GENETIC FACTORS INVOLVED WITH INCREASED ANTHOCYANIN PRODUCTION IN MAIZE

BY

MICHAEL PAULSMEYER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor John Juvik, Chair
Associate Professor Anthony Studer, Director of Research
Associate Professor Martin Bohn
Assistant Professor Sarah Hind

ABSTRACT

Anthocyanins are the pigments present in nearly every plant species. These pigments have numerous roles in the plant and are associated with visual signaling and free-radicle scavenging. Anthocyanins have numerous human applications as well. The antioxidant properties of these molecules confer numerous health benefits. In addition, the orange to red to violet hues of anthocyanins makes them suitable as natural colorants. In particular, anthocyanins are capable of replacing FD&C Red No. 40, which is the most common synthetic dye on the market. Replacing synthetic dyes is important for human health and for sustainability. Interest in maize as a source of natural colorants is growing due to the abundance and diversity of pigment production and due to the economy of scale of maize production. In Chapter 1, a population with novel characteristics was developed to assess the diversity of anthocyanin composition in maize and to see the potential of anthocyanin production in tissues outside the grain. Results of this study show that maize is an abundant source of natural colorants, especially in non-grain portions of the plant, and the diversity in hues makes it suitable for many natural colorant applications. Additionally, a novel anthocyanin type was discovered in this population that broadens our knowledge of anthocyanin synthesis in maize. Chapter 2 is also focused on understanding in detail how anthocyanins are synthesized in maize. In this chapter, recombinant *Anthocyanin acyltransferase1* (*Aat1*) was isolated to determine the specificity and reaction kinetics of this enzyme. *Aat1* is one of only two characterized anthocyanin dimalonyltransferases in the plant kingdom.

Integrating regulatory factors involved with increasing anthocyanin content in maize will assist in breeding for economical natural colorants. In all plants, the activation is stimulated by the core set of transcriptional regulators referred to as the MBW complex for the M̲yb, b̲HLH, and W̲D-repeat that must physically interact for anthocyanin-related genes to be transcribed. In maize, the MBW complex is composed of multi-allelic members that coordinate spatial and temporal patterning of anthocyanin synthesis in the plant. Aleurone pigment is activated by Myb *Colored aleurone1*, bHLH *Colored1*, and WD-repeat *Pale aleurone color1*. In mainly vegetative tissues, synthesis is regulated by Myb *Purple plant1*, bHLH *Booster1* (*B1*), and a currently unknown WD-repeat protein. Additionally, there are two genes known to have a profound effect on anthocyanin synthesis. The first is *Intensifier1*, a classic negative regulator of anthocyanin

synthesis in aleurone. The second is *Anthocyanin3* (*A3*), which is a known negative regulator of pigmentation in tissues where *B1* is most active. The gene or element involved with *A3* was unknown prior to this work. In Chapter 3, two genetic approaches are utilized to determine the gene involved. The first was a large-scale transposon-tagging population utilizing a *Dissociation* element in the upstream *Anthocyanin1* gene. The second was a bulk-segregant RNA-seq population segregating for the *a3* phenotype comprised of pools of green and purple husk tissues. Together, these populations pinpointed the gene to R3-Myb-like gene *Mybr97*. Anthocyanin and phenolic content increased 100-fold in mutant husks. Transcriptomic analysis revealed that the entire phenylpropanoid pathway was upregulated in mutants and revealed novel regulators that may be involved with anthocyanin synthesis.

An additional method for enhancing anthocyanin content is through the increased yield of anthocyanin-producing tissues in the grain. Pericarp and aleurone produce most of the anthocyanins in maize kernels. In Chapter 4, two populations with contrasting pericarp and aleurone yield were developed. Both populations utilized a unique trait referred to as the multiple aleurone layer (MAL) phenotype. Two populations were genotyped using genotyping-by-sequencing. It was found that the MAL trait is conferred mostly by a locus on chromosome 8 with several other minor loci modulating the number of aleurone layers. Anthocyanin content increased nearly 30 to 40% in MAL lines indicating an increase in aleurone yield. Genetic markers on Chromosome 8 were developed to make breeding for the MAL trait easier.

Finally, the classic way to increase anthocyanin content is by utilizing genetic resources in maize and breeding for higher anthocyanin content. Chapter 5 discusses methods for breeding for higher grain yield and anthocyanin yield in temperate dent varieties. As a result of the breeding program, three near-isogenic purple corn lines in a temperate inbred background and multiple breeding populations were created. Methods for breeding purple corn and protocols for marker assisted selection are presented in this chapter. The results of this breeding program demonstrate that purple corn can match the Midwest USA average grain yield without a reduction in anthocyanin content. Altogether, maize has potential as an economical source of natural colorants and there is extensive genetic diversity of purple corn that can be utilized to make high anthocyanin-yielding varieties.

ACKNOWLEDGMENTS

There were a lot of people at the University of Illinois that ensured my success, and I would like to thank everyone involved. First off, I want to thank my advisor, Jack Juvik, for giving me this amazing opportunity to work with such a fascinating project. As evidenced by the amount of work that I have done, I really enjoy working with purple corn and I was thankful to have the independence to do my own projects. I would also like to thank my committee for being there for me when I needed advice. I would also like to thank Dr. Megan West and Dr. Les West for initiating the project and giving me early guidance in my graduate school program. Finally, I want to thank the Illinois Corn Marketing Board for providing my fellowship. I was very fortunate to have a fellowship so that I could focus on the multiple facets of my research.

Grad school would not have been easy if I had not had the backing of some very important people in my life. I want to thank my family especially. I want to thank my parents for being supportive and pushing me to always work hard even when I was discouraged. Next, I want to thank my siblings. They are a source of constant comfort in my life and I am very fortunate to call them my best friends. Finally, I want to thank Aries Loumis. I will never forget the love and support I received. He truly pushed me to be a better person and a better scientist.

TABLE OF CONTENTS

CHAPTER 1: SURVEY OF ANTHOCYANIN COMPOSITION IN VARIOUS TISSUES OF MAIZE

1.1 Abstract

Anthocyanins are a diverse class of pigments produced by numerous plant species. These pigments are natural alternatives to common synthetic dyes used for many applications. As they are natural antioxidants, anthocyanins also have many human health benefits. Maize is growing as a source of natural colorants. Much research has been done to characterize anthocyanin production in the grain, but not much is known about other tissues in the maize plant. In this study, Apache Red Cob, which has unique diversity in pigment production, was crossed to a genetic stock harboring a recessive *anthocyanin3* locus known to enhance pigment in non-grain portions of the plant. The result was diverse pigment production in portions of the plant not normally pigmented. Samples were taken from anthers, cob glumes, husks, kernels, leaf sheaths, seedlings, silks, and tassels. The diversity in pigment production was assessed using UHPLC. A novel pigment type produced in anthers was determined to be anthocyanidin 3-*O*-6"-phenylacetylglucoside as confirmed by NMR and is the first report of such a pigment. Principal component analysis and hierarchical clustering established eight clusters representing the diversity in pigment production. The utility of these tissue sources as natural colorants was evaluated using representative plant extracts from each of those clusters. Results of this study show that anthocyanins are most abundant in the non-grain portions of the plant. Maize forage and stover may be a useful source of anthocyanins for the textile, cosmetic, or nutraceutical industries, which can utilize non-grain portions of the plant. Another interesting finding was that the diversity in pigment composition may, in part, be explained by regulatory factors controlling anthocyanin synthesis. Overall, this study demonstrates that maize is an appealing source of natural colorants in terms of pigment diversity and pigment concentration.

1.2 Introduction

Anthocyanins are diverse class of secondary metabolites that provide a range of hues in numerous plant species. Anthocyanins have many important functions in plants. They provide coloration in floral organs necessary for reproductive fitness and the molecules themselves are antioxidants important for free-radical scavenging (Hatier & Gould, 2008). These compounds have many important human applications as well. The water-soluble nature and diversity in hues

1

of these pigments makes them appealing natural replacements for many synthetic dyes. FD&C Red No. 40 is the most common synthetic dye approved in the United States, with over 2.5 million kilograms being certified for use in 2020 (Center for Food Safety and Applied Nutrition, 2020). The orange to red to violet hues of anthocyanins make them suitable replacements for Red No. 40. Consumer preference is shifting towards more natural ingredients in foods and beverages. Anthocyanins could provide a more sustainable and potentially safer alternative to synthetic dyes (Amchova et al., 2015). Anthocyanins are also well-characterized health-promoting compounds. They have been demonstrated to reduce biomarkers associated with inflammation, obesity, diabetes, cardiovascular disease, and certain cancers (He & Giusti, 2010). Supplementing foods and beverages with anthocyanins has potential for the promotion of human health. Natural colorants have applications outside of foods and beverages as well. Because of the numerous health benefits, anthocyanins have possible applications as nutraceuticals. Textile and cosmetic industries could also benefit from natural colorants as waste from processing is potentially damaging to the environment (De Nisi et al., 2021).

Maize in particular is gaining interest as a source of natural colorants due to the economy of scale of production. Pigmented maize can be integrated into the already existing specialty corn supply chain with these pigments utilized as a value-added co-product. After extraction of the water-soluble anthocyanins, the remaining portions of the kernel can be utilized for food, fuel, and feed (Somavat et al., 2016). In this way, maize may be a more economical source of natural colorants than other vegetable sources. Much attention has been given to assessing the diversity of anthocyanin production in kernel tissues of maize. Not much is known about tissues aside from those typically consumed.

Most maize cultivars grown around the world display reduced anthocyanin pigmentation in most tissues. Aerial portions of the plant are typically green while the grain is yellow or orange-yellow due to the presence of carotenoids. Anthocyanin pigmentation in maize requires the coordinate action of the *Colored1/Colored aleurone1 (R1/C1)* or the *Booster1/Plant color1 (B1/Pl1)* loci. The pair of *R1/C1* typically confers aleurone pigmentation in blue corn, while *B1/Pl1* confers plant pigmentation. The exception is with *R1* locus alleles *Leaf color1* and *Hopi (R1-cherry)* which are known to act in the pericarp along with a functional *Pl1* (Ludwig & Wessler, 1990; Petroni et al., 2000). Most purple corn is pigmented by these *R1* alleles, although it has been demonstrated that *B1* can also sometimes pigment pericarp (Lago et al., 2013, 2014).

2

Additional factors are also involved with enhancing pigmentation in maize. One such factor is *Anthocyanin3* (*A3*). This gene is a recessive intensifier of plant pigmentation and works alongside *B1/Pl1* to confer intensely purple leaf sheaths, husks, and tassels (Styles & Coe, 1986). Tissues that are the normal condition of green become intensely purple as the plant matures with a recessive *a3* locus. A genetic stock recessive for *a3* was combined with purple pericarp variety Apache Red Cob (PI 213730), which has demonstrated unique diversity in pericarp anthocyanin composition (Chatham et al., 2018). Crossing these two genotypes allows for a diverse array of anthocyanins across tissue types that do not normally express anthocyanins in maize. Images of these genotypes, selected progeny, and pigmented tissues is provided in Figure 1.1. The purpose of this current study is to assess the diversity of anthocyanin content and composition in these tissues that are not normally pigmented, to assess their value as sources of natural colorants, and to characterize novel anthocyanin types.

1.3 Materials and Methods

1.3.1 Plant Materials

Apache Red Cob (PI 213730) was crossed to genetic stock 320N from the Maize Genetics Cooperation Stock Center (Urbana, IL, USA). The Apache Red Cob parent used is a source of anther, kernel, leaf sheath, seedling, and silk pigmentation, along with a recessive *Red aleurone1* (*Pr1*) allele (Chatham & Juvik, 2021). Genetic stock 320N was the source for recessive *a3* and has a null allele of *r1*. Tassel, cob, and leaf sheath pigment was intensified in 320N due to the combination of *B1* and *Pl1* with *a3*. Over the course of two years, tissues were collected from deeply pigmented individuals to assess the diversity of anthocyanin composition in these tissues. A large F_2 population (approximately 480 members) was grown in 2018 at the University of Illinois Vegetable Crops Research Farm (Champaign, IL, USA). Kernels, husks, leaf sheaths, and tassels that appeared to have the *a3* dark purple phenotype were collected from this population (Figure 1.1A, C, G, I, and J). Many of the *a3* plants from this population were selfed and grown in 2019. Anthers, silks, seedlings, cob glumes, and additional tassels were collected from this population (Figure 1.1C–F, H, I, K, and L). Purple husks, cob glumes, and kernels were collected at maturity. Purple leaf sheaths from numerous seedlings derived from the same ear selfed in 2018 were collected at the third leaf collar stage. Tassels were collected after anthesis so that anther tissue was minimized. Silks and mature leaf sheaths were collected during the silking stage. Mature leaf sheath tissue was collected from the internode under the top ear for

3

consistency. Anthers and silks were flash frozen in liquid nitrogen during collection to prevent degradation. Purple kernels, cob glumes, and husks were dried after harvesting in a forced air dryer set to approximately 40 ˚C for 1 week. Anthers, seedlings, silks, tassels, and leaf sheaths were stored at -20 ˚C before being freeze-dried for preservation. Whole cobs were ground into a fine powder using a metal file to collect richly pigmented cob glumes. For anther, tassel, and leaf sheath tissue, intense shaking or vortexing with steel balls was used to grind the tissue to a fine powder. Kernel and husk samples were ground using a coffee grinder. Husk samples were then filtered through a No. 32 mesh (500 µm) to remove large particles. Bulked anther tissue used for identification of the novel anthocyanin was collected from a purple corn variety being developed in the Juvik Lab. Seed for the variety is available for research use upon request. Bulk anther tissue from this variety was freeze-dried and ground to a fine powder using a blender before extraction.

1.3.2 Anthocyanin Extraction

The solvent chosen for anthocyanin extractions was 2% (*v/v*) formic acid. Husks, silks, tassels, cob glumes, and leaf sheaths were extracted in a ratio of 0.1 g tissue to 10 mL solvent. Seedlings and anthers were extracted in a ratio of 50 mg tissue to 1.5 mL solvent. Kernels were extracted in a ratio of 2 g tissue to 10 mL solvent. All samples were extracted for 1 hour at 50 ˚C with mixing. All samples were centrifuged and filtered to 0.45 µm before injection into the UHPLC. Anthers for identification of the novel anthocyanin were extracted in a ratio of 139 g anther tissue to 3 L solvent. Extract after 1 hr of incubation at 50 ˚C was centrifuged in a 250 mL conical centrifuge tube and passed through a Whatman No. 2 filter. Anther tissue captured in the conical tubes was resuspended in solvent. The anthers were extracted a total of three times by filling the volume back to 3 L. Crude extract was then processed through the anthocyanin purification protocol below.

