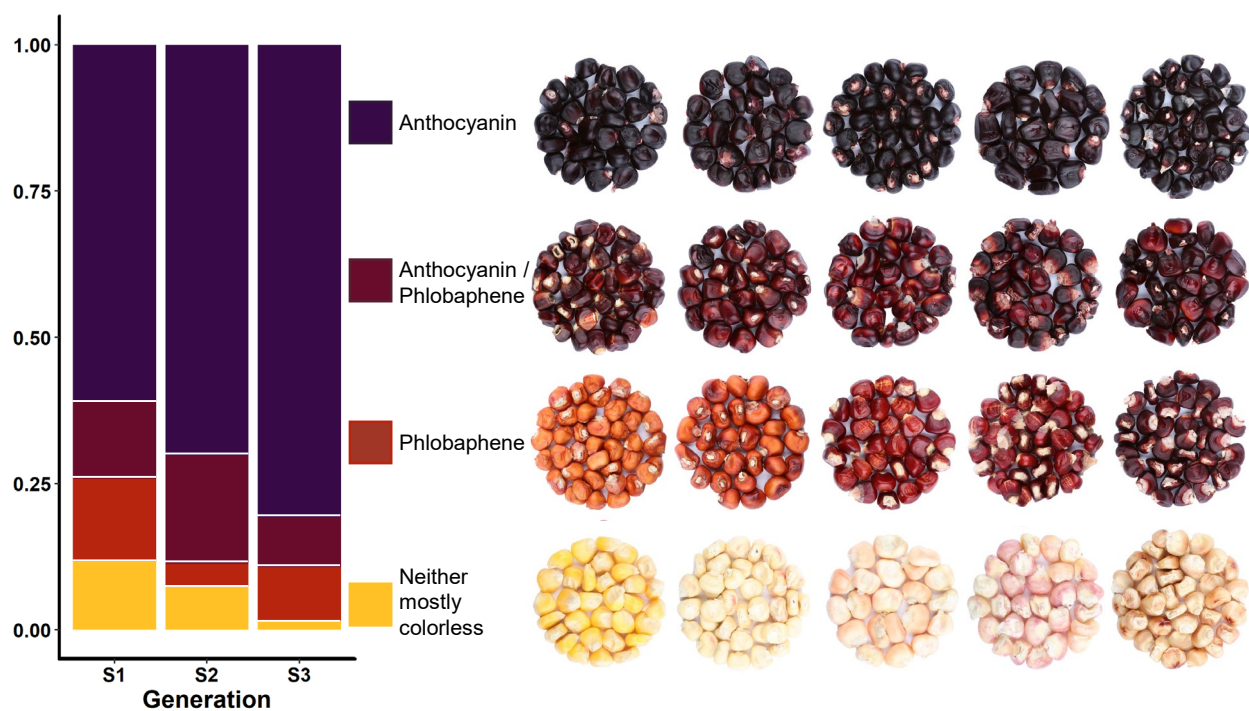
**Table 2:** Significant SNPs\* for each anthocyanin variable mapped

Variable	Range of significant SNPs			SNP with minimum p-value					
	Chr	Min position (Mb)	Max position (Mb)	Position (Mb)	P value	Effect	MAF		
<i>Total Anthocyanin Content</i>	**	1	181.98	190.39	190.39	4.11E-08	-14.20	0.08	
		1	219.17	219.83	219.83	2.19E-06	22.40	0.06	
	**	1	302.39	306.42	306.42	1.42E-19	-21.80	0.19	
	**	2	150.38	150.38	150.38	1.73E-07	-14.80	0.06	
	**	3	227.88	227.88	227.88	7.39E-07	-17.40	0.06	
	**	5	2.69	8.56	8.56	2.69E-08	9.81	0.31	
	**	6	20.15	41.68	21.26	1.27E-10	-25.90	0.07	
	**	6	112.79	116.37	113.98	2.24E-09	15.20	0.15	
	**	7	16.90	34.79	34.79	2.21E-07	15.30	0.08	
	**	7	117.64	128.93	121.75	1.20E-08	-15.70	0.17	
	**	7	144.83	146.65	145.45	8.33E-08	-14.40	0.10	
	**	9	132.37	145.32	143.53	2.64E-09	21.00	0.07	
		10	21.74	21.74	21.74	1.93E-06	-20.70	0.06	
		10	110.76	110.76	110.76	3.39E-06	17.40	0.07	
		**	10	136.88	140.84	139.50	1.03E-12	22.80	0.27
	<i>Abs<sub>Max</sub></i>	**	1	32.15	39.96	32.15	7.35E-08	-0.49	0.16
**		1	219.17	219.83	219.83	3.62E-08	0.78	0.06	
**		1	305.34	306.42	305.59	1.81E-14	-0.70	0.10	
**		2	150.38	150.38	150.38	3.08E-07	-0.43	0.06	
**		3	14.75	14.75	14.75	3.62E-07	0.60	0.07	
**		5	8.56	8.56	8.56	1.45E-07	0.28	0.31	
**		6	21.26	41.68	25.97	5.64E-09	0.84	0.06	
**		6	112.79	117.74	113.98	2.75E-09	0.45	0.15	
**		7	121.75	121.75	121.75	4.64E-07	-0.41	0.17	
**		9	11.75	11.75	11.75	5.58E-07	0.35	0.34	
**		9	132.37	145.32	143.53	1.50E-09	0.64	0.07	
**		10	21.74	21.74	21.74	6.00E-07	-0.65	0.06	
**		10	136.88	140.68	138.48	5.32E-11	-0.70	0.06	
		**	1	3.11	3.91	3.38	7.36E-09	-29.76	0.05
	**	1	19.59	39.96	27.29	7.48E-09	17.42	0.14	