1.3.3 Anthocyanin Purification

Anthocyanins were purified from crude extracts for NMR or mass spectrometry using the following procedure. First, crude extract was passed through Amberlite FPX-66 resin (Dupont) that was washed with ethanol and equilibrated with water. A ratio of 2 parts extract to 1 part resin was sufficient to capture all the phenolics. The resin was washed with 20 resin-bed volumes water for mass spectrometry samples and 6 resin-bed volumes water for the anther extract. Three resin-bed volumes of 95% ethanol were sufficient to elute semi-pure anthocyanins. This

4

ethanolic extract was freeze dried for mass spectrometry samples and resuspended in 0.1% HCl in methanol for analysis. For NMR samples, ethanol was evaporated using a rotary evaporator (Brücher) until it was a slurry. This extract was purified further with 6 washes of ethyl acetate in a 1:1 ratio of slurry to ethyl acetate. The semi-pure anthocyanin slurry was freeze-dried to a powder. This powder was resuspended in a ratio of 100 mg semi-pure anthocyanins to 1 mL 0.1% HCl in water and filtered using a 0.45 μm syringe filter. The extract was passed through an Agilent 1100 semi-preparative HPLC equipped with a Grace Prevail C_{18} (4.6 mm x 250 mm, 5 μm) column (Alltech Associates, Inc) maintained at 30 ˚C. The mobile phase was 2% formic acid (A) and acetonitrile (B) with a gradient of 10% B to 22.5% B in 5 min, 22.5% B to 26.5% B in 20 min, 26.5% to 100% B in 30 s, 100% B hold for 1 min, and back to 10% B in 30 s. The column was equilibrated for 6 min between samples. The detector was monitored at 520 nm and 280 nm. Fractions containing the novel anther anthocyanin were collected and concentrated with centrifugation under a vacuum. Due to the length of time between fractionation and pooling samples, a second round of stringent purification was run on concentrated extract. Only pure fractions (greater than 94% pure) as assessed by the UHPLC method below at a wavelength of 280 nm were collected in the second round. These extracts were concentrated once again and freeze-dried for submission for NMR.

1.3.4 Liquid Chromatography and Mass Spectrometry

Samples were analyzed on an Agilent 1290 Infinity II UHPLC with a Poroshell 120 SB-C_{18} (3 mm x 100 mm, 1.9 μm) column (Agilent Technologies) maintained at 50 ˚C. The mobile phase consisted of 5% formic acid (A) and acetonitrile (B) at a flow rate of 1.7 mL min^{-1} in a gradient of 4% B for 1 min, 4% to 14% B in 10 min, 14% to 25% in 3 min, 25% to 100% B in 30 s, and back to 4% B in 30 s. The column was equilibrated for 3 min between each sample. The detector was monitored at 520 nm and 280 nm. The UHPLC was calibrated using a linear gradient of 0.5 to 100 μg mL^{-1} for cyanidin 3-glucoside, pelargonidin 3-glucoside, and peonidin 3-glucoside standards dissolved in 0.1% HCl in water (Extrasynthese). Compounds were identified using liquid chromatography-tandem mass spectrometry (LC-MS/MS) at the Mass Spectrometry Lab at the University of Illinois (Urbana, IL, USA). Samples for compounds identification consisted of pools of representative tissue extracts. A Waters Synapt G2-Si ESI/LC-MS/MS was equipped with a Zorbax Eclipse XDB-C_{18} (4.6 mm x 150 mm, 5 μM) column (Agilent Technologies) kept at 50 ˚C. The mobile phase consisted of 0.2% formic acid

(A) and acetonitrile (B) at a flow rate of 0.7 mL min^{-1} in a gradient of 4% B to 25% B in 58 min, 25% to 100% B in 30 s, 100% B for 1 min, and back to 4% B in 30 s. The column was equilibrated for 10 min between each sample. Aliquots of 1 µL were used for analysis. Mass spectrometry data was interpreted in SeeMS (https://proteowizard.sourceforge.io/) (Chambers et al., 2012). A full list of compounds identified in this study is in Table 1.1.

1.3.5 NMR

NMR was conducted at the USDA National Center for Agricultural Utilization Research in Peoria, IL, USA. All NMR spectra were collected with a Bruker Avance 500 spectrometer (500.11 MHz) using a 5 mm BBO probe at 300.0 K, and a JEOL ECZ 600R spectrometer (600.17 MHz) with a 5 mm BBO probe at 298 K. The sample was dissolved in 95/5 methanol-d4/trifluoroacetic acid-d. Chemical shifts are reported as ppm from tetramethylsilane calculated from the lock signal. NMR predictions were made using ACD/C+H NMR Predictor DB v2017.1 (Advanced Chemistry Development, Inc., Toronto, ON, Canada).

1.3.6 Tristimulus colorimetric measurements

Samples used for tristimulus colorimetric measurements were determined by cluster analysis. Samples lying near the centroids of their clusters were used to represent their clusters. Extracts were combined and concentrated under a vacuum with centrifugation. Extracts were run on the UHPLC to determine anthocyanin content. Samples were diluted using either 0.1 M KCl (pH 1.0), 0.1 M citrate buffer (pH 3.0), or 0.1 M citrate buffer (pH 6.0). The pH was adjusted to 1.0, 3.0, or 6.0 using the appropriate buffer and 6N HCl or 2N NaOH to a final concentration of 50 µg mL^{-1}. Red No. 40 (Sigma) was dissolved in water and diluted to 50 µg mL^{-1} using the appropriate buffer. Pelargonidin 3-glucoside and cyanidin 3-glucoside standards (Extrasynthese) were dissolved in 0.1% HCl in water and also diluted to 50 µg mL^{-1} using the appropriate buffer. Tristimulus colorimetric measurements were conducted the same as previously described (Luna-Vital et al., 2017).

1.3.7 Statistics and Data Analysis

A major goal of this study was to assess the compositional diversity among tissue types and between samples within tissue types. For this reason, all data analyses were performed on the transformed UHPLC dataset where integrated peak areas of compounds were converted into percent of total peak area. Percent flavonol-anthocyanin condensed forms were calculated by summing the peak areas of all anthocyanins before cyanidin 3-glucoside by retention time.

6

Percent acylated anthocyanins were calculated by summing peak areas after the retention time of peonidin 3-glucoside. Principal component analysis (PCA) was calculated using the "princomp" function in R with the correlation matrix was used to calculate eigenvectors (R Core Team, 2019). Visualization of PCA was done using the "ggbiplot" (http://github.com/vqv/ggbiplot) package. Hierarchical cluster analysis used Ward's Minimum Variance Method (R Core Team, 2019). The number of significant clusters were chosen based on dendrogram branch length and also based on meaningful variation within the clusters. Many compounds were unknown and needed to be calibrated with the proper standard. The calibration standard of unknown anthocyanins was determined using cluster analysis on the correlation matrix of the transposed dataset and Ward's Minimum Variance Method. Three clusters were significant in this dataset and could differentiate pelargonidin forms from the cyanidin/peonidin forms of anthocyanins. Linear discriminant analysis was calculated using "lda" in R with defaults, but including the option for the leave-one-out cross-validation method (Venables & Ripley, 2002). Tukey's Honest Significant Differences were calculated using the "HSD.test" function (Mendiburu, 2020).

1.4 Results and Discussions

1.4.1 Total Anthocyanin Content Among Tissue Types

Intensely pigmented tissues were assayed across two summer seasons of samples to determine the concentration of anthocyanins in various tissues of maize. Previous studies demonstrate anthocyanin concentration is highly heritable and mainly related to genetic factors (Chapter 4; Chapter 5; Nankar et al., 2016; Paulsmeyer et al., 2017), but it is possible that concentration could change across years. Despite this, there were general trends that could be determined. Cob glumes had the highest maximum anthocyanin content of all tissues assayed, while leaf sheaths had the statistically highest average anthocyanin content (Table 1.2). This is in contrast with an earlier study that analyzed husk and cob tissue and found that husks had the highest anthocyanin content (Khamphasan et al., 2018). The current study did not analyze a whole cob sample, however. Therefore, total anthocyanins will be inflated due to the unpigmented cob pith not being sampled. Seedling tissue anthocyanin content may also be inflated since the whole plant was not sampled and only pigmented leaves. All tissues sampled had anthocyanin contents 30- to 100-fold the concentration of that found in whole kernel samples. One reason is that anthocyanin content of kernels in this study was less due to the

7

genetic background of the population. A previous study analyzing a large population of Apache Red Cob with the same extraction protocol found a maximum anthocyanin production of 1598 mg/kg in the whole kernel extract (Chatham & Juvik, 2021). Another reason kernel anthocyanin content was low compared to other tissue types was because the kernel consists mostly of the embryo and starch fractions by weight which are unpigmented by anthocyanins. If kernels have their pericarps removed either through the dry-grind or dry-milling process, a pericarp-enriched fraction would be anywhere from 10 to 20 times more concentrated given that a kernel is typically around 5 to 10% pericarp by weight (Somavat et al., 2016). Anthocyanin content of husks was also lower in this study than what has been observed in other genotypes in maize. Although it is hard to compare across the literature due to differences in extraction methods, husks have the potential to reach up to 19% anthocyanins by dry weight (Li et al., 2008). It can be expected that anthocyanin content can be improved in each tissue sampled by breeding given the high variability among the samples. Moreover, the extraction procedure reported here was by no means exhaustive. Organic solvents extract anthocyanins much more efficiently and much of the anthocyanins are still bound after the first extraction (Chatham et al., 2018). Therefore, the relative values for each tissue type are reported here instead of total anthocyanin content. For industries utilizing non-vegetable portions of the maize plant, forage and stover should provide more pigment per dry weight than purple kernels.

1.4.2 Compositional Diversity Among Tissue Types

The majority of the anthocyanins identified in this study have been previously described in maize. The most abundant compounds identified were the glucoside, malonylglucoside, and dimalonylglucoside forms of pelargonidin and cyanidin (Table 1.2). On average, these compounds averaged 80.4% of total anthocyanins based on integrated peak area. Generally, the remaining uncharacterized compounds were not novel compounds, but rather isoforms of known compounds such as hexose glycosides with or without malonic acid groups. Peonidin appears to be a minor component in most plant tissues. Kernels had the highest proportion of peonidin 3-glucoside of total anthocyanins, but the total peonidin content was around 6% on average. Some studies have detected minimal amounts of delphinidin anthocyanidins (Liu et al., 2018; Žilić et al., 2012). Maize has a predicted flavonoid 3',5'-hydroxylase that is expressed in various maize tissues (Liu et al., 2018; Stelpflug et al., 2016). Delphinidin or related petunidin and malvidin compounds were not detected in any tissue sampled in this study. One striking finding of this

8

study was the high proportion of cyanidin 3-dimalonylglucoside in tissues outside of husks, kernels, and cob glumes. In seedling and silk tissues specifically, this compound averaged nearly 55% of the total anthocyanins by peak area. These tissues had a less complex mixture of compounds than other tissue types.

Remaining compounds that could not be characterized were in minor quantities (less than 5% of total peak area on average). Those compounds that could not be determined were subject to hierarchical cluster analysis. In this analysis, three major clusters emerged. All the known pelargonidin anthocyanins clustered together. The second and third clusters contained a mixture of cyanidin and peonidin forms with no clear differentiation between the two anthocyanidins. Unknown compounds clustering with pelargonidin were quantified using pelargonidin 3-glucoside as a standard. The remaining unknown compounds were quantified using cyanidin 3-glucoside as a standard since their identities could not be definitively determined. Peonidin was a minor component in all tissues, so cyanidin is an appropriate approximation for all unknown peaks.

Acylation is an important modification for anthocyanins as they have interactions with stability and increasing total anthocyanins (Zhao et al., 2017). The most common acylation group was with aliphatic malonic acid. Acylation with succinic acid is not prevalent in maize due to the reduced affinity for this substrate with Anthocyanin acyltransferase1 (Aat1) (Paulsmeyer & Juvik, 2021). The ion with a 549 m/z and a fragment of 287 m/z consistent with cyanidin 3-succinylglucoside was more abundant in all tissues than the ion for peonidin 3-malonylglucoside, which is the same molecular weight. In tassels, seedlings, and silks, the ion for peonidin 3-malonylglucoside was not detected by LC-MS/MS. A minor compound with a 595 m/z ion and a 287 m/z fragment consistent with cyanidin 3-coumaroylglucoside was detected by LC-MS/MS. This compound was detected in another study as well (Bhatla & Pant, 1977). Aromatic acylation is not common in maize as Aat1 is only known to utilize aliphatic acids as substrates (Paulsmeyer & Juvik, 2021). Overall, tissue types differed greatly in the total amounts of acylated anthocyanins. Kernels had the least amount of acylated anthocyanins followed by husks and cob glumes (Table 1.2). One kernel sample had only 9.54% acylation, indicating that Aat1 or malonate anthocyanin transporters are spontaneously silenced in this population (Paulsmeyer et al., 2018; Zhao et al., 2011). Silencing is not due to variation at the Aat1 locus since only 7 samples of 180 contained less than 30% acylation. Paramutation has been demonstrated in maize

9

transporter *Multi-drug resistance-associated protein4* (*Mrpa4*) and affects phytic acid levels (Pilu et al., 2009). The mechanism of silencing is not understood for *Mrpa4*, but it is paramutagenic. Future studies will need to investigate whether this reduced acylation locus is also paramutagenic. Remaining tissues aside from kernels, husks, and cob glumes had an overwhelming majority of acylated anthocyanins and ranged from 79.96 to 86.49% acylation on average (Table 1.2).

Purple pericarp parent Apache Red Cob is known for producing considerable levels of flavanol-anthocyanin condensed forms (Chatham et al., 2018). Condensed forms were detected in every tissue analyzed with kernels having the highest amount on average. All other tissues averaged less than 0.5% total condensed forms (Table 1.2). One kernel sample produced 44.70% condensed forms of total anthocyanins. There were no significant differences in condensed form formation among tissues aside from kernel samples.

1.4.3 NMR identification of a novel pigment abundant in anthers

Anther tissue abundantly produced unknown compounds that elute after all known acylated compounds in maize. Mass spectrometry detected two compounds with a molecular weight of 567 *m/z* and 551 *m/z*. The fragment ions of these two compounds were 287 and 271 *m/z* corresponding to cyanidin and pelargonidin, respectively. No compound with a molecular weight of 581 *m/z* and a fragment ion of 301 m/z corresponding with peonidin was detected using mass spectrometry since peonidin is a minor component in anthers. No known anthocyanin compounds described to this date have a molecular weight corresponding to these three compounds. NMR has determined that the novel anthocyanin abundant in anthers is anthocyanidin 3-6"-phenylacetylglucoside (Figure 1.2). Predicted NMR spectra aligned well with the measured spectra indicating the structure was a good fit (Table 1.3). This novel anthocyanin is unique in that it is conjugated to an auxin compound widely distributed among all plant species (Sugawara et al., 2015). The auxin group on the anthocyanin molecule is unstable in solutions and easily reverts to the anthocyanidin 3-glucoside (data not shown). The aromatic breakdown product gives the anthocyanin a "sweet-like" off-odor that is characteristic of phenylacetic acid (Campo et al., 2012). The presence of this compound in anthers and the detection of cyanidin 3-coumaroylglucoside indicates that a secondary anthocyanin acyltransferase may be present in maize that can acylate with aromatic acyl groups. Aat1 and other dimalonyltransferase enzymes are not known to utilize aromatic acyl groups (Paulsmeyer

10

& Juvik, 2021; Suzuki et al., 2004). Future work should test if Aat1 is capable of utilizing aromatic acyl groups or if there is an alternative aromatic acyltransferase in maize.

1.4.4 Principal Component Analysis (PCA) finds compositional variation

Diversity in anthocyanin composition was assessed in deeply pigmented tissues across two years. All tissues aside from tassels were collected in mutually exclusive years, but anthocyanin composition is not expected to change much across years as it is a very heritable trait and controlled by genetic factors which will be mentioned later (Chapter 4; Chapter 5; Nankar et al., 2016; Paulsmeyer et al., 2017). Principal Component Analysis (PCA) could explain a large proportion of the variation. Using the covariance matrix method, PCA could explain 81.5% of the variation using two PCs. This variation was largely attributed to variation in the maize flavonoid 3'-hydroxylase (F3'H) locus, *Pr1*. Apache Red Cob segregates for a variant of *pr1* that is unable to synthesize cyanidin anthocyanins (Chatham & Juvik, 2021). Parent 320N has a functional *Pr1* since it synthesized mainly cyanidin forms of anthocyanins. While this finding is substantial, it is also known that this gene has a large effect on anthocyanin composition because it is a biosynthetic gene.

Using the correlation matrix, more finer details could be discerned that differentiated tissue types. The correlation matrix is an appropriate method to calculate PCs since percentage data is normalized and not influenced by the magnitude of variables. PCA with the correlation matrix found four significant components explaining 60.7% of the variance within the data. The first two components are shown in Figure 1.3. The first PC explained 27.4% of the variance and is largely a contrast between cyanidin and pelargonidin forms of anthocyanins. High scores in this PC indicate higher proportions of pelargonidin. This PC is explained by variation in *pr1* as was the first PC using the covariance matrix. The second PC explains 17.5% of the variation and is distinguished by the ability to synthesize flavanol-anthocyanin condensed forms and proportions of cyanidin 3-dimalonylglucoside. Seedlings, anthers, leaf sheaths, silks, and tassels generally produced a high proportion of cyanidin 3-dimalonylglucoside, while the others did not. In addition, non-modified glucosides of peonidin and cyanidin had high loadings in the second PC, indicating *Aat1* activity may explain compositional variation in some tissues. Three major distributions emerge in the second PC. The first distribution groups those individuals that abundantly produce cyanidin 3-dimalonylglucoside or pelargonidin 3-dimalonylglucoside. The second distribution was centered around zero and included husks, cob glumes, and *pr1* kernels.

11

These samples had less peonidin and cyanidin 3-glucoside anthocyanins. Finally, the third distribution within PC2 is *Pr1* kernels with high levels of condensed forms and cyanidin/peonidin 3-glucoside. The third PC only explained 8.6% of the variation. Peonidin 3-glucoside and acylated cyanidins were differentiated from the ability to make condensed forms. High values in this PC indicated an abundance of condensed forms. Finally, the last component with 7.2% variation discriminated between pelargonidin 3-6"-phenylacetylglucoside and pelargonidin 3-dimalonylglucoside from most of the minor anthocyanin types like cyanidin 3-malonylhexoside isoforms.

1.4.5 Hierarchical cluster analysis links anthocyanin regulation to pigment composition

In order to cluster similarly linked profiles based on composition and to assess the diversity in anthocyanin production in maize tissues, hierarchical clustering was performed. The most parsimonious number of clusters that minimized variance within clusters was with three clusters. The first cluster was all the samples affected by the *pr1* recessive variant. It appears that not every tissue is affected the same from this locus. Seedling, silks, and tassels had no representation in this cluster. Only four of the tassel samples had pelargonidin forms greater than 50% of total anthocyanins. Silks and seedlings only had 25 and 39 individuals, respectively, so it is possible that no *pr1* recessive alleles were sampled. However, a few seedling samples were taken from known recessive *pr1* individuals based on the previous year's phenotype and resulted in normal cyanidin levels. Moreover, very rarely did pelargonidin forms of anthocyanins comprise all the anthocyanins within a sample. This suggests that there are additional F3'H loci with varying expression in tissues. The second and third clusters determined by hierarchical clustering differentiate the kernels, husks, and cob glumes from the remaining tissues as was the case with the second PC described above. In general, the proportion of cyanidin anthocyanins determined the split between these clusters. Kernels, husks, and cob glumes produce more cyanidin 3-glucoside and cyanidin 3-malonylglucoside. The similarity in compositions may indicate a broader regulatory role.

Variation within the three clusters was minimized; however, there was a large amount of meaningful variation in anthocyanin composition within these clusters. To explain more of this variation, additional clusters were formed based on branch length separation in the dendrogram and based on pigment profiles contained within these clusters. Eight clusters were chosen to represent the total variability in anthocyanin profiles. Tissue types did not separate cleanly into

12

the clusters, but the trends matched the PCA clustering (Figure 1.3). The representation of each tissue type within the clusters is shown in Table 1.4. Each cluster had signature compounds that defined the compositional profile for the group. Chromatograms for each cluster are in Figure 1.4. A visual summary of the average proportion of each quantified compound within the clusters is in Figure 1.5. Cluster 1 was a mixture of mainly tassel, anther, and leaf sheath tissues. This cluster had the highest levels of cyanidin 3-6"-phenylacetylglucoside and a substantial amount of cyanidin 3-malonylglucoside relative to the other clusters. Cluster 2 was very similar to Cluster 1 in that it contains many tassel, anther, and leaf sheath tissue samples; however, it appears Cluster 2 is an intermediate or weak *Pr1* class. This cluster had on average 37.4% pelargonidin forms of anthocyanins, especially relatively high amounts of phenylacetylglucosides and pelargonidin 3-dimalonylglucoside. The anther, silk, tassel, and leaf sheath samples within this cluster may be *pr1* recessive, but the additional F3'H loci in maize makes the phenotype hard to score. Cluster 3 included all seedling samples, nearly all silk samples, and some anther samples. This cluster was differentiated by very high levels of cyanidin 3-dimalonylglucoside. Clusters 4 through 7 contained the husk, kernel, and cob glume samples. A majority of husks and cob glumes were included with Cluster 4. This cluster was noteworthy because it had the highest level of cyanidin 3-malonylglucoside and an appreciable amount of peonidin 3-glucoside. Cluster 5 was the most unique cluster and contained a majority of the kernel samples. It had the highest amount of cyanidin 3-glucoside, peonidin 3-glucoside, and condensed forms. Some husk and cob glume samples incidentally fit into this cluster. Cluster 6 was a *pr1* class for husks and cobs and contained on average 66.0% pelargonidin forms of anthocyanins. This cluster was distinctive in that it had the highest proportion of pelargonidin 3-malonylglucoside of total anthocyanins. The total proportion of cyanidin and peonidin forms in this cluster ranged from 13 to 55% giving further evidence that additional F3'H genes are active in samples from Cluster 2 and Cluster 6. Cluster 7 and Cluster 8 were typical *pr1* recessive clusters with 86.7% and 90.9% pelargonidin forms of total anthocyanins, respectively. Cluster 8 only contained anther samples, while Cluster 7 was mainly kernel samples with a couple cob samples. The differentiation between the two clusters was the high proportion of pelargonidin 3-6"-phenylacetylglucoside and pelargonidin 3-dimalonylglucoside in anthers and condensed forms and pelargonidin 3-glucoside in Cluster 7. A unique finding of this study was the diversity of anthocyanin production within the female tissues sampled. The hypothesis was that all female tissues were going to be similar

13

compositionally since they are part of the same organ. However, it appears that regulatory differences among the tissues have an effect in altering pigment profiles. It is unknown which allele of *R1* is conferring pericarp pigmentation in this population (Chatham & Juvik, 2021). It is known that *Hopi* or *Lc1* and genes together confer kernel, cob, leaf sheath, and husk anthocyanins to varying degrees, but the pigment intensity outside the grain may be impeded by *A3* (Ludwig & Wessler, 1990; Petroni et al., 2000). The *B1* allele in 320N can confer husk and cob pigmentation as well, but not kernel pigmentation. Husks and cob glumes were clustered distinctly from kernel samples with a small degree of overlap. It could be that the differentiation between Cluster 4 and Cluster 5 might be the additional activation of biosynthetic genes by *B1*. Acylated anthocyanins were much at a higher proportion in husks and cob glumes than kernels, but less so in other tissues where *B1* is dominant (Table 1.2). Furthermore, Cluster 2 and Cluster 6, which are pelargonidin dominant clusters affected by *B1*/*Pl1*, had higher proportions of cyanidin than true *pr1* recessive clusters like Cluster 7 and Cluster 8, which were 86.7 to 90.9% pelargonidin forms on average. It may be that *B1* is capable of activating F3'H genes in these tissues, but *R1* is not. Furthermore, condensed form pigments were much more abundant in kernels (Table 1.2), which this was one major differentiation between Cluster 4 and Cluster 5. Factors within pericarp might be driving these compositional differences. A gene that might be involved with this difference is *Pericarp color1* (*P1*), which is known to act in pericarp and is linked with condensed form formation (Chatham & Juvik, 2021). Moreover, silk tissue was very distinct from any female tissue sampled. Studies have shown that certain *R1* alleles separate from *Lc1* and *Hopi1* can confer silk, seedling, and anther pigmentation (Lago, 2014; Ludwig & Wessler, 1990). These alleles are common in modern maize cultivars as the reference genome B73 contains this allele of *R1* (Andorf et al., 2010). *Pl1* is also capable of activating pigment synthesis in anthers or silks (Patterson et al., 1993; Yonemaru et al., 2018). The allele of in 320N does not pigment anthers or silks, however, so alleles of *Pl1* and/or *R1* in Apache Red Cob may equally be responsible for pigmenting these tissues. Many anther samples were split between weak *pr1* Cluster 2 and strong *pr1* Cluster 8 (Table 1.4). This may be due to a dosage effect of *pr1* alleles in anthers or due to the alternate expression of alleles of *R1* and in these samples. Additionally, two silk samples clustered with anthers, tassels, and leaf sheaths (Table 1.4), giving more evidence that activation by may be possible in anthers and silks. Future studies should expand the representation of silk samples to see where they cluster, and which genes are

14

responsible for their anthocyanin expression. Overall, results from hierarchical cluster analysis demonstrate that anthocyanin profiles may assist in inferring anthocyanin regulatory networks.

1.4.6 Anthocyanin composition and its relationship to color

The application of maize tissues as natural colorants will be ultimately decided by the hue and stability of the matrix in which they are placed. Extracts from several individuals representing their clusters as determined above were combined to calculate tristimulus color values at a pH of 1, 3, and 6. The tristimulus color results show how the diversity in hues expected from these pigment sources (Figure 1.6). Tristimulus color readings changed dramatically from pH 1 to pH 6 as was expected. The change in color (ΔE) from a pH of 1 to pH 3 was on average 19.7 and from a pH 1 to pH 6 was 67.3. A difference of 2 or greater in ΔE is a noticeable change for most observers and a value of 3 to 6 is acceptable reproducibility (Mokrzycki & Tatol, 2011). Anthocyanins are most stable at a pH of 1 to 2 and then change to colorless chalcone or blue/violet quinonoidal bases between a pH of 3 to 6 (He & Giusti, 2010). The CIELAB b* value decreased at pH increased indicating a bathochromic shift due to blue/violet quinonoidal base formation and the L* value increased indicating colorless chalcone formation increased lightness (Table 1.5). The hues from all extracts were different from the standard Red No. 40. The most similar extracts to Red No. 40 were from *pr1* recessive anthers in Cluster 8 at pH 1 (Table 1.5). The extract representing Cluster 8 was 91.8% pelargonidin forms of total anthocyanins. Pelargonidin 3-glucoside standard by itself is not sufficient to match Red No. 40, but cyanidin 3-glucoside standard was only a ΔE of 6.02 from Red No. 40 at pH 1 (Table 1.5). Co-pigments within the anther samples must be interacting to produce the bathochromic shift necessary to closely match Red No. 40 (Chatham et al., 2020). Cluster 6, Cluster 7, and Cluster 8 maintained a ΔE under 20 from Red No. 40 at a pH of 3 (Table 1.5). Red No. 40 is not a perfect match to a true red (RGB 255, 0, 0) and its measured tristimulus values here differ at a ΔE of 23.88 at a pH of 1 (Table 1.5). The closest cluster to a true red was Cluster 6 at pH 1, which is a balance of pelargonidin forms and cyanidin forms from husks and cobs. This cluster was even closer to a true red than Red No. 40. Corn stover recessive for *pr1* might be a good alternative to Red No. 40 for a true application of red at a low pH. Natural water-soluble orange colors are also sought after in the natural colorant industry. Pelargonidin 3-glucoside standard at pH 3 is a ΔE of 23.02 from true orange (RGB 255, 165, 0) (Table 1.5). Varieties recessive for *pr1* produce mainly pelargonidin forms of anthocyanins. Removing co-pigments such as

15

flavones might match a truer orange color (Chatham et al., 2020). Overall, the diversity in hues from all tissue types make maize a suitable natural colorant replacement for acidic matrices.

1.4.7 Linear discriminant analysis accurately classifies tissue types solely on pigment composition

Linear discriminant analysis was performed to determine if tissue types could be discerned accurately by composition alone. Using compound proportions resulted in 90.3% classification accuracy (Table 1.6). Classifying using compound concentrations did not improve the analysis, nor did log-transformation. The two models had 74.2% and 84.5% accuracy, respectively. The linear discriminant model was accurate despite the large amount of diversity in compositions within tissues. It does not appear that *Pr1* variants have an effect on tissue classification. The lowest accuracy in classification was for cob glumes, with only 79.8% accuracy. The linear discriminant model had issues discriminating husks and cob glumes, since their pigment profiles tended to overlap. Husks were misclassified as cob glumes 13.83% of the time and *vice versa* 20.24% of the time. These results confirm the classification from PCA and cluster analysis and imply that husks and cob glumes may be regulated similarly in maize. Some kernel samples were outliers in that they clustered with husks and cob glumes (5.56% and 1.11%, respectively). Kernels had abundant representation in the dataset and therefore more likely to demonstrate outliers. Anthers were most often misclassified as tassel tissues, and vice versa, which is logical since these tissues are from the same organ. Tassel sampling was done so that anther tissue was minimized; however, it was tedious to fully separate the two. Results of linear discriminant analysis prove that major compositional differences can be found among tissue types.