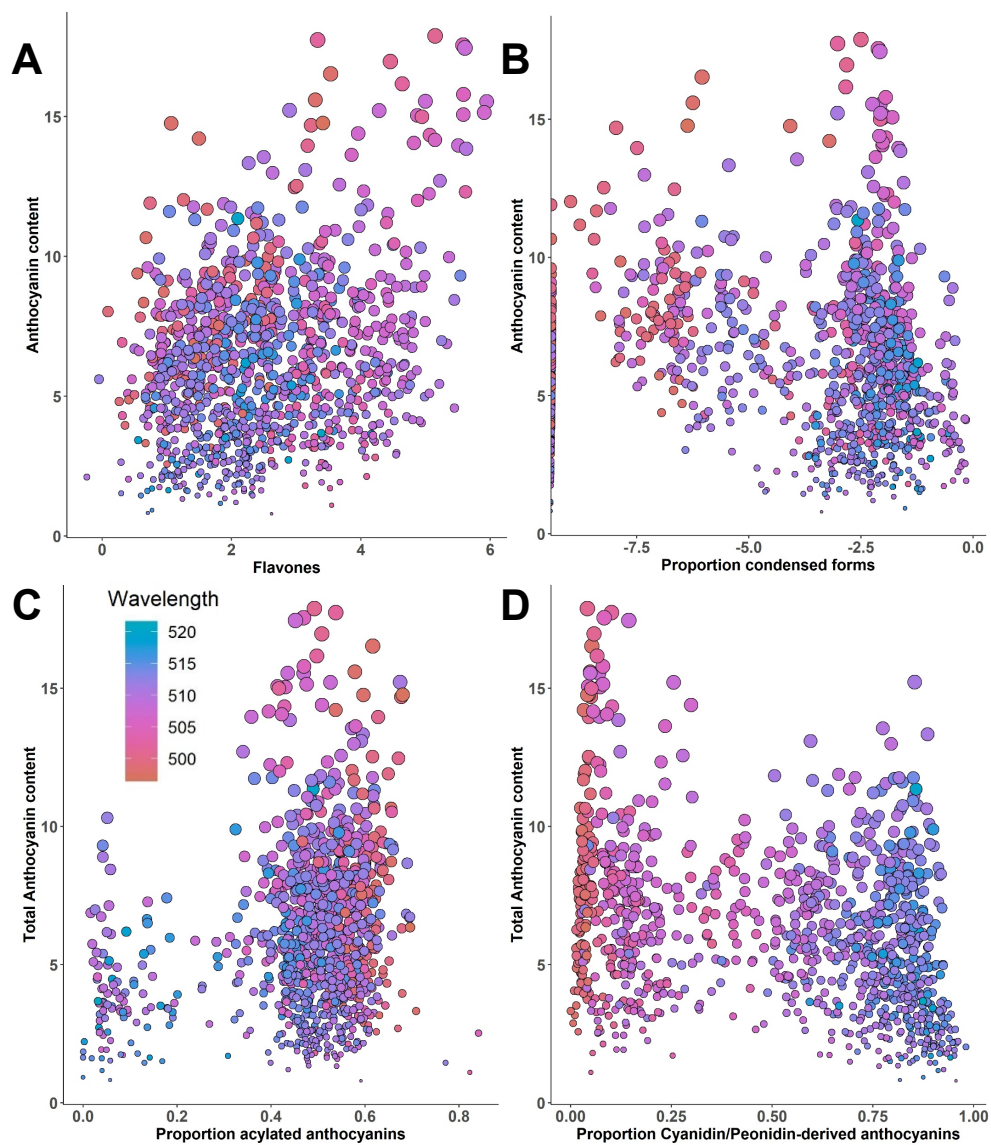
Full Data Set

		Truncated Data Set							
<b>Portions - Full Data Set</b>	<i>Total Anthocyanin Content</i>	1	181.98	190.39	190.39	1.88E-06	-13.57	0.08	
		1	219.17	219.17	219.17	1.10E-06	-29.01	0.05	
		3	1.40	1.40	1.40	3.72E-06	-50.68	0.07	
		**	6	113.00	113.00	113.00	1.66E-07	-10.37	0.17
		**	7	18.11	34.79	22.61	1.07E-07	9.27	0.48
		**	7	117.64	130.47	121.75	2.41E-10	-18.08	0.16
		**	7	177.72	177.72	177.72	5.18E-08	25.59	0.05
		**	9	11.47	11.47	11.47	3.13E-07	17.22	0.08
			10	35.48	35.48	35.48	1.14E-06	-14.22	0.13
		**	10	69.68	69.68	69.68	1.13E-07	-15.06	0.13
		10	83.96	83.96	83.96	7.98E-07	17.63	0.08	
	**	10	137.30	142.86	139.50	3.73E-15	28.51	0.28	
	<i>Abs<sub>Max</sub></i>	**	1	19.59	51.84	39.96	3.84E-16	-1.03	0.23
		**	1	181.98	190.39	190.39	4.65E-07	-0.42	0.08
		**	1	219.17	219.17	219.17	2.52E-07	-0.90	0.05
			3	53.87	53.87	53.87	2.29E-06	-0.43	0.06
			7	21.06	21.06	21.06	2.39E-06	-0.27	0.47
		**	7	121.75	121.75	121.75	3.06E-09	-0.52	0.16
		**	9	10.91	11.75	11.47	2.16E-09	0.60	0.08
			10	19.51	19.51	19.51	2.14E-06	0.46	0.11
**		10	35.48	35.48	35.48	1.41E-08	-0.50	0.13	
**		10	66.12	69.68	69.68	3.42E-09	-0.51	0.13	
**	10	139.50	141.74	139.50	7.72E-12	0.74	0.28		
<i>Anthocyanins</i>	**	1	27.09	53.37	35.49	2.62E-22	0.17	0.10	
	**	1	299.44	306.42	306.42	7.50E-23	-0.12	0.19	
		2	36.11	36.11	36.11	1.04E-06	0.05	0.39	
		3	174.87	174.87	174.87	3.51E-06	0.07	0.06	
	**	4	33.19	33.19	33.19	6.45E-07	0.08	0.10	
		5	183.34	183.60	183.60	1.64E-06	-0.04	0.44	
		5	213.51	213.51	213.51	3.38E-06	-0.17	0.01	
	**	6	112.48	116.37	113.98	8.34E-16	0.10	0.15	
		6	170.40	170.40	170.40	3.63E-06	0.05	0.13	
		7	156.44	156.44	156.44	2.48E-06	0.08	0.08	
**	9	23.85	23.85	23.85	9.81E-08	-0.23	0.03		
	10	130.74	149.20	130.74	1.87E-06	0.09	0.05		

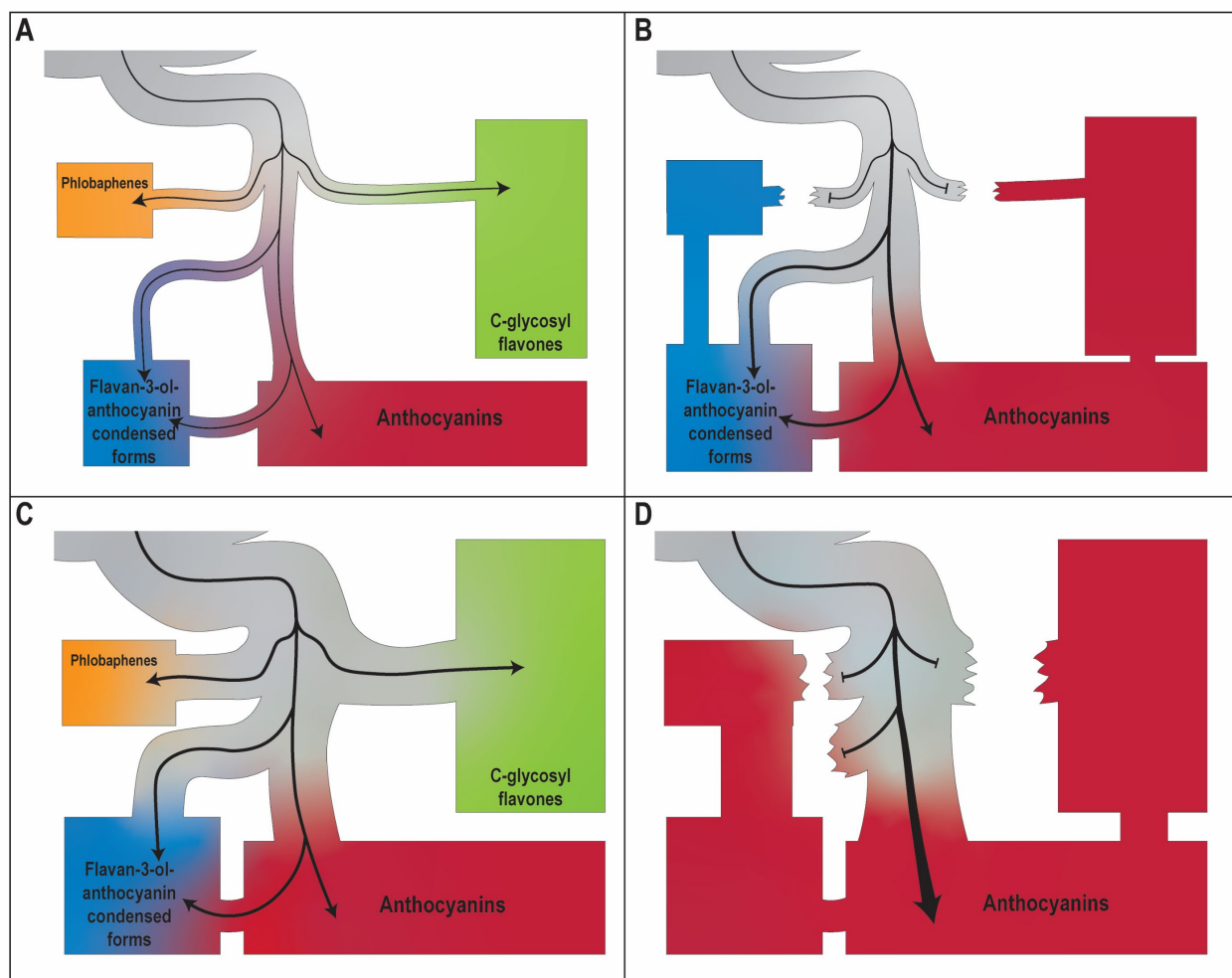
		Molar Prop							
		**							
<i>Flavanol- Anthocyanin Condensed Forms</i>	**	1	14.36	53.46	32.51	9.09E-19	0.01	0.27	
	**	1	297.07	306.42	297.07	1.71E-07	0.01	0.05	
	**	3	1.40	1.40	1.40	6.03E-07	-0.01	0.10	
		3	52.09	52.09	52.09	1.14E-06	0.01	0.03	
	**	3	174.37	174.37	174.37	3.22E-07	-0.01	0.04	
	**	4	176.64	176.64	176.64	1.31E-08	0.01	0.07	
		7	142.26	142.26	142.26	1.79E-06	-0.01	0.10	
	**	9	16.63	34.51	16.63	2.59E-08	-0.01	0.08	
		9	59.96	81.44	76.07	1.10E-06	0.01	0.08	
	**	9	130.35	130.35	130.35	7.70E-08	-0.01	0.14	
<i>Anthocyanin</i>	**	1	32.36	52.94	47.14	2.83E-19		0.48	
	**	2	35.25	36.11	36.11	2.43E-07		0.39	
	**	4	93.54	93.54	93.54	4.07E-09		0.08	
	**	5	183.60	184.03	183.60	1.68E-07		0.41	
		5	205.09	205.09	205.09	2.35E-06		0.14	
	**	6	112.79	115.20	113.98	4.05E-08		0.12	
		7	24.10	24.10	24.10	1.72E-06		0.31	
	**	7	121.24	121.38	121.24	2.84E-07		0.19	
	**	8	168.48	168.48	168.48	5.60E-07		0.05	
	**	9	115.42	115.42	115.42	7.98E-09		0.30	
		9	138.95	138.95	138.95	2.83E-06		0.14	
	<i>Flavanol- Anthocyanin Condensed Forms</i>	**	1	19.59	54.03	39.96	5.24E-22		0.23
		**	2	3.10	5.46	3.37	1.76E-07		0.29
		**	9	11.75	34.51	19.94	9.24E-11		0.07
			9	42.41	42.41	42.41	1.58E-06		0.06
		**	9	59.96	59.96	59.96	2.80E-07		0.06
		**	9	76.07	76.11	76.11	2.74E-07		0.06
		**	9	81.44	81.44	81.44	1.65E-07		0.07
		**	9	94.21	94.21	94.21	1.54E-07		0.05
**		9	115.42	115.42	115.42	4.41E-08		0.30	



**Figure 1:** Proportions of lines from each generation containing anthocyanin and phlobaphenes based on visual assessment. Kernel pictures show ranges for each category. Generations are labeled based on maternal pericarp (i.e. S2, S3, and S4 plants have pericarp have maternal pericarp representative the S1, S2, and S3 generations, respectively)