1.5 Conclusions

Maize is a rich source of anthocyanins, not only in concentration, but in diversity. The array of compounds found in each tissue type provide a range of color and stability for natural colorants. Tissues outside of the grain are abundant in acylated forms and therefore may be more stable counterparts to kernel anthocyanins. Furthermore, compositional variation may be attributed to differences in regulation by the anthocyanin transcription factors. Results of this study demonstrate how anthocyanin composition data may be able to assist in predicting anthocyanin regulatory networks. In this survey, cob glumes held the most anthocyanins followed by leaf sheath tissue. These results indicate forage or corn stover consisting of cob,

16

husk, and stalk tissue from purple corn are valuable sources of natural colorants. The integration of *B1*, *Pl1*, and *a3* is vital to making a strongly pigmented purple corn variety. Non-vegetable portions of the maize plant can be valuable for the textile, cosmetic, and nutraceutical industries, while grain anthocyanins can be used as a co-product in food and beverage industries. The associations of anthocyanins with human health add value to these pigments. Overall results of this study show how the diversity of maize anthocyanins can provide an abundant source of pigment and a wide range of hues for the natural colorant industry.

1.6 Figures and Tables

Figure 1.1 Anthocyanin pigmentation in the various tissue types of pigmented maize

A) Genetic stock 320N demonstrating the *a3* phenotype. Husks, tassels, and leaf sheaths are intensely purple. B) Apache Red Cob (PI 213730) with weak leaf sheath, silk, and tassel pigmentation. C) Member of the Apache Red Cob × 320N population demonstrating the *a3* phenotype with intense leaf sheath, anther, husk, tassel, and silk pigmentation. D) Member of the Apache Red Cob × 320N population with silk and anther pigmentation, but weak stalk and tassel pigmentation. E) Ear with the *a3* phenotype and dark silks. F) Ear with dark silks and weak husk pigmentation. G) Typical *a3* husk. H) Tassel with the *a3* phenotype and dark anthers. I) Pigmented anther tissue. J) Purple corn. K) Purple cobs. L) Purple seedling. Photo Credit A–K UI Public Affairs: L. Brian Stauffer

18
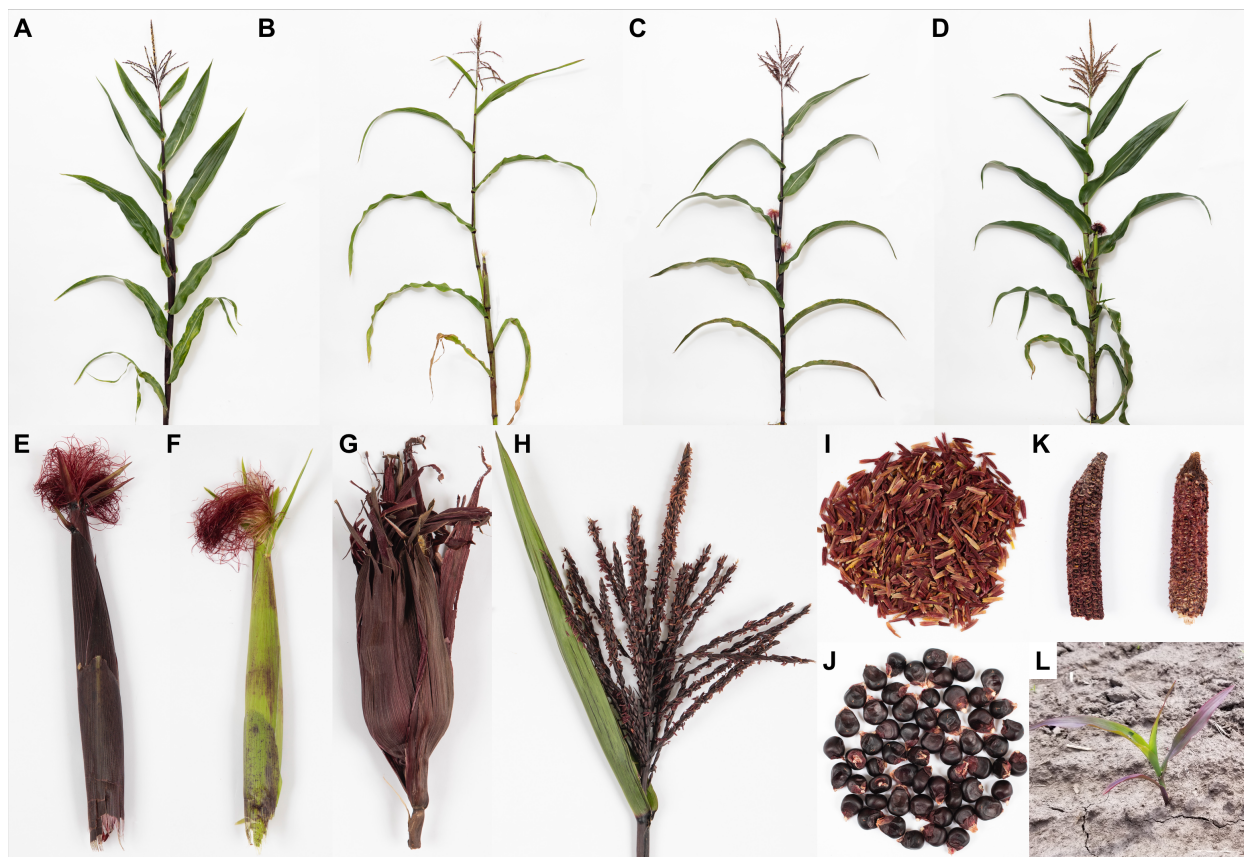
Figure 1.2 Structure of the novel anthocyanin cyanidin 3-6"-phenylacetylglucoside with fragmentation patterns
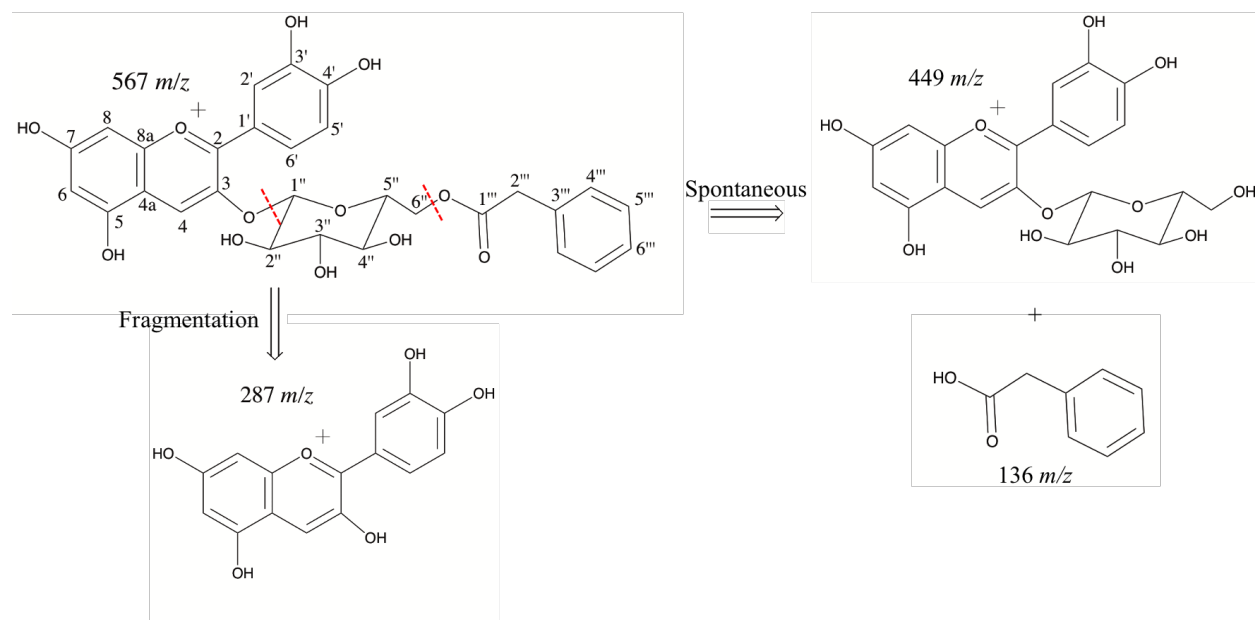
Figure 1.3 Principal component analysis (PCA) and hierarchical cluster analysis of various pigmented maize tissues

A) The first two principal components (PCs) grouped by tissue type. Each point represents the anthocyanin composition of a single sample as analyzed on the UHPLC. Arrows indicate eigen vectors for compounds with identities in Table 1.1. B) The same PCA grouped by Ward's Minimum Variance Method for clustering. Each star represents the composition of the extract used to represent each cluster for chromatograms (Figure 1.4) and tristimulus color values (Figure 1.6).
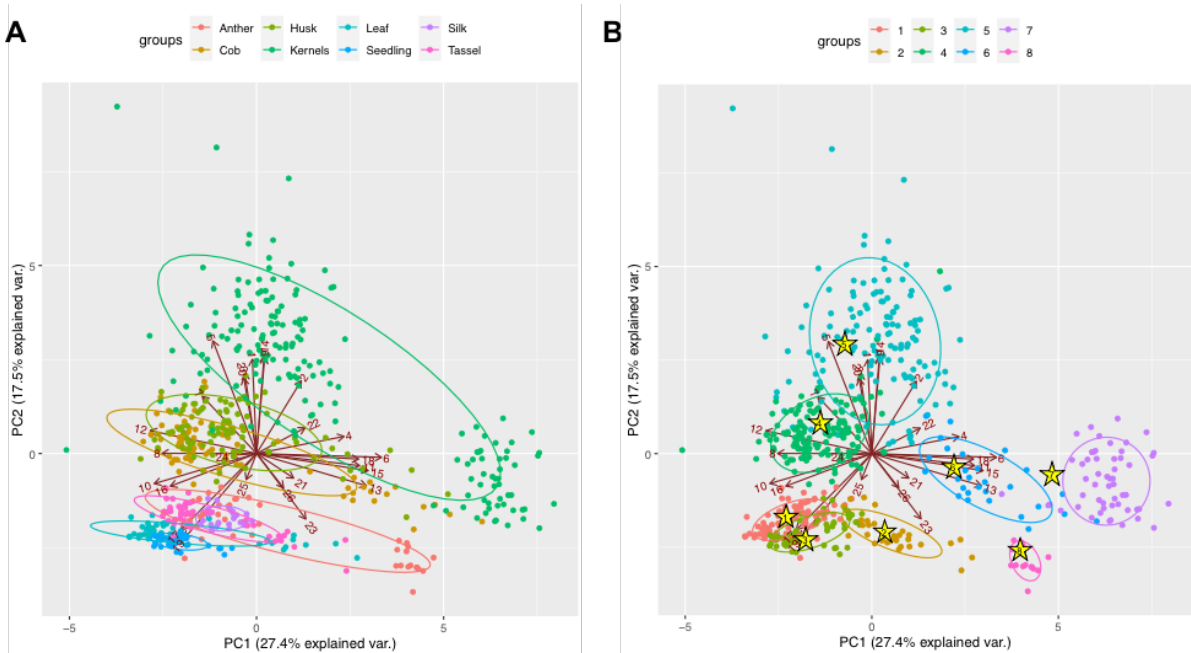
Figure 1.4 Chromatograms of tissue extracts representing the diversity in maize pigment production