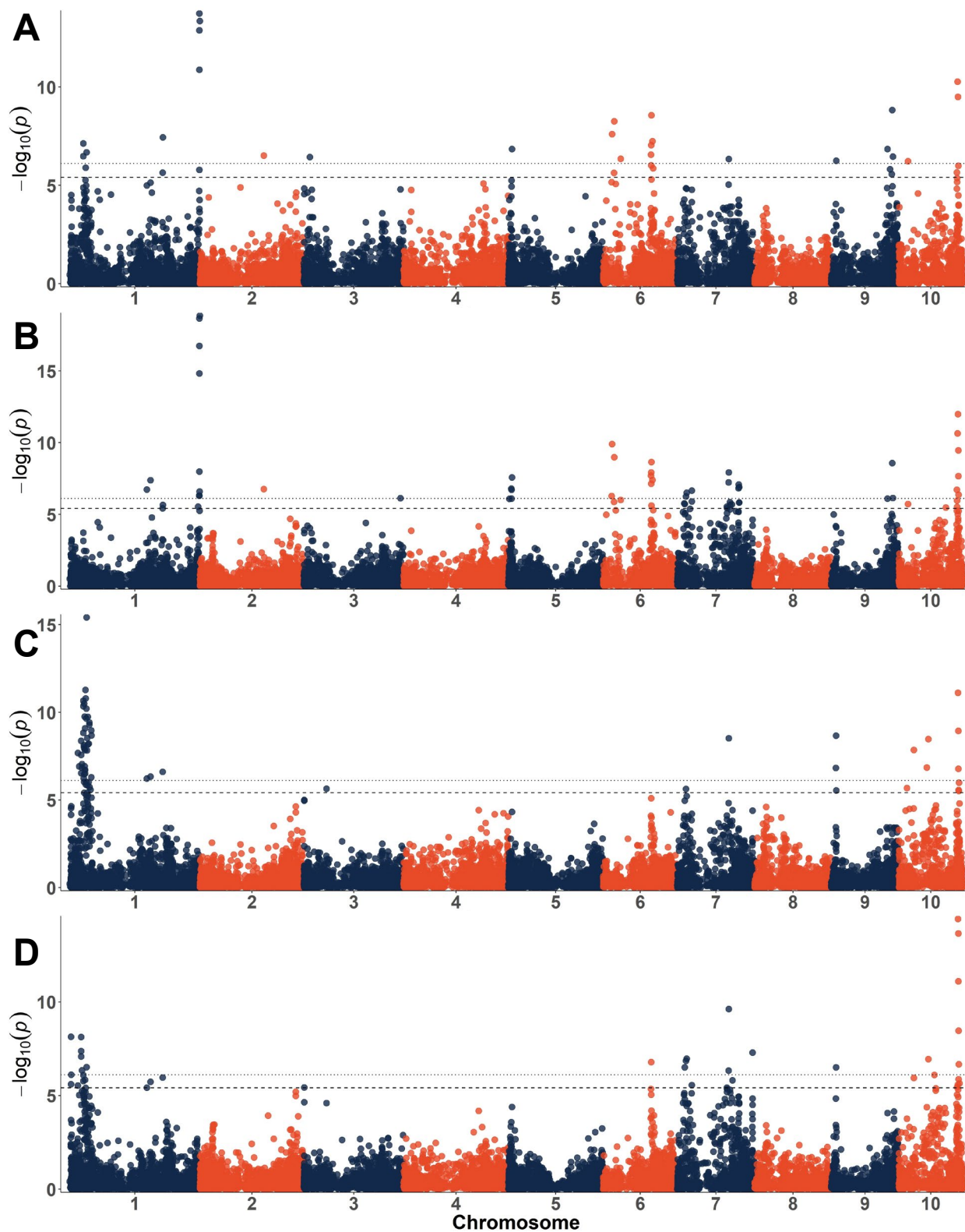


**Figure 2:** Effects of flavone content (A, log transformed,  $\mu\text{g/ml}$ ), and proportions of condensed (B, log transformed), acylated (C), and cyanidin/peonidin-derived (D) anthocyanins on normalized total anthocyanin content. Points are sized according to normalized total anthocyanin content (square root transformed,  $\mu\text{g/ml}$ ). Points are colored according to wavelength.



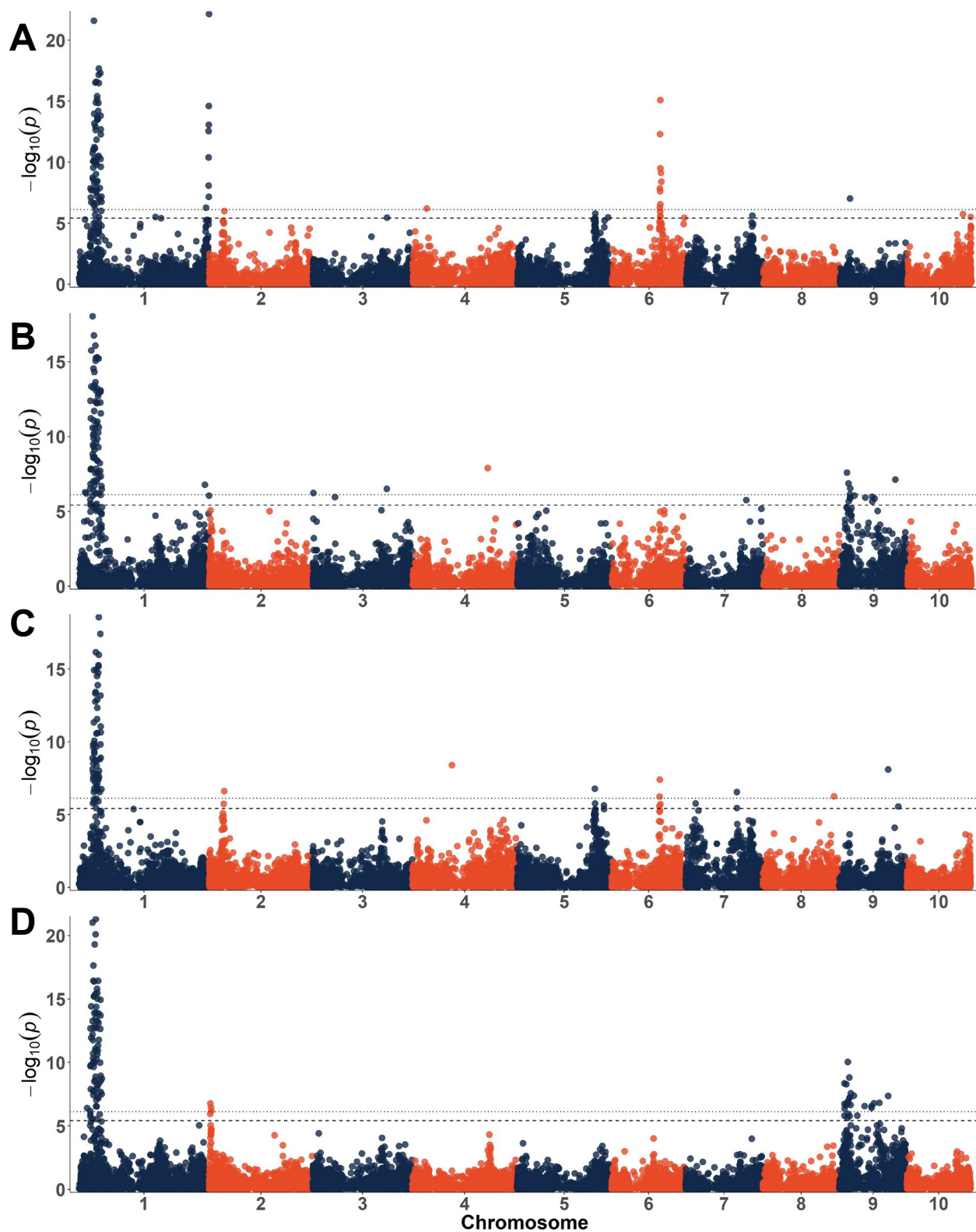
**Figure 3:** Schematic showing simplified potential strategies for increasing anthocyanin content. Diagrams are laid out to resemble the pathway presented in Figure 6.x. (A) Assumed normal flux through the pathway with all major products made. (B) Loss of structural genes or regulatory factors required for phlobaphene (orange) and flavone (green) biosynthesis (e.g. P1) shunts flux toward flavonols and anthocyanins. (C) Increased total flux through the entire pathway, and corresponding increases in all end products. (D) Combined increased flux (C) and partitioning of flux (B) toward anthocyanin biosynthesis.