Compound identities correspond to Table 1.1.
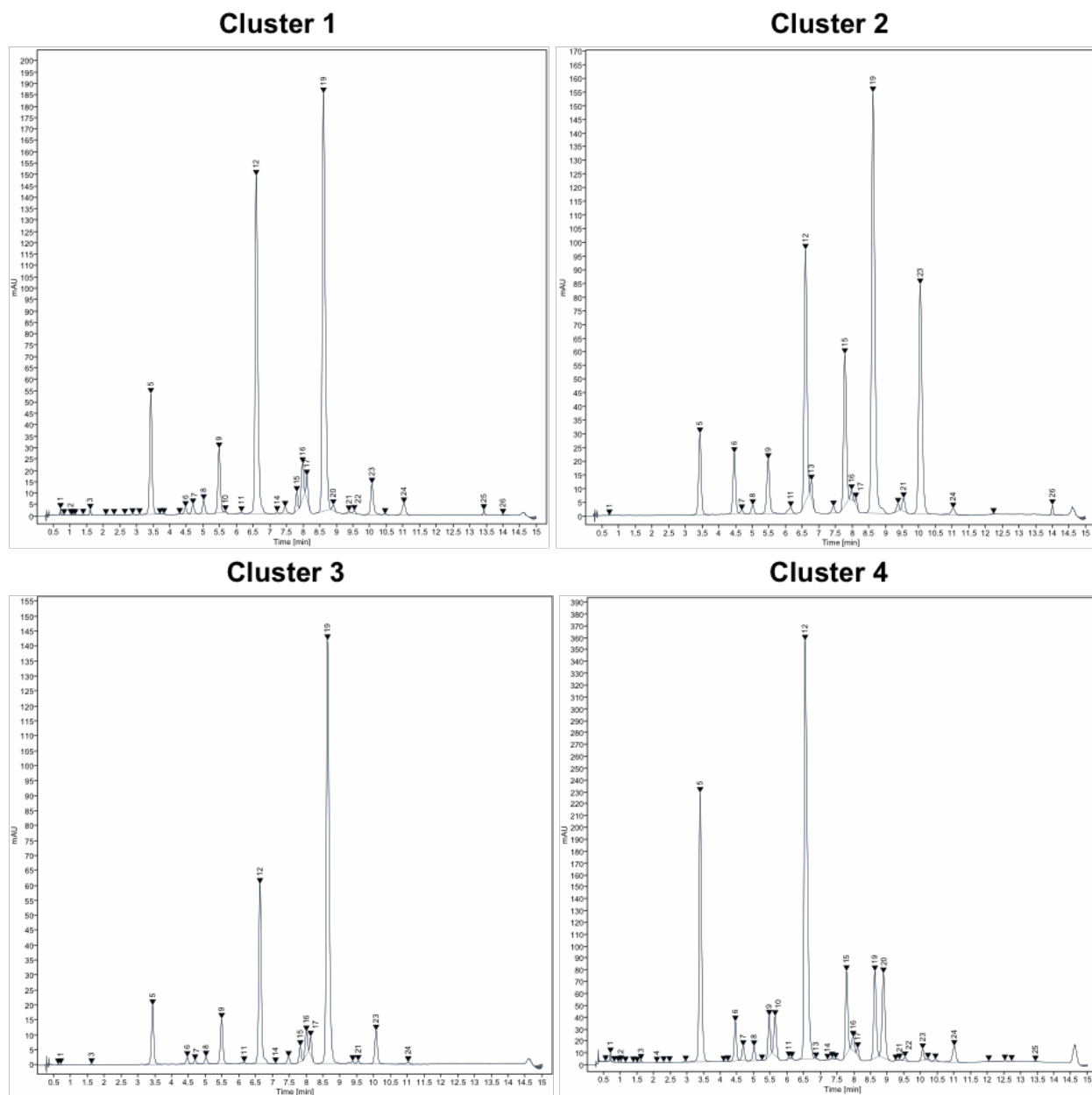
Figure 1.4 (Cont.)
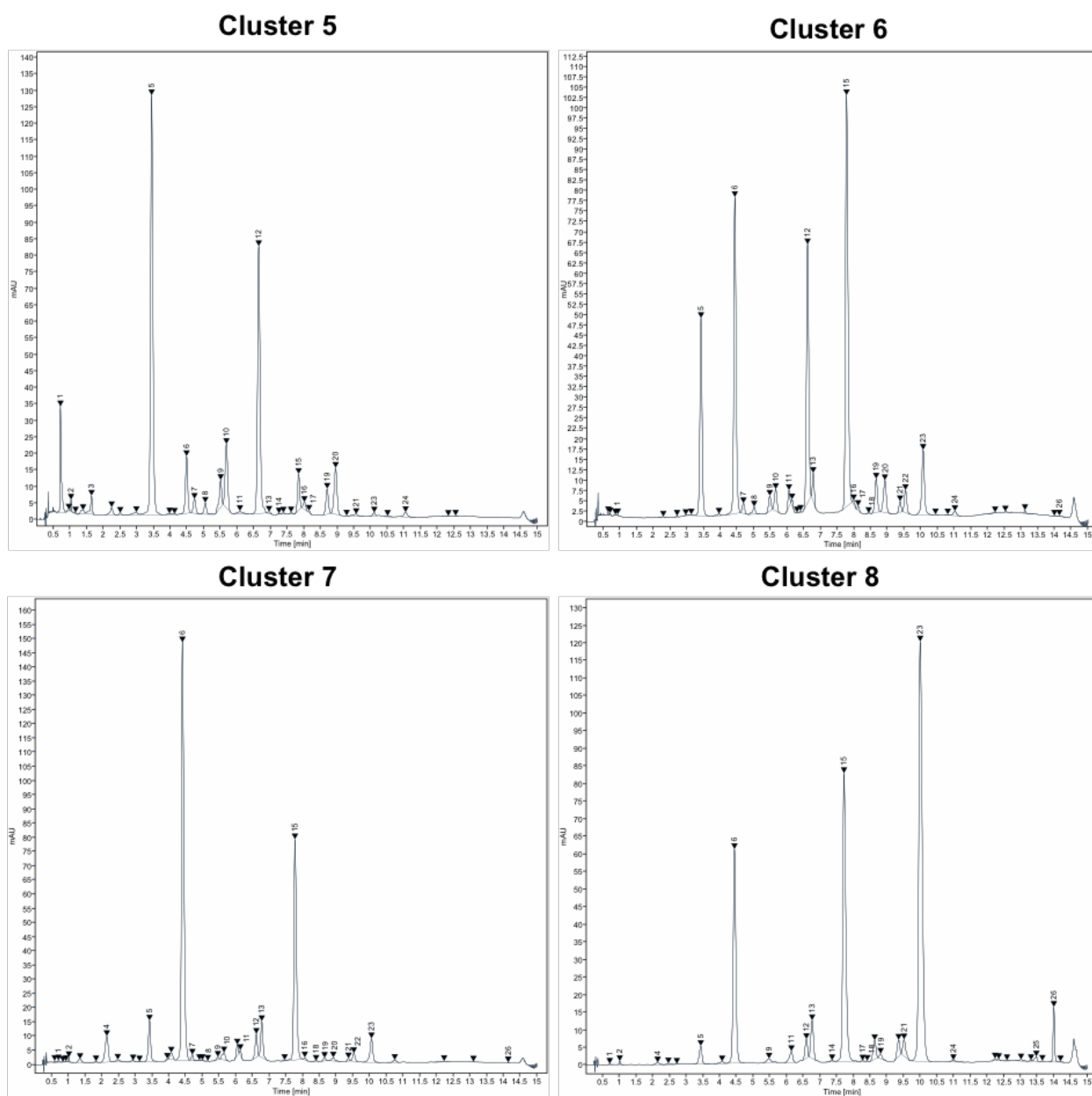
Figure 1.5 Heatmap of the average proportion of each compound separated by cluster

Rows indicate compound identities in Table 1.1. The color of each compound is scaled according to the legend.
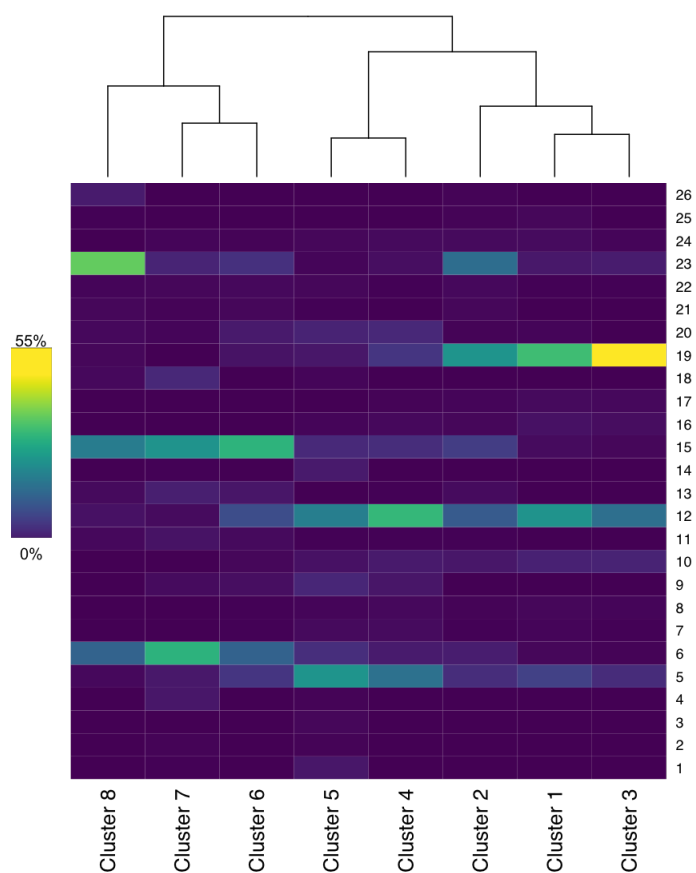
Figure 1.6 Tristimulus color values for extracts representing the diversity in anthocyanin composition in maize tissues

Color squares are from representative extracts of clusters determined by Ward's Minimum Variance Method compared to cyanidin 3-glucoside (C3G), pelargonidin 3-glucoside (Pg3G), and FD&C Red No. 40. Each sample was adjusted to a pH of 1, 3, or 6 and normalized to 50 µg mL^{-1}.

Table 1.1 Compound identities and mass spectrometry results

ID	Compound	Rt (min)	[M+H]+ *m/z*
1	Catechin-(4,8)-cyanidin 3,5-diglucoside	0.69	899
2	Catechin-(4,8)-pelargonidin 3,5-diglucoside	0.99	883
3	Catechin-(4,8)-cyanidin 3-6"-malonylglucoside-5-glucoside	1.56	985
4	Afzelechin-(4,8)-pelargonidin 3,5-diglucoside	2.12	867
5	Cyanidin 3-glucoside	3.38	449
6	Pelargonidin 3-glucoisde	4.41	433
7	Unknown Cyanidin	4.64	
8	Unknown Cyanidin	4.98	
9	Peonidin 3-glucoside	5.43	463
10	Cyanidin 3-3"-malonylglucoside	5.62	535
11	Unknown Pelargonidin	6.04	
12	Cyanidin 3-6"-malonylglucoside	6.56	535
13	Unknown Pelargonidin	6.76	
14	Unknown Cyanidin	7.14	
15	Pelargonidin 3-6"-malonylglucoside	7.74	519
16	Unknown Cyanidin	7.92	
17	Unknown Cyanidin	8.06	
18	Unknown Pelargonidin	8.39	
19	Cyanidin 3-3",6"-dimalonylglucoside	8.59	621
20	Peonidin 3-6"-malonylglucoside / Cyanidin 3-6"-succinylglucoside	8.88	549
21	Unknown Pelargonidin	9.35	
22	Unknown Cyanidin	9.50	
23	Pelargonidin 3-3",6"-dimalonylglucoside	10.00	605
24	Peonidin 3-3",6"-dimalonylglucoside	10.96	535
25	Cyanidin 3-6"-phenylacetylglucoside	13.33	567
26	Pelargonidin 3-6"-phenylacetylglucoside	13.93	551

25

Table 1.2 Summary of anthocyanin content and composition in the various tissue types of pigmented maize

	Total Anthocyanin Content			Proportion of Acylation (%)			Proportion of Condensed Forms (%)		
	Min	Average*	Max	Min	Average*	Max	Min	Average*	Max
Anther	21.04	196.39 c	852.67	73.97	79.96 c	88.14	0	0.30 b	0.96
Cob	13.64	608.65 b	3214.28	46.24	68.02 d	85.34	0	0.55 b	1.98
Husk	14.47	225.78 c	1931.55	59.13	69.30 d	83.08	0	0.07 b	0.80
Kernels	0.54	7.91 c	30.58	9.54	53.39 e	72.86	0	6.15 a	44.70
Leaf	42.37	913.56 a	3018.58	75.27	82.05 bc	89.41	0	0.28 b	0.89
Seedling	39.54	500.91 b	1208.81	83.11	86.49 a	89.98	0	0.19 b	0.54
Silk	14.11	511.82 b	1087.71	76.81	85.53 ab	90.15	0	0.16 b	0.90
Tassel	28.42	588.59 b	1744.81	76.97	82.74 abc	89.50	0	0.35 b	1.83

* Letters represent significant differences at $p<0.05$ using the Tukey's Honest Significant Differences method

26

Table 1.3 NMR shift results

Pos*	¹H (ppm)			¹³C (ppm)			HSQC	HMBC	CoSY
	ACD prediction	Bruker @27C	JEOL @25C	ACD prediction	Bruker @27C	JEOL @25C		pos (¹H) to pos (¹³C)	pos (¹H) to pos (¹H)
2				163.1	163.2	163.2			
3				144.1	142.6	144.1			
4	8.97	8.95 (s)	8.92 (s)	135.3	137.5	135.3	Y	3, 5, 8a, 2	
4a				111.9	113.9	111.6			
5				155.9	158.5	156.4			
6	6.85	6.67 (d, 1.8)	6.64 (d, 0.8)	102	103.8	?	Y	4a	
7				169.1	168.8				
8	6.93	6.92 (d, 1.8)	6.90 (d, 0.8)	93.9	95	?	Y		
8a				156.3	157.1	156.1			
1'				119.8	119.3	119.9			
2'	7.94	8.05 (d, 2.3)	8.02 (d, 2.4)	117	117.4	117.1	Y	6', 3', 4'	6' (w)
3'				146.1	147.3	146.2			
4'				154.6	154.1	154.6			
5'	7.34	7.05 (d, 8.8)	7.02 (d, 8.8)	116	111.8	116.1	Y	2', 1', 3'	6'
6'	7.74	8.27 (dd, 8.7, 2.3)	8.24 (dd, 8.7,2.6)	127	127.6	127.1	Y	2', 4', 2	5', 2' (w)
1"	5.43	5.31 (d, 7.8)	5.28 (d,7.8)	102.1	101.3	102.2	Y	3	2"
2"	3.94	3.67 (dd, 9.2, 7.9)	3.63 (dd, 9.2, 7.7)	73.2	74	73.3	Y	3", 1"	1", 3"
3"	3.86	3.56 (t, 9.1)	3.52, (t, 9.0)	76.4	78.1	76.5	Y	2"	2", 4"
4"	3.42	3.40 (t, 9.4)	3.37 (ddd, 9.7, 9.1, 3.4)	69.9	71.7	70	Y	5"	4", 5"
5"	3.6	3.80 (ddd, 9.9, 6.9, 2.1)	3.78 (ddd, 9.6, 7.0, 2.0)	74.3	76.1	74.7	Y		4", 6b", 6a" (w)
6a"	4.48	4.53 (dd, 12.0, 2.1)	4.49 (dd, 11.9, 2.0)	63.5	65.2	63.6	Y	1'''	6b", 5" (w)
6b"	4.23	4.24 (dd, 12.1, 7.0)	4.21 (dd, 11.9, 7.0)	63.5	65.2	63.6	Y		5", 6a"
1'''				171.9	173.8	172			
2'''	3.66	3.65 (2H, s)	3.62 (2H, s)	40.2	41.8	40.3	Y	4''', 3''', 1'''	
3'''				134	135.3	134			
4'''	7.34	7.16 (2H, m)	7.12 (m)	128.8 (*)	130.2	128.8	Y	6''' (*)	
5'''	7.33	7.16 (2H, m)	7.13 (m)	128.0 (*)	129.4	128.1	Y	3''' (*)	
6'''	7.29	7.16 (m)	7.12 (m)	126.6	127.9	126.7	Y		

* Position names refer to assignments in Table 1.1

Table 1.4 Representation of tissue types in each cluster determined by Ward's Minimum Variance Method

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Anther	19	8	0	0	0	0	0	15
Cob	0	0	0	59	8	15	0	0
Husk	0	0	0	76	7	11	0	0
Kernels	0	0	0	8	118	2	52	0
Leaf	40	9	0	0	0	0	0	0
Seedling	0	0	40	0	0	0	0	0
Silk	1	1	23	0	0	0	0	0
Tassel	34	4	14	0	0	0	0	0

Table 1.5 Tristimulus color experiment results

Clusters are representative extracts from groupings determined by Ward's Minimum Variance Method and compared to cyanidin 3-glucoside (C3G), pelargonidin 3-glucoside (Pg3G), and FD&C Red No. 40. Each sample was normalized to 50 µg mL^{-1} and analyzed using the same conditions in Luna-Vital et al. (2017).

	pH	L*	A*	B*	ΔE Red No. 40 pH1	ΔE True Red	ΔE True Orange
Cluster 1	1	67.81	59.82	34.22	23.30	41.37	57.79
Cluster 2	1	68.27	57.73	44.36	13.98	35.34	48.82
Cluster 3	1	66.61	59.25	32.63	24.89	42.53	58.84
Cluster 4	1	62.42	61.80	46.31	11.59	29.26	51.54
Cluster 5	1	58.49	65.60	46.88	13.22	25.51	55.10
Cluster 6	1	58.65	62.32	71.80	16.28	19.15	42.30
Cluster 7	1	65.53	57.48	72.94	16.46	26.38	35.35
Cluster 8	1	65.49	57.49	60.56	6.04	26.56	39.41
C3G	1	66.64	61.43	63.18	6.02	23.34	41.51
Pg3G	1	73.30	49.91	98.07	43.31	47.62	32.29
Red 40	1	66.03	62.54	57.29	0.00	23.88	45.15
Cluster 1	3	72.27	50.94	18.41	41.05	59.94	66.35
Cluster 2	3	71.11	51.21	28.53	31.33	51.48	57.46
Cluster 3	3	70.15	51.66	20.89	38.21	56.92	64.52
Cluster 4	3	67.07	53.24	25.95	32.71	51.13	61.07
Cluster 5	3	61.87	58.58	28.77	29.09	44.88	62.37
Cluster 6	3	60.73	55.84	48.13	12.53	31.76	46.58
Cluster 7	3	67.62	51.62	51.04	12.68	35.77	39.99
Cluster 8	3	67.81	51.89	44.03	17.10	39.31	45.30
C3G	3	71.07	56.78	39.13	19.71	40.62	51.77
Pg3G	3	76.09	44.07	67.86	23.54	42.68	23.02
Red 40	3	65.85	63.54	55.32	2.22	23.97	47.01
Cluster 1	6	90.71	6.76	3.12	81.58	104.35	79.34
Cluster 2	6	90.31	7.27	5.45	79.57	102.43	76.92
Cluster 3	6	86.34	6.32	11.81	75.11	98.02	70.35
Cluster 4	6	82.44	13.45	5.38	73.31	95.48	74.70
Cluster 5	6	79.07	13.51	7.76	70.90	92.92	72.07
Cluster 6	6	71.72	22.24	18.91	55.94	77.60	60.16
Cluster 7	6	82.55	13.79	11.60	68.83	91.37	68.54
Cluster 8	6	82.54	12.18	15.35	67.58	90.34	65.13
C3G	6	89.46	11.43	0.72	79.76	102.21	80.55
Pg3G	6	90.34	11.93	3.50	77.75	100.41	77.94
Red 40	6	66.85	58.13	60.32	5.41	26.75	39.77

29

Table 1.6 Confusion matrix for linear discriminant analysis

Values above the diagonal represent false positive (Type I error) values, while values below the diagonal represent false negative (Type II error) values.

	Anther	Cob	Husk	Kernels	Leaf	Seedling	Silk	Tassel
Anther	88.10%	0%	0%	0%	0%	0%	0%	11.90%
Cob	0%	79.76%	20.24%	0%	0%	0%	0%	0%
Husk	0%	13.83%	86.17%	0%	0%	0%	0%	0%
Kernels	0%	1.11%	5.56%	93.33%	0%	0%	0%	0%
Leaf	0%	0%	0%	0%	93.88%	0%	0%	6.12%
Seedling	0%	0%	0%	0%	0%	100.00%	0%	0%
Silk	0%	0%	0%	0%	0%	4.00%	88.00%	8%
Tassel	3.85%	0%	0%	0%	0%	0%	0%	96.15%

Above the diagonal represents false positive (Type I error) values, and below the diagonal represents false negative (Type II error) values.

1.7 References

Amchova, P., Kotolova, H., & Ruda-Kucerova, J. (2015). Health safety issues of synthetic food
colorants. *Regulatory Toxicology and Pharmacology*, *73*(3), 914–922.
https://doi.org/10.1016/j.yrtph.2015.09.026

Andorf, C. M., Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Campbell, D. A., & Sen, T. Z.
(2010). The Locus Lookup tool at MaizeGDB: Identification of genomic regions in maize
by integrating sequence information with physical and genetic maps. *Bioinformatics*,
26(3), 434–436. https://doi.org/10.1093/bioinformatics/btp556

Bhatla, S., & Pant, R. (1977). Isolation and characterisation of anthocyanin pigment from
phosphorus-deficient maize plants. *Current Science*, *46*, 700–702.

Campo, E., Saenz-Navajas, M. P., Cacho, J., & Ferreira, V. (2012). Consumer rejection threshold
of ethyl phenylacetate and phenylacetic acid, compounds responsible for the sweet-like
off odour in wines made from sour rotten grapes: CRT of ethyl phenylacetate and
phenylacetic acid. *Australian Journal of Grape and Wine Research*, *18*(3), 280–286.
https://doi.org/10.1111/j.1755-0238.2012.00198.x

Center for Food Safety and Applied Nutrition. (2020). *Report on the Certification of Color
Additives: 4th Quarter, Fiscal Year 2020, July 1-September 30*. Food and Drug
Administration. https://www.fda.gov/industry/color-certification-reports/report-
certification-color-additives-4th-quarter-fiscal-year-2020-july-1-september-30

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto,
L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N.,
Frewen, B., Baker, T. A., Brusniak, M.-Y., Paulse, C., Creasy, D., … Mallick, P. (2012).
A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*,
30(10), 918–920. https://doi.org/10.1038/nbt.2377

Chatham, L. A., Howard, J. E., & Juvik, J. A. (2020). A natural colorant system from corn:
Flavone-anthocyanin copigmentation for altered hues and improved shelf life. *Food
Chemistry*, *310*, 125734. https://doi.org/10.1016/j.foodchem.2019.125734

Chatham, L. A., & Juvik, J. A. (2021). Linking anthocyanin diversity, hue, and genetics in purple
corn. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa062

31

Chatham, L. A., West, L., Berhow, M. A., Vermillion, K. E., & Juvik, J. A. (2018). Unique flavanol-anthocyanin condensed forms in apache red purple corn. *Journal of Agricultural and Food Chemistry*, *66*(41), 10844–10854. https://doi.org/10.1021/acs.jafc.8b04723

De Nisi, P., Borlini, G., Parizad, P. A., Scarafoni, A., Sandroni, P., Cassani, E., Adani, F., & Pilu, R. (2021). Biorefinery approach applied to the valorization of purple corn cobs. *ACS Sustainable Chemistry & Engineering*, *9*(10), 3781–3791. https://doi.org/10.1021/acssuschemeng.0c08717

Hatier, J.-H. B., & Gould, K. S. (2008). Anthocyanin function in vegetative organs. In C. Winefield, K. Davies, & K. Gould (Eds.), *Anthocyanins* (pp. 1–19). Springer New York. http://link.springer.com/10.1007/978-0-387-77335-3_1

He, J., & Giusti, M. M. (2010). Anthocyanins: Natural colorants with health-promoting properties. *Annual Review of Food Science and Technology*, *1*(1), 163–187. https://doi.org/10.1146/annurev.food.080708.100754

Khamphasan, P., Lomthaisong, K., Harakotr, B., Ketthaisong, D., Scott, M. P., Lertrat, K., & Suriharn, B. (2018). Genotypic variation in anthocyanins, phenolic compounds, and antioxidant activity in cob and husk of purple field corn. *Agronomy*, *8*(11), 271. https://doi.org/10.3390/agronomy8110271

Lago, C. (2014). *Study of Maize Genotypes Rich in Anthocyanins for Human and Animal Nutrition* [Dissertation, Università Degli Studi Di Milano]. http://air.unimi.it/handle/2434/230010

Lago, C., Cassani, E., Zanzi, C., Landoni, M., Trovato, R., & Pilu, R. (2014). Development and study of a maize cultivar rich in anthocyanins: Coloured polenta, a new functional food. *Plant Breeding*, *133*(2), 210–217. https://doi.org/10.1111/pbr.12153

Lago, C., Landoni, M., Cassani, E., Doria, E., Nielsen, E., & Pilu, R. (2013). Study and characterization of a novel functional food: Purple popcorn. *Molecular Breeding*, *31*(3), 575–585. https://doi.org/10.1007/s11032-012-9816-6

Li, C.-Y., Kim, H.-W., Won, S., Min, H.-K., Park, K.-J., Park, J.-Y., Ahn, M.-S., & Rhee, H.-I. (2008). Corn husk as a potential source of anthocyanins. *Journal of Agricultural and Food Chemistry*, *56*(23), 11413–11416. https://doi.org/10.1021/jf802201c

Liu, X., Li, S., Yang, W., Mu, B., Jiao, Y., Zhou, X., Zhang, C., Fan, Y., & Chen, R. (2018). Synthesis of seed-specific bidirectional promoters for metabolic engineering of

anthocyanin-rich maize. *Plant and Cell Physiology*, *59*(10), 1942–1955.
https://doi.org/10.1093/pcp/pcy110

Ludwig, S. R., & Wessler, S. R. (1990). Maize R gene family: Tissue-specific helix-loop-helix
proteins. *Cell*, *62*(5), 849–851. https://doi.org/10.1016/0092-8674(90)90259-H

Luna-Vital, D., Li, Q., West, L., West, M., & Gonzalez de Mejia, E. (2017). Anthocyanin
condensed forms do not affect color or chemical stability of purple corn pericarp extracts
stored under different pHs. *Food Chemistry*, *232*, 639–647.
https://doi.org/10.1016/j.foodchem.2017.03.169

Mendiburu, F. de. (2020). *agricolae: Statistical Procedures for Agricultural Research*.
https://CRAN.R-project.org/package=agricolae

Mokrzycki, W. S., & Tatol, M. (2011). Color difference ΔE: A survey. *Machine Graphics and
Vision*, *Vol. 20, No. 4*, 383–411.

Nankar, A. N., Dungan, B., Paz, N., Sudasinghe, N., Schaub, T., Holguin, F. O., & Pratt, R. C.
(2016). Quantitative and qualitative evaluation of kernel anthocyanins from southwestern
United States blue corn. *Journal of the Science of Food and Agriculture*, *96*(13), 4542–
4552. https://doi.org/10.1002/jsfa.7671

Patterson, G. I., Kaszás, E., Cone, K. C., Coe, E. H., Jr., & Chandler, V. L. (1993). The pl gene
undergoes paramutation. *Maize Genetics Cooperation Newsletter*, *67*, 42–43.

Paulsmeyer, M., Brown, P., & Juvik, J. (2018). Discovery of Anthocyanin Acyltransferase1
(AAT1) in maize using genotyping-by-sequencing (GBS). *G3: Genes, Genomes,
Genetics*, *8*(11), 3669–3678. https://doi.org/10.1534/g3.118.200630

Paulsmeyer, M., Chatham, L., Becker, T., West, M., West, L., & Juvik, J. (2017). Survey of
anthocyanin composition and concentration in diverse maize germplasms. *Journal of
Agricultural and Food Chemistry*, *65*(21), 4341–4350.
https://doi.org/10.1021/acs.jafc.7b00771

Paulsmeyer, M., & Juvik, J. (2021). Functional characterization of an anthocyanin
dimalonyltransferase in maize. *Molecules*, *26*(7), 2020.
https://doi.org/10.3390/molecules26072020

Petroni, K., Cominelli, E., Consonni, G., Gusmaroli, G., Gavazzi, G., & Tonelli, C. (2000). The
Developmental expression of the maize regulatory gene hopi determines germination-
dependent anthocyanin accumulation. *Genetics*, *155*(1), 323–336.

Pilu, R., Panzeri, D., Cassani, E., Badone, F. C., Landoni, M., & Nielsen, E. (2009). A paramutation phenomenon is involved in the genetics of maize low phytic acid1-241 (lpa1-241) trait. *Heredity*, *102*(3), 236–245. https://doi.org/10.1038/hdy.2008.96

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Somavat, P., Li, Q., de Mejia, E. G., Liu, W., & Singh, V. (2016). Coproduct yield comparisons of purple, blue and yellow dent corn for various milling processes. *Industrial Crops and Products*, *87*, 266–272. https://doi.org/10.1016/j.indcrop.2016.04.062

Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., Leon, N. de, & Kaeppler, S. M. (2016). An expanded maize gene expression atlas based on rna sequencing and its use to explore root development. *The Plant Genome*, *9*(1), plantgenome2015.04.0025. https://doi.org/10.3835/plantgenome2015.04.0025

Styles, E. D., & Coe, E. H., Jr. (1986). Unstable expression of an R allele with a3 in maize: A recessive intensifier of plant color. *Journal of Heredity*, *77*(6), 389–393. https://doi.org/10.1093/oxfordjournals.jhered.a110267

Sugawara, S., Mashiguchi, K., Tanaka, K., Hishiyama, S., Sakai, T., Hanada, K., Kinoshita-Tsujimura, K., Yu, H., Dai, X., Takebayashi, Y., Takeda-Kamiya, N., Kakimoto, T., Kawaide, H., Natsume, M., Estelle, M., Zhao, Y., Hayashi, K., Kamiya, Y., & Kasahara, H. (2015). Distinct characteristics of indole-3-acetic acid and phenylacetic acid, two common auxins in plants. *Plant and Cell Physiology*, *56*(8), 1641–1654. https://doi.org/10.1093/pcp/pcv088

Suzuki, H., Nakayama, T., Yamaguchi, M., & Nishino, T. (2004). CDNA cloning and characterization of two Dendranthema × morifolium anthocyanin malonyltransferases with different functional activities. *Plant Science*, *166*(1), 89–96.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. http://www.stats.ox.ac.uk/pub/MASS4/

Yonemaru, J., Miki, K., Choi, S., Kiyosawa, A., & Goto, K. (2018). A genomic region harboring the allele from the Peruvian cultivar JC072A confers purple cob on Japanese flint corn (*Zea mays* L.). *Breeding Science*, *68*(5), 582–586. https://doi.org/10.1270/jsbbs.18090

Zhao, C.-L., Yu, Y.-Q., Chen, Z.-J., Wen, G.-S., Wei, F.-G., Zheng, Q., Wang, C.-D., & Xiao, X.-L. (2017). Stability-increasing effects of anthocyanin glycosyl acylation. *Food Chemistry*, *214*, 119–128. https://doi.org/10.1016/j.foodchem.2016.07.073

Zhao, J., Huhman, D., Shadle, G., He, X.-Z., Sumner, L. W., Tang, Y., & Dixon, R. A. (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. *The Plant Cell*, *23*(4), 1536–1555. https://doi.org/10.1105/tpc.110.080804

Žilić, S., Serpen, A., Akıllıoğlu, G., Gökmen, V., & Vančetović, J. (2012). phenolic compounds, carotenoids, anthocyanins, and antioxidant capacity of colored maize (Zea mays L.) Kernels. *Journal of Agricultural and Food Chemistry*, *60*(5), 1224–1231. https://doi.org/10.1021/jf204367z

CHAPTER 2: FUNCTIONAL CHARACTERIZATION OF AN ANTHOCYANIN DIMALONYLTRANSFERASE IN MAIZE[1]

2.1 Abstract

Anthocyanins are pigments with appealing hues that are currently being used as sources of natural colorants. The interaction of acylation on the stability of anthocyanin molecules has long been known. Maize is an abundant source of malonylglucoside and dimalonylglucoside anthocyanins. The enzyme Aat1 is an anthocyanin acyltransferase known to synthesize the majority of acylated anthocyanins in maize. In this paper, we characterize the substrate specificity and reaction kinetics of Aat1. It was found that Aat1 has anthocyanin 3-*O*-glucoside dimalonyltransferase activity and is only the second enzyme of this type characterized to this date. Our results indicate that Aat1 can utilize malonyl-CoA, succinyl-CoA, and every anthocyanin 3-*O*-glucoside tested. Results of this study provide insight into the structure–function relations of dimalonyltransferases and give a unique insight into the activity of monocot anthocyanin acyltransferases.