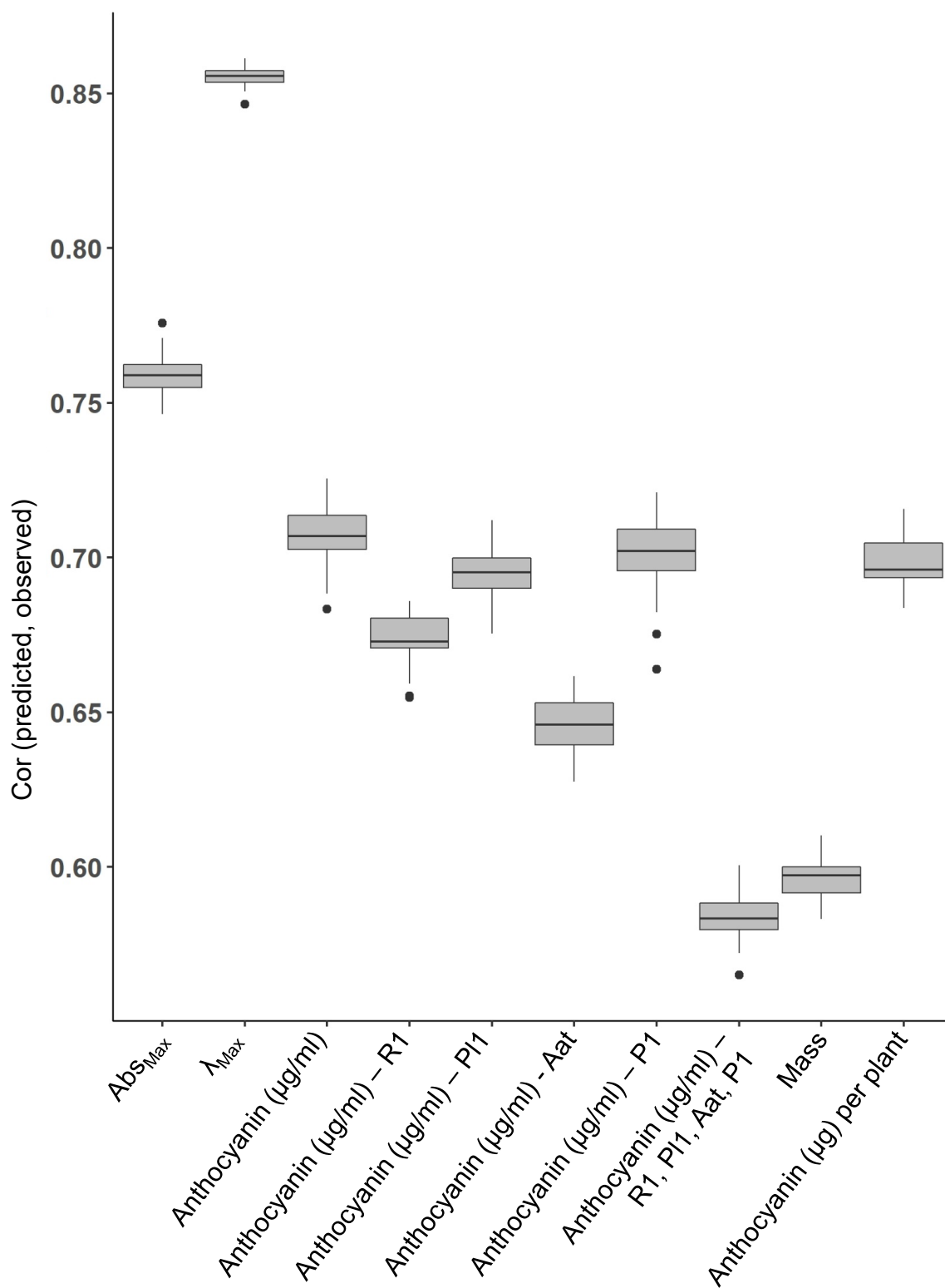


**Figure 4:** Manhattan plots for maximum absorbance (A, C) and total anthocyanin content in  $\mu\text{g/ml}$  (B, D). A and B contain data from the full population while C and D contain only the lines with proportions of acylated anthocyanins between 0.35 and 0.75.

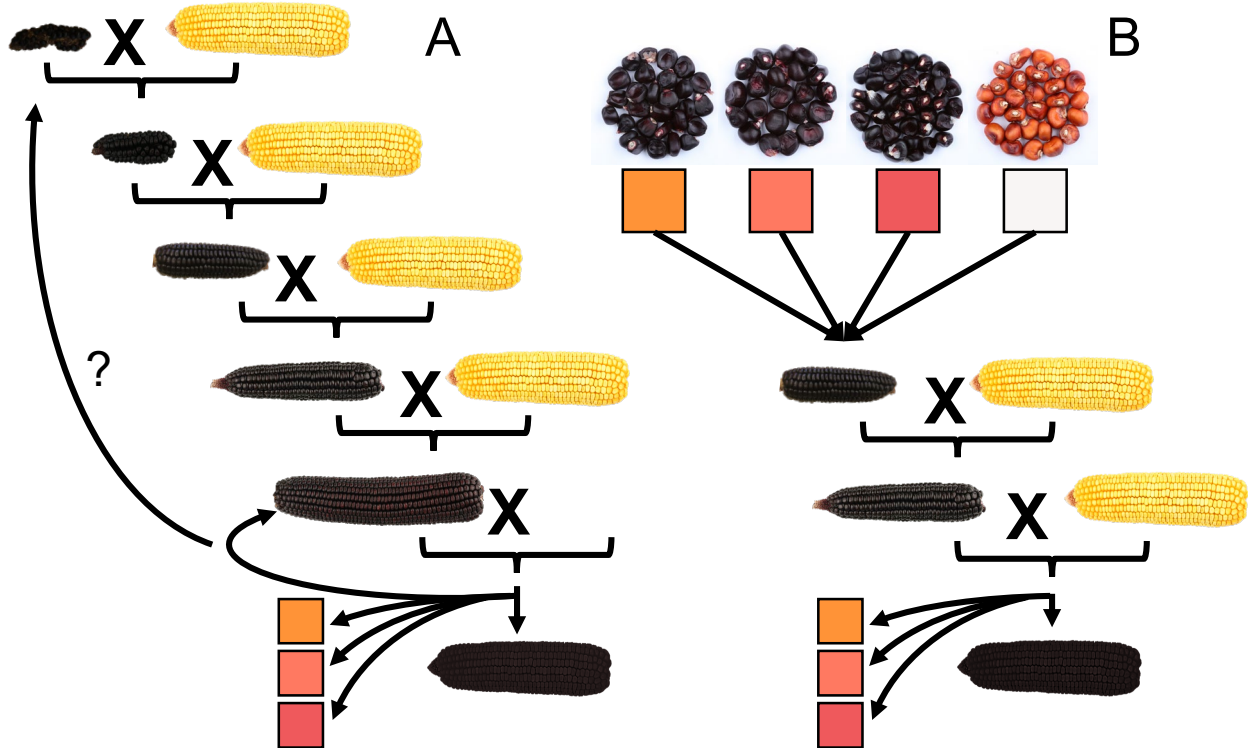




**Figure 5:** Manhattan plots for the partitioning of flux through the flavonoid pathway, showing (A) proportion anthocyanins and the (B) proportion of condensed forms



**Figure 6:** Prediction accuracy of genomic selection.



**Figure 7:** Traditional introgression (A) compared to landrace improvement (B). When multiple traits are required from a landrace, improvement of landrace materials prior to introgression may reduce linkage drag and the number of backcrosses required to regain necessary agronomic traits while maintaining the landrace sourced trait.