2.2 Introduction

Anthocyanins are a diverse group of pigments that are ubiquitous in the plant kingdom. They impart most of the orange to red and purple to blue colors present in plant species. Due to the prevalence of these compounds in many plant species and due to the attractiveness of their hues, anthocyanins are becoming popular natural substitutes for synthetic dyes. Purple corn is becoming an increasingly popular choice for anthocyanin extraction. The economy of scale for purple corn and the utilization of anthocyanins as co-products in the maize supply chain make purple corn an attractive alternative to other natural colorant sources currently on the market (Paulsmeyer et al., 2017; Somavat et al., 2016). Anthocyanins are also well-known health-promoting compounds. Purple corn extracts have a much higher antioxidant capacity than any other fruit and vegetable extract (Cevallos-Casals & Cisneros-Zevallos, 2003). In a review of the literature, purple corn extracts are shown to have associations with anti-inflammatory, antimutagenic, anticarcinogenic, and anti-angiogenesis (Lao et al., 2017). Incorporation of maize

---

extracts as sources of natural colorants will have dietary implications for increasing human health.

A major obstacle for adopting anthocyanins as natural colorants, however, is the inferior stability compared to synthetic dyes. Stability of anthocyanins is controlled in part by the inherent structure of the molecule. Adding acyl groups to the anthocyanin glycone is one method plants use to enhance the stability of these pigments. Acylation is defined as the esterification of organic acids to the sugar of the anthocyanin molecule. Two types of acylation are possible: aliphatic and aromatic. Aromatic acylation generally confers the greatest stability probably due to intramolecular stacking of pigment molecules (Zhao et al., 2017). Aliphatic acylation has also been shown to be important as it enhances total anthocyanin content in maize (Paulsmeyer et al., 2018). Aliphatic acylation has also been shown to decrease enzymatic degradation and increase the uptake of anthocyanins into the vacuole, all while maintaining the same hues of the molecule (Suzuki et al., 2002; J. Zhao et al., 2011). In vitro studies show that acylation protects the anthocyanin molecule during heat stress, intense light, and high pH (Bąkowska-Barczak, 2005; Zhao et al., 2017). Manipulating proportions of acylated anthocyanins in extracts seems to be an effective way of enhancing stability of natural colorants.

In maize, acylated anthocyanins account for a majority of the total anthocyanins produced, with aliphatic malonic acid derivatives being the most represented form (Paulsmeyer et al., 2017). In particular, the 6"-position of the glucoside is the first acylated position followed by the 3"-position for dimalonylglucosides (Figure 2.1). An enzyme catalyzing this reaction in maize has been discovered and is named *Anthocyanin acyltransferase1* (*Aat1*) (Paulsmeyer et al., 2018). Nearly all the acylated anthocyanins were depleted in dysfunctional mutants of this gene. However, it is unknown if the enzyme itself is capable of producing dimalonylglucoside anthocyanins present in the maize grain or what sort of specificity the enzyme has in terms of substrates. Only one enzyme has been characterized to this date that has dimalonyltransferase activity—chrysanthemum (*Dendranthema × morifolium*) Dm3Mat2 (Suzuki et al., 2004). In this study, we characterize the activity of recombinant Aat1 and determine specificity and reaction kinetics.

2.3 Materials and Methods

2.3.1. Substrates

Analytical standards of cyanidin 3-*O*-glucoside, delphinidin 3-*O*-glucoside, pelargonidin 3-*O*-glucoside, and peonidin 3-*O*-glucoside were purchased from Extrasynthese (Genay, France). Malonyl-CoA, succinyl-CoA, and acetyl-CoA were purchased from CoALA Biosciences (Austin, TX, USA). Quercetin 3-*O*-glucoside standard was purchased from Sigma-Aldrich. Anthocyanins were reconstituted in 0.1% HCl in water, while acyl donor substrates were reconstituted in water and used within a day. To prepare malonylglucoside standards, an enzyme assay was performed under standard conditions and run on a semi-preparative HPLC to isolate the malonylglucoside peak. The purified aliquots were combined, concentrated in a SpeedVac® AES2010 concentrator (Thermo Savant, Waltham, MA, USA) at 43 °C and then lyophilized. Powder was reconstituted with 0.1% HCl in water and the concentration was adjusted to 600 µM based on the A520 of cyanidin 3-*O*-glucoside given that the molar extinction coefficients are similar (Suzuki et al., 2002). Purity at A520 was 91.95% according to the UHPLC protocol described below.

2.3.2. Cloning

Full length *Aat1* transcript (NCBI Reference Sequence NP_001148286.2) was amplified using PCR with primers 5'-CATGATTC<u>GAATTC</u>ATGGCGGCAGCAACGGCAACT-3' and 5'-GCTACGAT<u>AAGCTT</u>TCACAGCAACCGGAGCCACTCCA-3' that introduce *EcoR*I and *Hind*III sites (underlined). Restriction enzyme cloning introduced the transcripts into the pET-30a(+) vector (MilliporeSigma, Burlington, MA, USA) as described by the supplier (EMD Chemicals Inc., 2011). Vectors were transformed into chemically competent Rosetta-gami™ 2(DE3)pLysS Competent Cells (MilliporeSigma, Burlington, MA, USA). LB media supplemented with 1 mM $MgSO_4$ to assist growth (Christensen et al., 2017) and 50 µg/mL kanamycin for selection were inoculated with cultures stored at −80 °C in 25% (*v/v*) glycerol. After cultivating overnight at 37 °C, two mL of overnight culture was added to 100 mL fresh media and grown at 37 °C to an OD600 of 0.6 to 0.8. Cultures were put on ice for five minutes before being induced with 0.4 mM IPTG overnight at 18 °C. Overnight cultures were centrifuged at 1600 ×*g* for 10 min. Media were decanted and pellets were resuspended in 100 mM potassium phosphate (pH 8.0), 10% glycerol, 1 mM EDTA, and 5 mM 2-mercaptoethanol (Buffer A) with

38

an addition of 1 mM PMSF. Bacterial lysis by sonication was done on ice with a 20 kHz probe set to 50% power with five 10 s bursts with 20 s rests between pulses.

2.3.3. Affinity Tag Purification

To capture 6x-His-tagged recombinant Aat1, one mL HisPur™ Ni-NTA agarose resin (Thermo Fisher Scientific, LLC, Waltham, MA, USA) was equilibrated with 10 column volumes ultrapure water and 20 column volumes of Buffer A + 10 mM imidazole before the addition of the protein extract. Whole crude protein extract was filtered in a 0.45 µm PES syringe filter (MilliporeSigma, Burlington, MA, USA) and equilibrated with constant shaking for 1 hr. The resin with captured protein was washed with 30 column volumes of Buffer A + 10 mM imidazole. The 6x-His-tagged protein was eluted in 3 column volumes of Buffer A + 500 mM imidazole. Purified protein was desalted in a G-25 MidiTrap (Cyteva, Marlborough, MA, USA) desalting column according to manufacturer's instruction using 100 mM potassium phosphate (pH 8.0) + 10% glycerol as the equilibration buffer. SDS-PAGE analysis was done on a 10% Mini-Protean® TGX™ gel with 50 µL wells (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

2.3.4. Enzyme Assays

A standard reaction consisted of 60 µM malonyl-CoA as the acyl donor and 120 µM cyanidin 3-*O*-glucoside as the acyl acceptor in 100 mM potassium phosphate (pH 8.0). To calculate reaction kinetics, acyl acceptors were varied between 5 and 120 µM, acyl donors were varied between 5 and 30 µM, and cyanidin 3-*O*-6"-malonylglucoside was varied between 2.5 and 30 µM. Anthocyanins, which were diluted with 0.1% HCl in water, were no more than 5% of the reaction volume so as to not change the pH of the reaction. The final reaction volume was 100 µL and each assay was replicated three times. Reactions were run for 30 min at 30 °C before being halted with 100 µL ice cold 4% (*v/v*) formic acid. Enzyme assays determining pH preference used 100 mM sodium acetate (pH 3.0 to 5.5), 100 mM potassium phosphate (pH 6.0 to 8.0), 100 mM Tris (pH 8.5 and 9.0), and 100 mM CAPS (pH 10 and 11). Kinetics constants were determined using the rearranged Michaelis–Menten equation proposed by K. A. Johnson (2019) (Johnson, 2019). Constants K_{cat} and K_{sp} (K_{cat}/K_m) were solved using the "nls" non-linear modeling function in R (R Core Team, 2019). Velocity was calculated as the summation of micromoles of product formed over seconds.

2.3.5. Liquid Chromatography and Mass Spectrometry

Anthocyanins in enzyme assays were quantified using an Agilent 1290 series UHPLC. An aliquot of 20 µL of enzyme assay was separated in an Agilent InfinityLab Poroshell 120 SB-C$_{18}$ (4.6 mm × 100 mm, 1.9 µm) column (Santa Clara, CA, USA) kept at 50 °C. The wavelengths used for detection were 520 nm for all anthocyanins or 280 nm for quercetin 3-*O*-glucoside. The mobile phase consisted of 2% formic acid (A) and acetonitrile (B) at a rate of 1.7 mL/min in a gradient from 4% B from 0 min to 1 min, 4% to 20% B from 1 to 10 min, then 20% to 100% in 30 s and back to 4% B in 30 s. The column was equilibrated with 4% B for 4 min between samples. Anthocyanin standards were varied in a linear range to calibrate peak areas. To confirm compounds formed in enzyme assays, representative samples were analyzed on a Waters Synapt G2-Si ESI/LC-MS/MS at the Mass Spectrometry Lab at the University of Illinois (Urbana, IL, USA) using the same column and temperature as the enzyme assays. The gradient was adjusted to run on the machine. The mobile phase consisted of 0.1% formic acid (A) and acetonitrile (B) at a rate of 1 mL/min in a gradient of 4% B from 0 to 1 min, 4% B to 25% B from 1 to 25 min, then a 95% B hold for 5 min and back to equilibration with 4% B for 5 min. Aliquots of 10 µL were injected for each sample. To prepare semi-pure malonylglucoside standards, standard enzyme assays were run on an Agilent 1100 Series HPLC with a semi-preparative fraction collector. Aliquots totaling 75 µL were separated in a Grace Prevail RP-C$_{18}$ (4.6 mm × 250 mm, 5 µm) column (Columbia, MD, USA) heated to 30 °C. The diode array detector was monitored at 280 nm. Mobile phases used were 2% formic acid (A) and acetonitrile (B) with a flowrate of 1 mL/min in a gradient of 10% B at 0 min, 30% B at 35 min, 100% B from 36 to 38 min, and equilibration for 5 min at 10% B.

2.4 Results and Discussion

2.4.1. Phylogenetic Analysis

Phylogenetic analysis of known flavonoid malonyltransferases is similar to previously reported analyses with other acyltransferase enzymes (Figure 2.2). Another monocot flavonoid malonyltransferase in *Oryza sativa*, OsMat2, is most similar to Aat1 in maize and shares a clade in the phylogenetic tree. However, they only share 51% sequence identity. Maize Aat1 shares 28% to 31% identity with other malonyltransferases in eudicots, which is typical of acyltransferase members. Previous studies report that the minimal amount of sequence similarity among acyltransferases is 25% to 34% (D'Auria, 2006). Despite this high sequence diversity,

40

flavonoid acyltransferases share unique signatures that are responsible for catalytic activity. The His-X-X-X-Asp domain (Motif 1) is a motif found in every acyltransferase member (Figure 2.3). Through protein crystallography, it was found that the His-residue in particular is responsible for catalyzing the acyl transfer (Manjasetty et al., 2012; Unno et al., 2007). The kinetic mechanism for anthocyanin malonyltransferases is most likely via a ternary complex in which the His residue from Motif 1 deprotonates the hydroxyl group of the acyl acceptor initiating a nucleophilic attack on the thioester of the acyl donor (Suzuki et al., 2003). Motifs Tyr-[Phe/Lys]-Gly-Asn-Cys (Motif 2) and Asp-Phe-Gly-[Trp/Phe]-Gly (Motif 3) are anthocyanin specific (Figure 2.3). Not many flavonoid acyltransferases have been characterized in monocots. The only other represented member is OsMat2, which has no activity with anthocyanin substrates (Kim et al., 2006). Both Aat1 and OsMat2 share the Phe variant of Motif 3, while all eudicot flavonoid acyltransferases contain the Trp residue (Figure 2.3). More monocot acyltransferases need to be characterized to test whether this variant is shared among the monocot lineage. As for the Motif 2 variant present in maize, it was proposed in Suzuki et al. (2004) that the Lys substitution may be involved with dimalonyltransferase activity as it may alter the binding pocket (Suzuki et al., 2004). Manjasetty et al. (2012) confirm in a crystallographic study that Motif 2 does in fact shape the binding site of the acyl donor (Manjasetty et al., 2012). However, OsMat2 also contains the Lys substitution, but does not have dimalonyltransferase activity. Unlike other flavonoid malonyltransferases, however, OsMat2 is able to freely accept both 3-*O*-glucoside and 7-*O*-glucoside forms equally, which is unique for a flavonoid malonyltransferase (Kim et al., 2009). Future studies should focus on the effect the Lys substitution has on the binding pocket and dimalonyltransferase activity.