**Supplementary Table 1** : List of potential gene candidates.

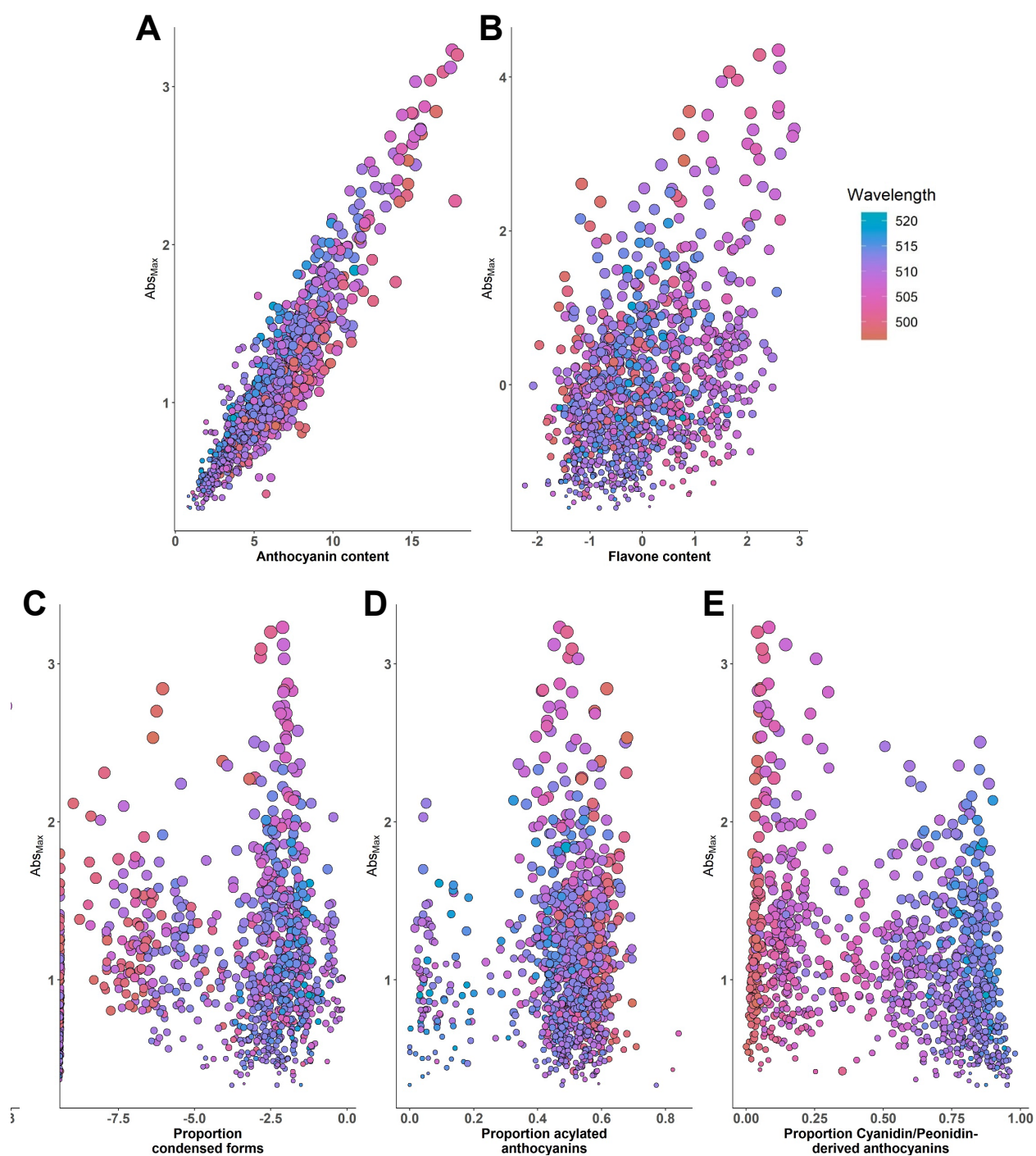
Chromosome	Mb	Gene model	Description
1	6.43	Zm00001d027490	blast hit peutnia ATPase
1	19.76	Zm00001d027997	predicted glucosyltransferase; UDP-D-glucose 5-O-b-D-glucosyltransferase; anthocyanidin 5,3-O glycos
1	34.65	Zm00001d028426	UDP-D-glucuronic acid 4-epimerase, regulated by P1
1	34.96	Zm00001d028436	vacuolar proton ATPase a3
1	41.83	Zm00001d028653	UDP-D-xylose synthase, regulated by P1
1	41.84	Zm00001d028654	rhamnose synthase gene, regulated by P1
1	48.42	Zm00001d028842	P1, MYB
1	51.84	Zm00001d028930	myb75, pza03168
1	62.75	Zm00001d029218	Fht2 , F3H (flavonone 3-hydroxylase)
1	63.57	Zm00001d029243	flavonoid 3'-monooxygenase; predicted
1	74.78	Zm00001d029526	flavonoid 3'-monooxygenase; predicted
1	85.07	Zm00001d029744	flavanone 3-beta hydroxylase (fnsi1), regulated by P1
1	143.16	Zm00001d030548	leucoanthocyanidin dioxygenase, regulated by P1
1	172.56	Zm00001d030996	vpp9 - vacuolar proton pump9
1	191.77	Zm00001d031488	LAR/ANR ortholog?
1	191.77	Zm00001d031489	LAR/ANR ortholog?
1	202.21	Zm00001d031802	dihydroflavonol 4-reductase; predicted
1	209.09	Zm00001d031997	LAR/ANR ortholog?
1	209.72	Zm00001d032024	myb38, Zm 38, regulation of phenylpropanoid pathway
1	210.34	Zm00001d032042	flavonoid 3'-monooxygenase; predicted
1	213.13	Zm00001d032103	4-coumarate-CoA ligase, predicted
1	215.90	Zm00001d032178	myb83; blast hit for <i>AtMYB60</i>
1	245.01	Zm00001d032969	Bz2, GST (glutathione S-transferase), regulated by P1
1	249.13	Zm00001d033062	umc1991; FaTT12 blast match; transparent testa 12 protein
1	259.13	Zm00001d033327	flavanone 3-dioxygenase, predicted
1	266.97	Zm00001d033552	blast hit peutnia ATPase
1	283.16	Zm00001d034072	UDP-D-glucose dehydrogenase, regulated by P1
1	290.53	Zm00001d034357	LAR/ANR ortholog?
1	298.58	Zm00001d034635	Chi1, chalcone isomerase , regulated by P1
1	305.70	Zm00001d034925	Aat1 , AAT (anthocyanin acyltransferase)regulated by P1
2	3.02	Zm00001d001924	flavanone 3-dioxygenase, predicted
2	3.06	Zm00001d001927	flavanone 3-dioxygenase, predicted
2	3.56	Zm00001d001960	Fht1 , F3H (flavonone 3-hydroxylase),
2	4.28	Zm00001d002006	blast hit for AtTT13; plasma membrane ATPase
2	13.59	Zm00001d002476	myb13, regulation of phenylpropanoid pathway
2	19.04	Zm00001d000236	B1, bHLH (basic helix-loop-helix)
2	32.09	Zm00001d003102	UDP-glucosyltransferase
2	56.99	Zm00001d003741	flavonoid 3'-monooxygenase; predicted
2	117.76	Zm00001d004555	naringenin 2-hydroxylase, predicted
2	135.50	Zm00001d004744	myb64, Blast hit for NtMYB6, PtrMYB142
2	154.11	Zm00001d005018	MATE blast hit
2	158.48	Zm00001d005106	Rif1 similar
2	179.27	Zm00001d005577	flavonoid 3'-monooxygenase; predicted
2	188.23	Zm00001d005784	myb111
2	189.70	Zm00001d005821	flavonoid 3'-monooxygenase; predicted
2	189.80	Zm00001d005823	flavonoid 3'-monooxygenase; predicted
2	201.26	Zm00001d006193	flavonoid 3'-monooxygenase; predicted
2	202.58	Zm00001d006236	myb31, regulation of phenylpropanoid pathway
2	208.46	Zm00001d006446	Sm2 , RHT (rhamnosyl transferase)
2	208.54	Zm00001d006449	predicted glucosyltransferase; UDP-D-glucose 5-O-b-D-glucosyltransferase; anthocyanidin 5,3-O glycos
2	209.82	Zm00001d006493	Blast hit for MtMATE1, AtFFT, AtTT12
2	209.83	Zm00001d006494	putative mate
2	214.24	Zm00001d006684	flavonoid 3-O-glucosyltransferase, regulated by P1
2	224.25	Zm00001d007191	myb110, blast hit for MtMYB2
2	230.95	Zm00001d007403	whp1 , CHS (chalcone synthase),

2	239.39	Zm00001d007775	mha1 - membrane H(+)-ATPase1, blast hit for petunia ATPase
2	242.41	Zm00001d007918	flavonoid 3'-hydroxylase, predicted
2	242.87	Zm00001d007924	naringenin 2-hydroxylase; predicted
3	2.71	Zm00001d039371	tcptf33 - TCP-transcription factor 33; blast match for Arabidopsis TCP3
3	30.59	Zm00001d040173	irl1-isoflavone reductase-like1, Medicago LAR blast match
3	46.20	Zm00001d040483	naringenin-chalcone synthase, predicted
3	54.36	Zm00001d040621	myb40
3	58.96	Zm00001d040701	anthocyanidin synthase, predicted
3	105.72	Zm00001d041214	blast hit peutnia ATPase
3	128.15	Zm00001d041580	myb118
3	133.01	Zm00001d041679	UDP-D-apiose/UDP-D-xylose synthase 2, regulated by P1
3	135.60	Zm00001d041740	wrky58
3	135.77	Zm00001d041746	LAR/ANR ortholog?
3	136.45	Zm00001d041763	flavonoid 3-O-glucosyltransferase, regulated by P1
3	140.88	Zm00001d041853	myb8
3	142.23	Zm00001d041882	atp1-ATPase1
3	158.15	Zm00014a017739	flavanone 3-beta hydroxylase, regulated by P1
3	159.07	Zm00001d042287	myb2 / zm 1, regulation of phenylpropanoid pathway
3	173.06	Zm00001d042594	anthocyanidin reductase, predicted; LAR/ANR candidate
3	175.96	Zm00001d042665	myb32
3	176.72	Zm00001d042679	flavonoid 3'-monooxygenase; predicted
3	178.59	Zm00001d042740	flavonoid 3-O-glucosyltransferase, regulated by P1
3	184.61	Zm00001d042943	UDP-D-xylose synthase, regulated by P1
3	185.75	Zm00001d042980	flavanone 3- beta hydroxylase; anthocyanidin synthase (predicted)
3	186.84	Zm00001d043025	wrky43-WRKY-transcription factor 43 - blast hit for <i>PH3</i> (Petunia wrky) and <i>Tig2</i> (Arabidopsis wrky)
3	211.32	Zm00001d043837	myb blas match
3	219.87	Zm00001d044122	A1 , DFR / F4R (bifunctional dihydroflavonol 4-reductase / flavanone 4-reductase),
3	221.63	Zm00001d044194	mybr97, MYB-like transcription factor ETC3
3	234.82	Zm00001d044683	chalcone isomerase, regulated by P1
3	205.4 - 215.9	NA	A3, anthocyaninless3, negative regulator, location based on B73_REFGEN_V2
4	5.74	Zm00001d048800	anthocyanidin synthase, predicted
4	10.62	Zm00001d048954	GDP-D-mannose 3,5-epimerase, regulated by P1
4	52.07	Zm00001d049910	VvAMT anthomate hit, blast match for AtFFT
4	73.91	Zm00001d050214	blast hit peutnia ATPase
4	76.14	Zm00001d050247	wrky97 - WRKY-transcription factor 97
4	133.31	Zm00001d050955	flavanonoid-3'hydroxylase, regulated by P1
4	153.48	Zm00007a00015593	rhamnose synthase gene, regulated by P1
4	161.21	Zm00001d051520	myb19, regulation of phenylpropanoid pathway
4	161.59	Zm00001d051529	4-coumarate-CoA ligase, predicted
4	169.86	Zm00001d051788	myb142
4	173.66	Zm00001d051886	VvAMT anthomate hit, MATE blast match for AtFFT
4	176.75	Zm00001d052022	blast hit for <i>AtTT13</i> ; plasma membrane ATPase
4	177.83	Zm00001d052053	UDP-D-glucuronic acid 4-epimerase, regulated by P1
4	191.25	Zm00001d052492	anthocyanidin 3-O-glucosyltransferase, predicted
4	196.89	Zm00001d052673	C2 , CHS (chalcone synthase),
4	197.24	Zm00001d052683	AOMT ortholog
4	197.34	Zm00001d052684	AOMT ortholog
4	202.39	Zm00001d052841	AOMT ortholog
4	202.4	Zm00001d052842	AOMT ortholog
4	202.4	Zm00001d052843	AOMT ortholog
4	202.45	Zm00001d052847	wrky87
4	211.44	Zm00001d053060	myb67
4	214.61	Zm00001d053124	myb61
4	216.04	Zm00001d053144	anthocyanidin synthase, predicted
4	219.69	Zm00001d053202	flavonoid 3'-monooxygenase; predicted
4	220.72	Zm00001d053220	myb42, regulation of phenylpropanoid pathway
4	221.21	Zm00001d053234	putative MATE efflux family protein

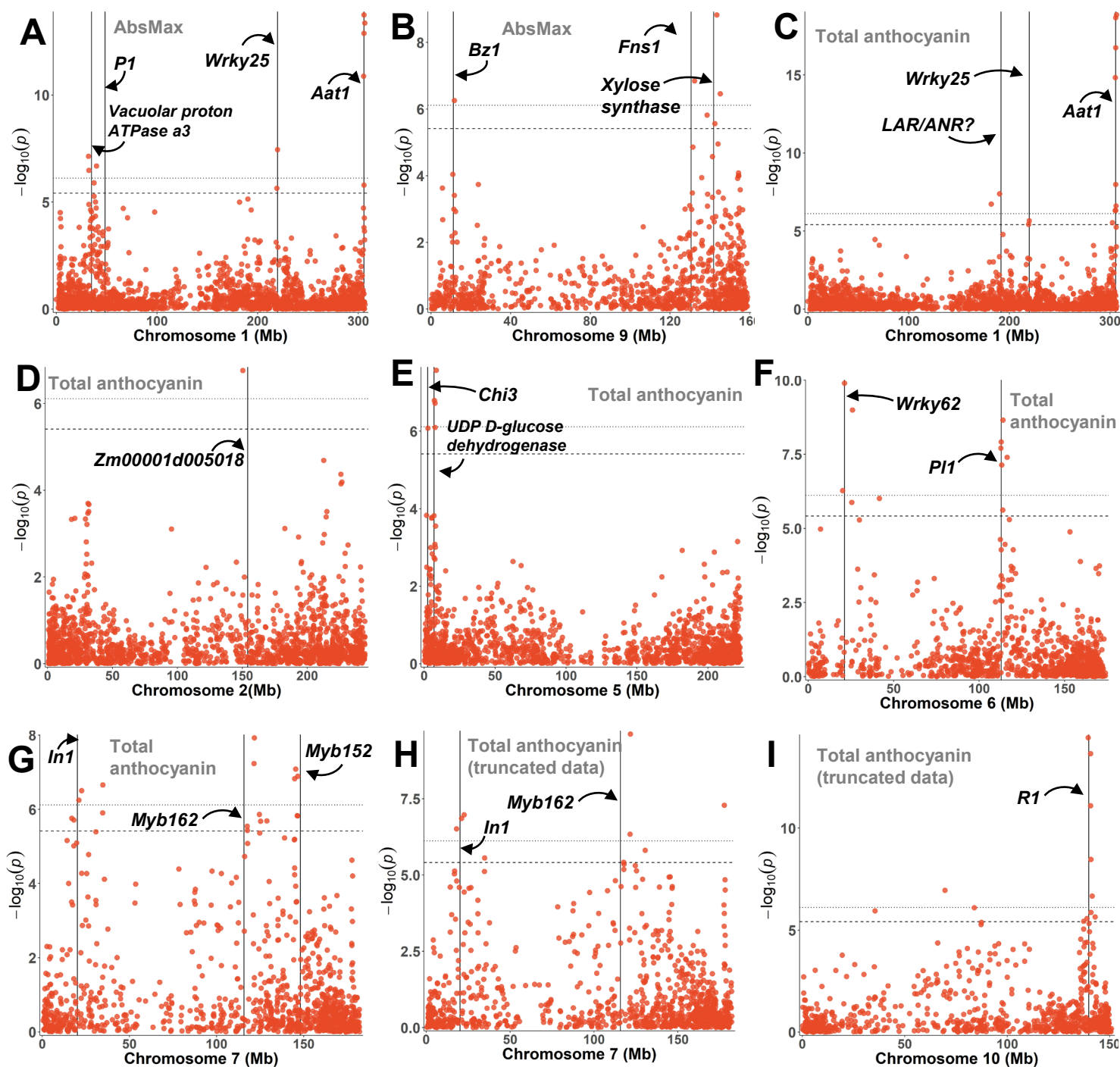
4	240.70	Zm00001d053765	vpp3 - vacuolar proton pump3
5	2.58	Zm00001d012972	chi3 , CHI (chalcone isomerase)
5	6.95	Zm00001d013245	UDP-D-glucose dehydrogenase, regulated by P1
5	21.18	Zm00001d013810	mate2, blast hit for TT12
5	27.69	Zm00001d013991	naringenin-chalcone synthase, predicted
5	49.73	Zm00001d014481	dihydroflavonol 4-reductase; predicted
5	68.03	Zm00001d014914	A2 , ANS (Anthocyanidin Synthase / Leucoanthocyanidin Dioxygenase)
5	91.46	Zm00001d015459	4-coumarate-CoA ligase, predicted
5	98.40	Zm00001d01556	vpp7 - vacuolar proton pump7
5	137.12	Zm00007a00031002	flavonol synthase (FLS1), regulated by P1
5	146.54	Zm00001d016144	chi5 , CHI (chalcone isomerase)
5	147.26	Zm00001d016151	naringenin 2-hydroxylase; predicted
5	161.97	Zm00001d016424	anthocyanidin 3-O-glucosyltransferase, predicted
5	171.35	Zm00001d016653	myb3, Mp1, pac1b, (WD repeat protein)
5	184.49	Zm00001d017077	pr1 , F3'H (flavonoid 3'-hydroxylase)
5	201.31	Zm00001d017601	rhamnose synthase gene, regulated by P1
5	201.80	Zm00001d017617	pac1, WDR (WD repeat protein)
5	206.34	Zm00001d017771	LAR / ANR candidate
5	216.31	Zm00001d018197	ppp1 - pyrophosphate-energized proton pump1)
5	217.82	Zm00001d018278	chi4 , CHI (chalcone isomerase)
6	27.97	Zm00001d035462	anthocyanidin synthase, predicted
6	60.27	Zm00001d035918	myb1
6	82.19	Zm00001d036293	AOMT ortholog
6	99.24	Zm00001d036728	blast hit for Petunia <i>PH5</i> (P-type H <sup>+</sup> ATPase)
6	112.98	Zm00001d037118	pl1, R2R3-MYB
6	123.79	Zm00001d037382	CGT (c-glucosyl transferase), regulated by P1
6	123.80	Zm00001d037383	ugt1, flavonoid 3-O-glucosyltransferase, regulated by P1
6	123.83	Zm00001d037384	flavonoid 3-O-glucosyltransferase, regulated by P1
6	127.03	Zm00001d037487	leucoanthocyanin dioxygenase, regulated by P1
6	127.21	Zm00001d037492	vpp1 - vacuolar proton pump homolog1
6	130.12	Zm00001d037576	phyrophosphate-energized vacuolar membrane proton pump
6	137.68	Zm00001d037784	Sm1, rhamnose synthase, regulated by P1
6	145.53	Zm00001d038023	wrky83 - WRKY-transcription factor 83, blast hit for <i>PH3</i> (Petunia wrky)
6	146.92	Zm00001d038073	anthocyanidin synthase, predicted
6	154.81	Zm00001d038338	myb85
6	164.06	Zm00001d038765	predicted glucosyltransferase; UDP-D-glucose 5-O-b-D-glucosyltransferase; anthocyanidin 5,3-O glycos
6	165.39	Zm00001d038833	flavonoid 3-O-glucosyltransferase, regulated by P1
7	6.96	Zm00001d018839	flavonoid 3'-hydroxylase, predicted
7	14.88	Zm00001d019062	mha3 - membrane H(+)-ATPase3, blast hit for petunia ATPase
7	20.11	Zm00001d019170	In1, bHLH-like inhibitor
7	24.21	Zm00001d019256	Ufgt2
7	76.56	Zm00001d019922	Rif1, EMSY-like / ENT domain containing protein
7	115.74	Zm00001d020457	myb162
7	125.51	Zm00001d020628	flavonoid 3'-monooxygenase; predicted
7	138.21	Zm00001d020970	LAR/ANR ortholog?
7	144.66	Zm00001d021167	Glycosyltransferase
7	144.74	Zm00001d021168	Glycosyltransferase
7	148.15	Zm00001d021296	myb152, regulation of pheyylpropanoid pathway
7	158.56	Zm00001d021629	TT12 blast match, TT12-like protein
7	161.42	Zm00001d021721	vpp8 - vacuolar proton pump8
7	169.19	Zm00001d022060	anthocyanidin reductase, predicted
7	169.51	Zm00001d022075	flavonoid 3'-monooxygenase; predicted
7	174.34	Zm00001d022278	1-Cys peroxiredoxin, regulated by P1regulated by P1
7	178.22	Zm00001d022467	Glycosyltransferase
7	181.72	Zm00001d022628	myb11, myb-related protein hv1,
8	11.88	Zm00001d008528	myb22
8	17.38	Zm00001d008699	flavonoid 3'-monooxygenase; predicted

8	25.25	Zm00001d008919	tcptf24 - TCP-transcription factor 24; blast match for Arabidopsis TCP3
8	30.34	Zm00001d009022	Medicago LAR blast match
8	34.06	Zm00001d009066	myb90
8	72.89	Zm00007a00006475	flavonoid-3'-hydroxylase, regulated by P1
8	88.82	Zm00001d009908	UDP-D-xylose synthase, regulated by P1
8	102.88	Zm00001d010201	myb25
8	112.88	Zm00001d010399	wrky92
8	119.04	Zm00001d010521	flavonoid-3'-hydroxylase, TT7 similar, , regulated by P1
8	134.04	Zm00001d010933	myb147
8	139.10	Zm00001d011107	anthocyanidin synthase, predicted
8	150.48	Zm00001d011438	A4 , DFR / F4R (bifunctional dihydroflavonol 4-reductase / flavanone 4-reductase)
8	155.95	Zm00001d011614	myb24
8	157.43	Zm00001d011645	myb10
8	171.09	Zm00001d012255	myb163, Myb-related protein 308
8	174.83	Zm00001d012456	anthocyanidin synthase, predicted
8	178.28	Zm00001d012663	VvAMT anthomate hit
9	8.98	Zm00001d044975	c1, R2R3-MYB
9	11.22	Zm00001d045055	Bz1 , UFGT (UDP-glucose flavonol glycosyltransferase),
9	13.45	Zm00001d045122	blast hit for AtTT13; plasma membrane ATPase
9	16.08	Zm00001d045206	AOMT ortholog
9	17.15	Zm00001d045254	anthocyanidin 5,3-O glucosyltransferase (ufgt3), regulated by P1
9	17.51	Zm00001d045269	mrp3, MRP (multidrug resistance-associated protein, ABC transporter)
9	26.70	Zm00001d045560	myb148
9	39.27	Zm00001d045788	naringenin dibenzoylmethane tautomer C-glucosyltransferase; predicted
9	44.29	Zm00001d045862	naringenin dibenzoylmethane tautomer C-glucosyltransferase; predicted
9	81.44	Zm00001d046336	vpp2 - vacuolar proton pump2
9	97.60	Zm00001d046591	vpp5 - vacuolar-type H <sup>+</sup> -pyrophosphatase5
9	106.80	Zm00001d046811	dihydroflavonol 4-reductase; predicted
9	108.42	Zm00001d046855	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase, regulated by P1LAR/ANR blast match
9	121.10	Zm00001d047182	myb52
9	125.38	Zm00001d047279	flavonoid 3'-monooxygenase; predicted
9	126.27	Zm00001d047309	wrky61
9	130.65	Zm00001d047452	Fns1 , F2H (flavone 2-hydroxylase),
9	136.19	Zm00001d047579	myb29
9	141.98	Zm00001d047796	rhamnose synthase gene, regulated by P1
9	142.00	Zm00001d047797	UDP-D-xylose synthase, regulated by P1
9	11.78 - 23.27	NA	Bp1, Brown pericarp1
10	2.01	Zm00001d023277	MATE blast hit
10	2.05	Zm00001d023282	myb69, blast hit for AtMYB60
10	2.21	Zm00001d023284	myb123
10	2.22	Zm00001d023287	myb107
10	9.51	Zm00001d023542	vpp4 - vacuolar proton pump4
10	14.01	Zm00001d045139	vpp6 - vacuolar proton pump6
10	14.46	Zm00001d026371	anthocyanidin reductase, predicted; LAR/ANR candidate
10	79.83	Zm00001d024596	AOMT ortholog
10	82.87	Zm00007a00022102	anthocyanidin reductase, regulated by P1
10	85.36	Zm00001d024726	myb134, blast hit for MtMYB2
10	96.10	Zm00001d024943	Flavone synthase typeII
10	96.10	Zm00001d024943	naringenin 2-hydroxylase, predicted
10	96.33	Zm00001d024946	Fnsii1 , FNS (flavone synthase); predicted naringenin 2-hydroxylase
10	139.78	Zm00001d026147	r1, hopi1, lc1, sn1, bHLH (basic helix-loop-helix)
10	140.87	Zm00001d026203	myb5, regulation of phenylpropanoid pathway
10	144.44	Zm00001d026369	anthocyanin reductase, predicted; LAR/ ANR candidate
10	144.45	Zm00001d026370	anthocyanidin reductase, predicted; LAR/ANR candidate
10	146.95	Zm00001d026490	mha4 - proton-exporting ATPase4, blast hit for petunia ATPase and AtTT13

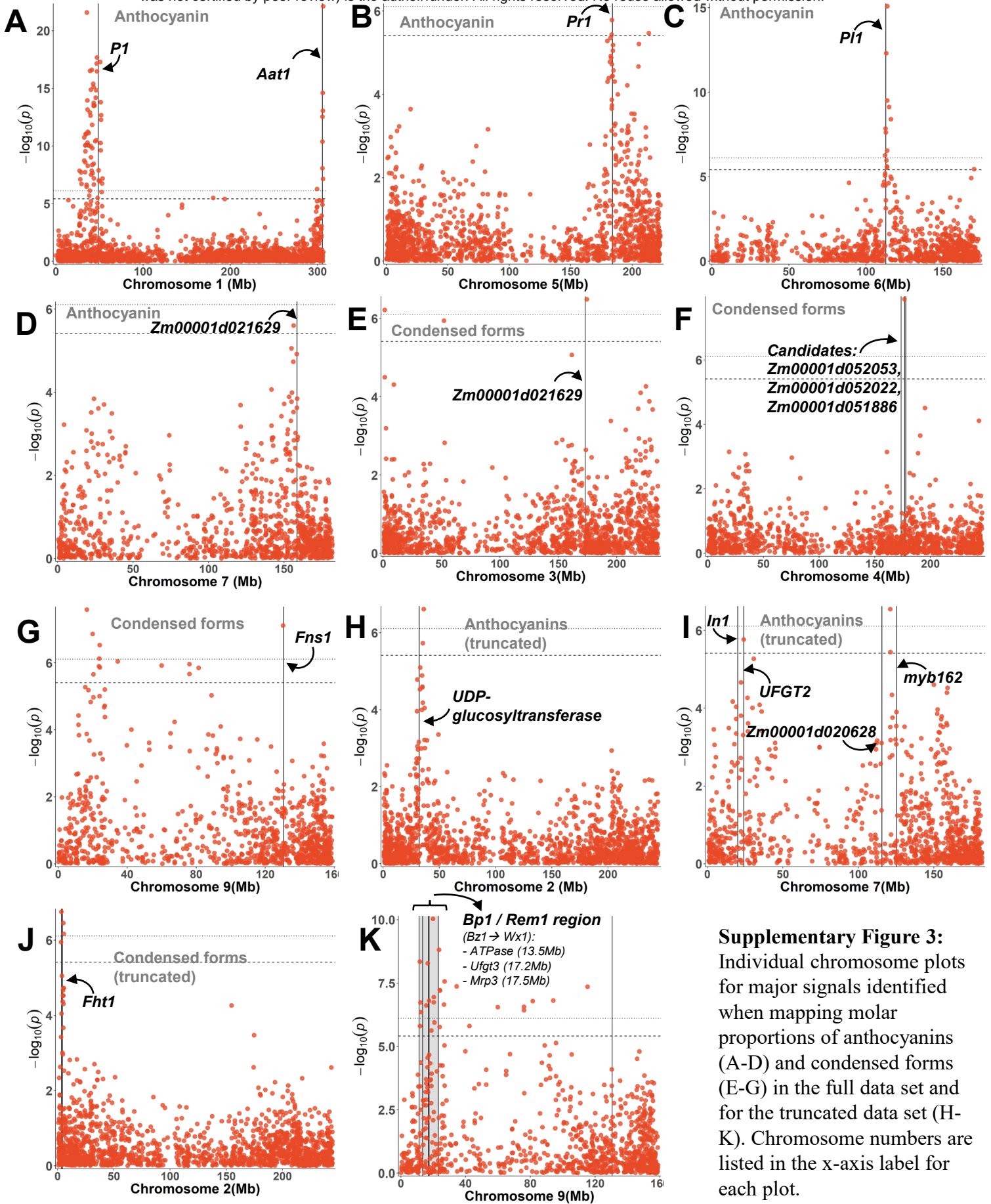




**Supplemental Figure 1:** Effects of anthocyanin composition on maximum absorbance ( $Abs_{Max}$ ), a proxy for color intensity. Anthocyanin content was square root transformed from  $\mu\text{g/ml}$ , and flavone content ( $\mu\text{g/ml}$ ) and condensed form proportion were log transformed.



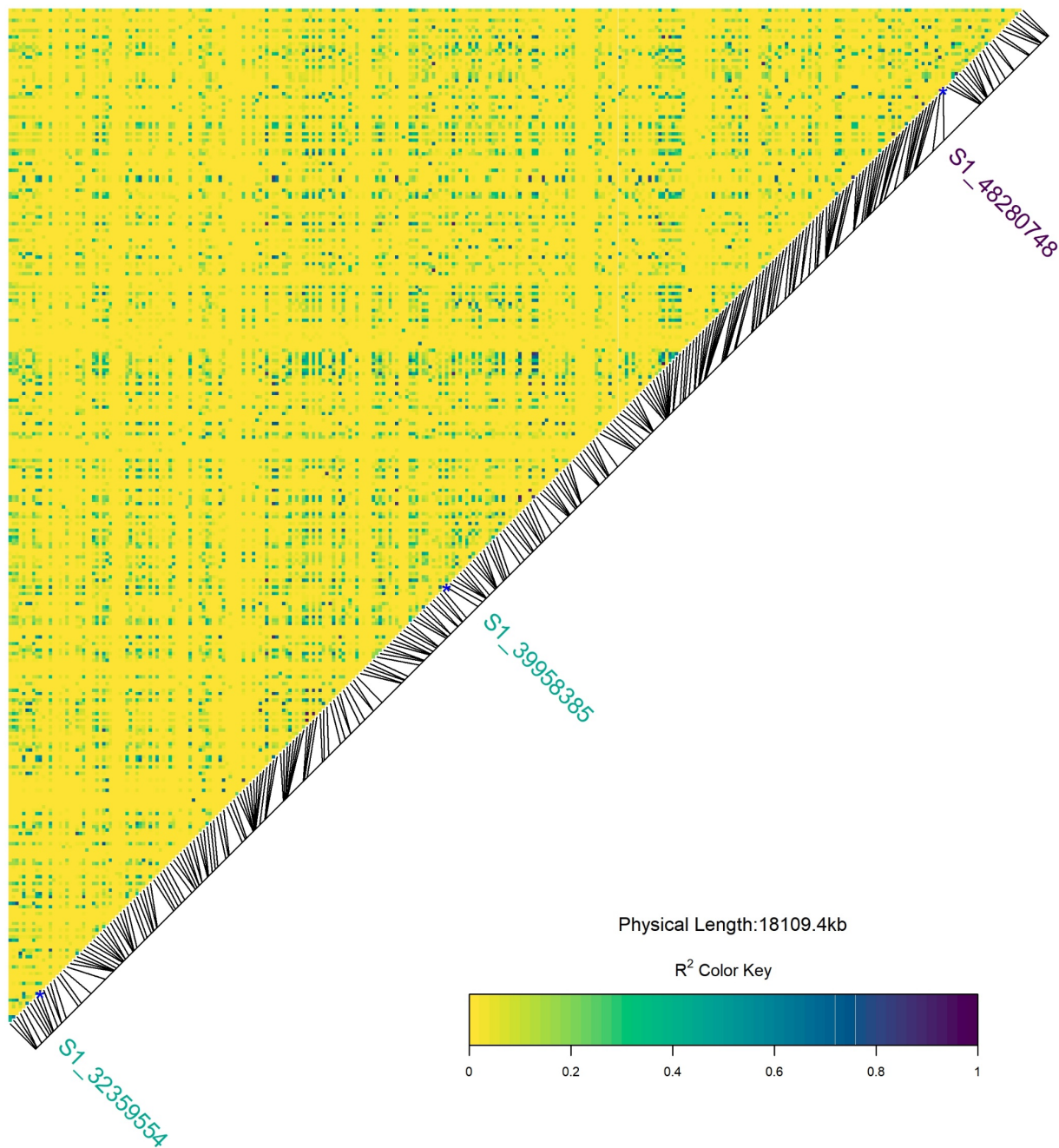
**Supplementary Figure 2:** Individual chromosome plots for major signals identified for total absorbance (A, B), anthocyanin content (C-G), and anthocyanin content for the truncated data set (H,I; tails of percent acylation distribution removed). Chromosome numbers are listed in the x-axis label for each plot.



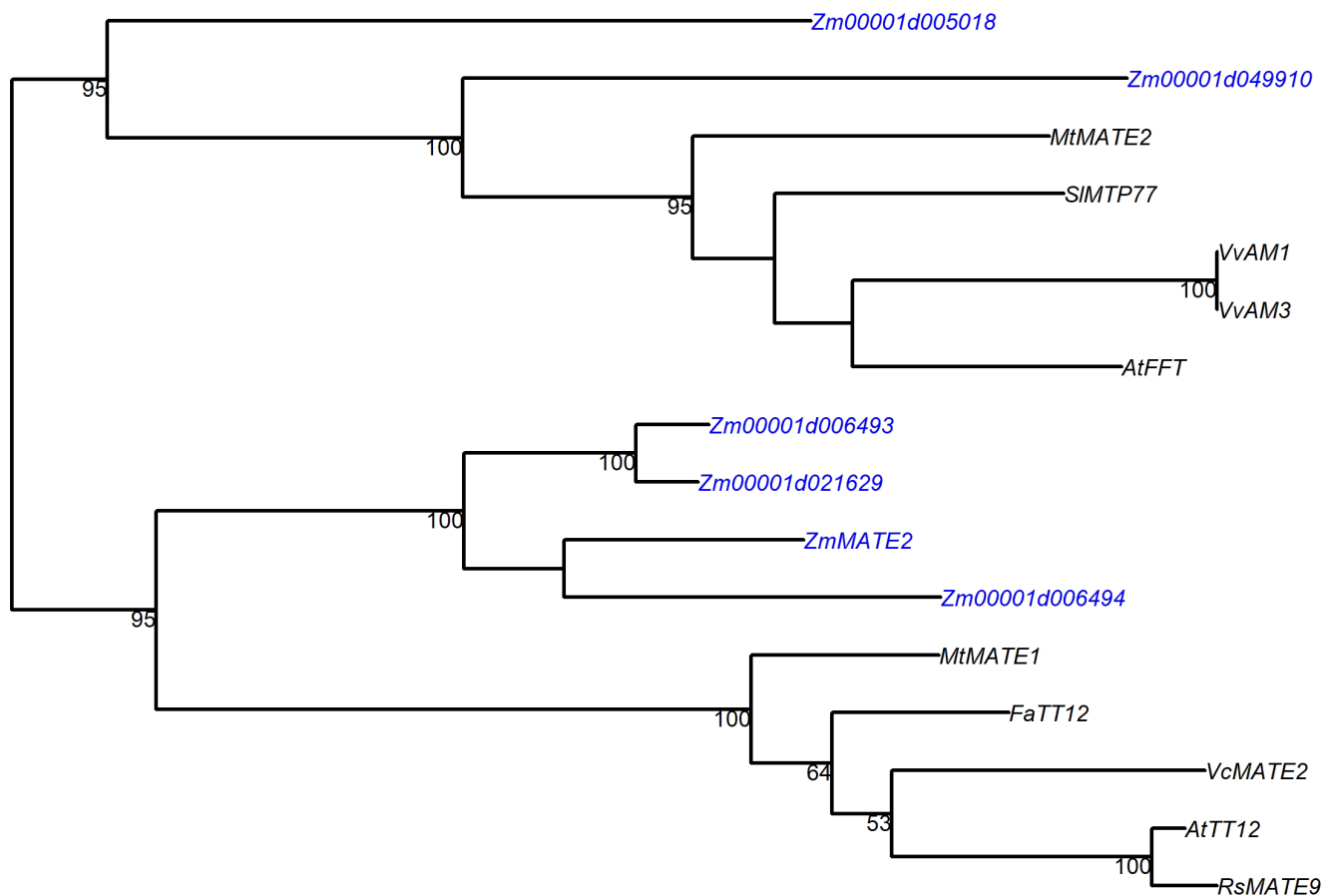
**Supplementary Figure 3:** Individual chromosome plots for major signals identified when mapping molar proportions of anthocyanins (A-D) and condensed forms (E-G) in the full data set and for the truncated data set (H-K). Chromosome numbers are listed in the x-axis label for each plot.



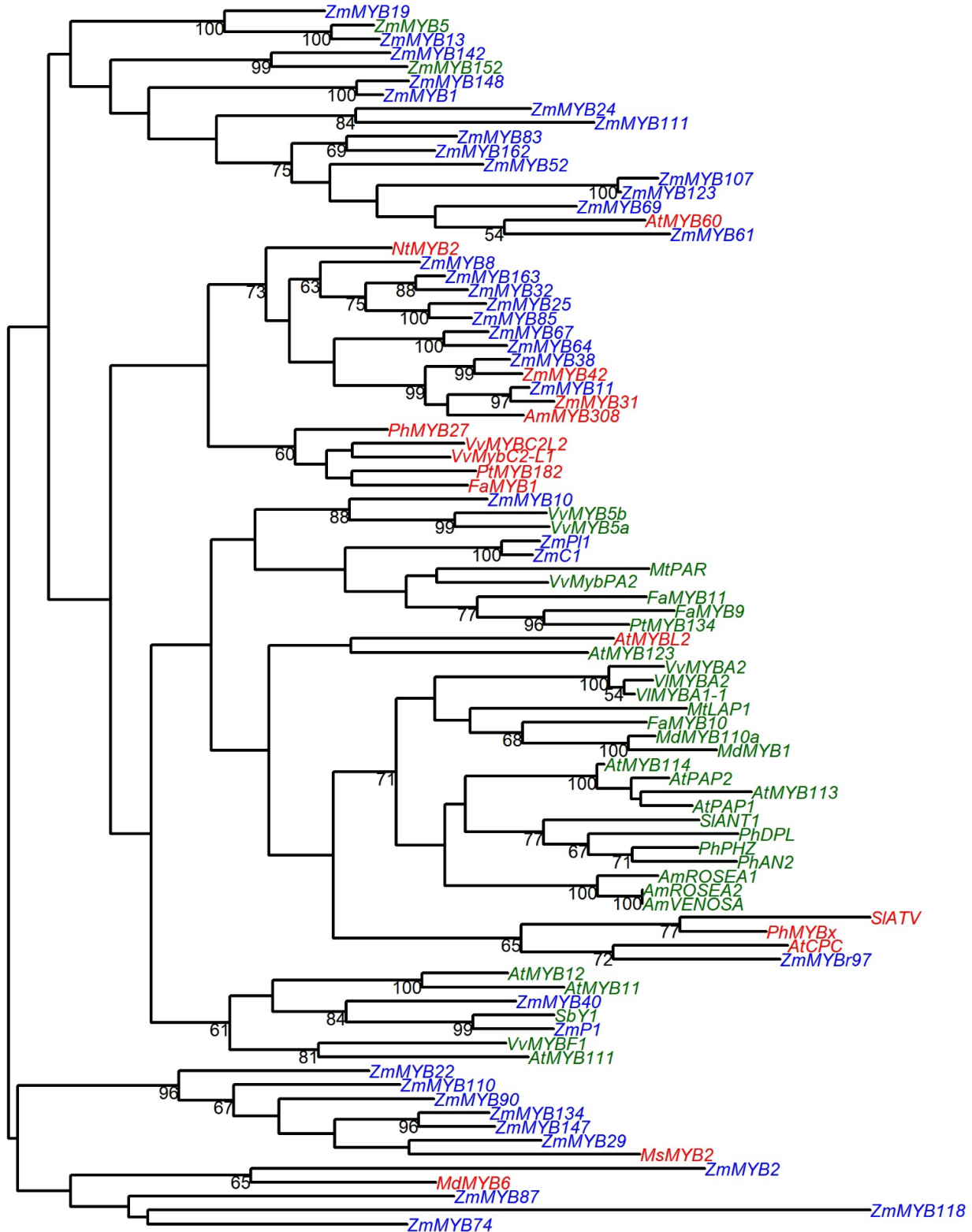
Pairwise LD



**Supplementary Figure 4:** 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 5:** Maximum likelihood phylogeny based on peptide sequences of flavonoid transporting MATEs and candidate MATEs from maize. Branch numbers represent bootstrap values based on 100 replicates.



**Supplementary Figure 6:** Phylogram of MYBs. Red labels indicate negative regulation while green labels indicate positive regulation of anthocyanin or flavonoid biosynthesis. Maize MYBs with unknown function are colored in blue. Sequences correspond to Supplementary Table 3, and branch numbers represent bootstrap values based on 100 replicates.