2.4.2. Characterization of Aat1 Recombinant Protein

The gene model for *Aat1* has been revised since it was first discovered (Paulsmeyer et al., 2018). The revised transcript adds 94 amino acids missing due to an incorrectly annotated transcription start site. The corrected transcript was inserted into a pET-30a(+) 6x-His-tag expression vector and transformed into Rosetta-gami™ 2 cells (MilliporeSigma, Burlington, MA, USA), which alleviates codon bias and enhances disulfide bond formation for eukaryotic proteins. Protein was purified to apparent purity after affinity tag purification (Figure 2.4). The protein showed strong malonyltransferase activity and could catalyze the malonyl transfer to the 6"-position of the 3-*O*-glucoside. In addition, maize Aat1 can catalyze the subsequent addition of

41

a malonyl group to form dimalonylglucosides (Figure 2.5). This makes Aat1 only the second anthocyanin dimalonyltransferase characterized to this date. The formation of 3"-malonylglucosides is possible, but only detected in very small amounts (Figure 2.5). The dimalonyltransferase reaction for Aat1 occurs much more slowly than the initial malonylation reaction (Figure 2.6), but the specificity constant is slightly higher for cyanidin 3-O-6"-malonylglucoside than the glucoside form (Table 2.1). The enzyme has an activity range between a pH of 5.5 and 10 with a pH between 6.6 and 8.0 being maximal. Although not explicitly tested here, other anthocyanin malonyltransferases are inhibited by the presence of Cd^{2+}, Cu^{2+}, Fe^{2+}, Hg^{2+}, Mg^{2+}, and Zn^{2+} (Suzuki et al., 2001, 2002; Suzuki, Sawada, et al., 2003). Future studies should investigate the effect of these cations on Aat1 enzyme function.

Specificity and reaction kinetics of Aat1 varied widely from the other anthocyanin dimalonyltransferase, Dm3Mat2 (Suzuki et al., 2004). The most striking difference is the apparently slow rate of enzyme activity (Figure 2.6 and Table 2.1). The estimate for K_{cat} ranged from 7.25×10^{-4} to 6.23×10^{-3} s^{-1} (Table 2.1). This is in comparison with Dm3Mat2 which had rates ranging from 0.27 to 11.1 s^{-1} (Suzuki et al., 2004). The slow turnover rate may be the reason a typical blue corn sample does not completely convert all available 3-O-glucoside forms to malonylglucosides. In a survey of 98 diverse blue corn varieties, the average percentage of acylation was 63.19% (Paulsmeyer et al., 2017). This is in comparison to chrysanthemum, which has nearly 100% acylated anthocyanins present in the flowers (Hong et al., 2015). In addition, Aat1 differs in that it cannot effectively utilize delphinidin 3-O-glucoside as well as other 3-O-glucoside anthocyanins. No maize lines to this date have been found to produce appreciable amounts of delphinidin, malvidin, or petunidin, especially in the grain (Paulsmeyer et al., 2017). Cyanidin 3-O-glucoside appears to be the most preferred substrate for Aat1, but not in Dm3Mat2 (Suzuki et al., 2004). Similar to other anthocyanin acyltransferases, Aat1 can utilize quercetin 3-O-glucoside (results not shown) (Suzuki et al., 2002, 2004). Moreover, the enzyme can utilize succinyl-CoA as an acyl acceptor, but not acetyl-CoA. Acetyl-CoA may actually be an inhibitor of enzyme function according to other studies (Suzuki et al., 2001, 2002; Suzuki, Sawada, et al., 2003). The turnover rate and specificity constant for succinyl-CoA was much lower than that for malonyl-CoA, enforcing the notion that the enzyme is indeed a malonyltransferase (Table 2.1). Moreover, the enzyme cannot utilize rutinoside or 3,5-diglucoside anthocyanins (data not shown) probably due to steric hinderance of the additional sugars in the binding pocket.

42

2.5 Conclusions

Maize Aat1 is an anthocyanin dimalonyltransferase responsible for synthesizing the majority of anthocyanins in maize grain. It can utilize malonyl-CoA, succinyl-CoA and a variety of 3-*O*-glucoside anthocyanin substrates. Future studies should investigate the effect of the residues that may be important for dimalonyltransferase activity, especially the Lys substitution in Motif 2 (Figure 2.3). In addition, more malonyltransferase members in other monocot species need to be explored to determine phylogenetic relationships and structure–activity relationships. Overall, investigations into Aat1 have implications on increasing anthocyanin content in the grain and therefore implications on human health. Stabilizing anthocyanins, especially in purple and blue corn, has implications on making more economical sources of natural colorants and increasing the health-promoting aspects of maize.

2.6 Acknowledgements

2.7 Figures and Table

Figure 2.1 Structure of an anthocyanin 3-*O*-glucoside molecule with common substitutions

	R_1=H R_2=H Pelargonidin R_1=OH R_2=H Cyanidin R_1=OCH$_3$ R_2=H Peonidin R_1=OH R_2=OH Delphinidin R_1=OCH$_3$ R_2=OH Petunidin R_1=OCH$_3$ R_2=OCH$_3$ Malvidin
	Left: Acetyl-CoA Middle: Malonyl-CoA Right: Succinyl-CoA

Figure 2.2 Phylogenetic analysis of flavonoid malonyltransferases

From top to bottom: *Glandularia × hybrida* Vh3Mat1 (AAS77402.1), *Lamium purpureum* Lp3Mat (AAS77404.1), *Nicotiana tabacum* NtMat1 (BAD93691.1), *Salvia splendens* Ss5Mat (AAL50566.1), *Perilla frutescens* Pf5Mat (AAL50565.1), *Chrysanthemum × morifolium* Dm3Mat3 (BAF50706.1), *Dahlia pinnata* Dv3Mat (Q8GSN8.1), *Pericallis cruenta* Sc3Mat (AAO38058.1), *Chrysanthemum × morifolium* Dm3Mat1 (AAQ63615.1), *Chrysanthemum × morifolium* Dm3Mat2 (AAQ63616.1), *Oryza sativa* OsMat1 (NP_001046855.1), *Zea mays* ZmAat1 (NP_001148286.2). The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl & Pauling, 1965). Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).
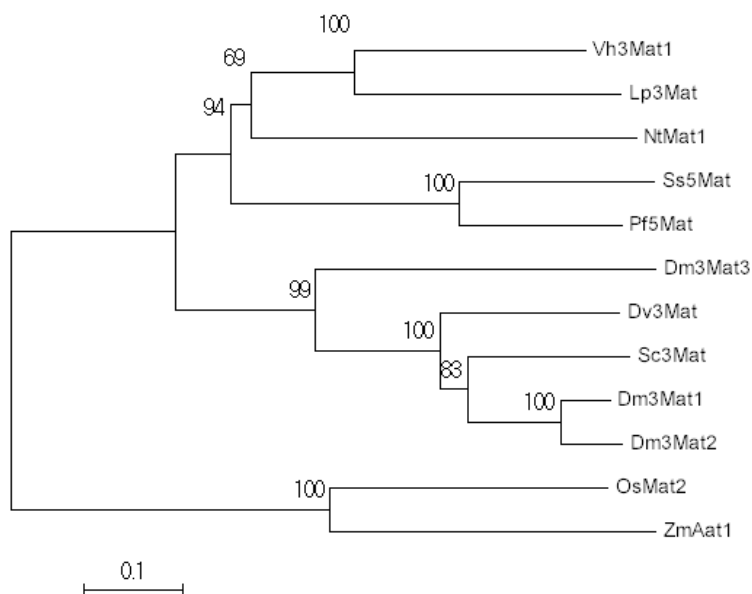
Figure 2.3 Alignment of known anthocyanin malonyltransferases

From top to bottom: *Glandularia × hybrida* Vh3Mat1 (AAS77402.1), *Lamium purpureum* Lp3Mat (AAS77404.1), *Nicotiana tabacum* NtMat1 (BAD93691.1), *Salvia splendens* Ss5Mat (AAL50566.1), *Dahlia pinnata* Dv3Mat (Q8GSN8.1), *Pericallis cruenta* Sc3Mat (AAO38058.1), *Perilla frutescens* Pf5Mat (AAL50565.1), *Chrysanthemum × morifolium* Dm3Mat1 (AAQ63615.1), *Chrysanthemum × morifolium* Dm3Mat2 (AAQ63616.1), *Chrysanthemum × morifolium* Dm3Mat3 (BAF50706.1), *Oryza sativa* OsMat1 (NP_001046855.1), *Zea mays* ZmAat1 (NP_001148286.2). Motifs 1 to 3 are indicated above their respective amino acids.

Figure 2.4 Purification of Recombinant Aat1

1) Uninduced culture, 2) resuspended culture before sonication, 3) soluble protein fraction after sonication, 4) flow-through of the Ni-NTA column, 5) purified Aat1 with an arrow indicating the protein band.
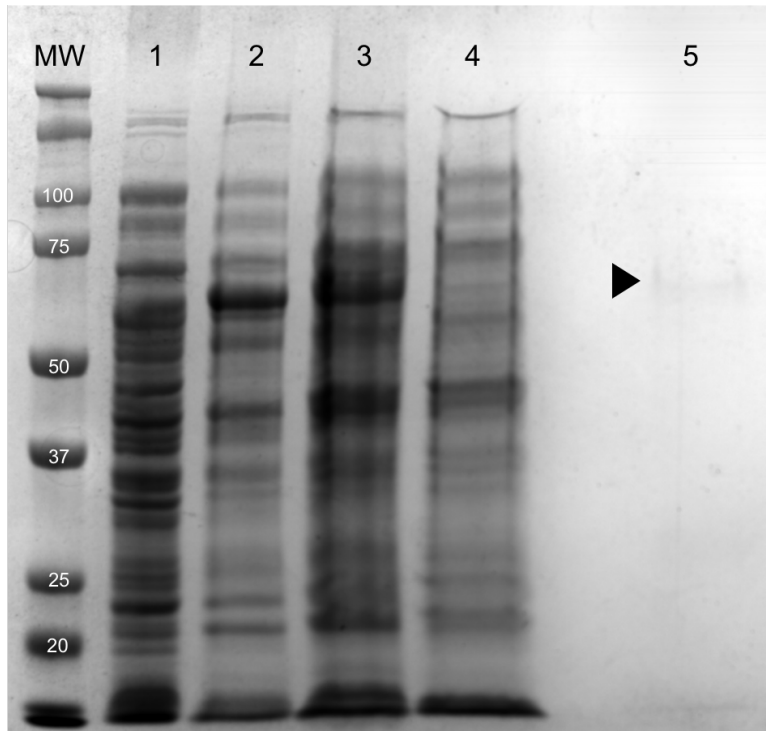
Figure 2.5 Aat1 is capable of synthesizing dimalonylglucoside anthocyanins

(a) Cyanidin 3-*O*-glucoside standard (Extrasynthese), (b) enzyme reaction with cyanidin 3-*O*-glucoside with enzyme products expanded inset. (c) Cyanidin 3-*O*-6"-malonylglucoside semi-preparative HPLC fraction. (d) Enzyme reaction with cyanidin 3-*O*-6"-malonylglucoside. Compound identities: 1) Cyanidin 3-*O*-glucoside, 2) Cyanidin 3-*O*-6"-malonylglucoside, 3) Cyanidin 3-*O*-3",6"-dimalonylglucoside, 4) Cyanidin 3-*O*-3"-malonylglucoside.
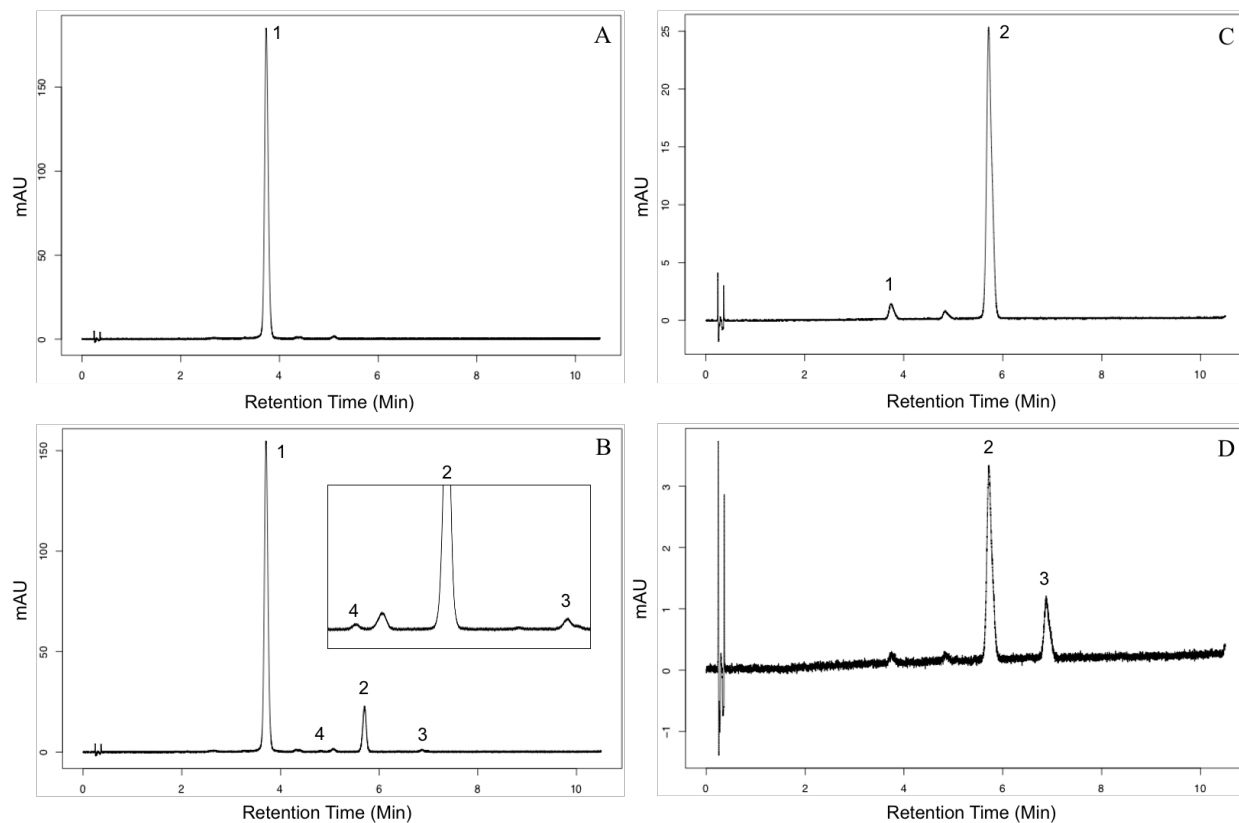
Figure 2.6: Michaelis-Menten Plots

Dots indicate measured velocity over initial substrate concentration. Velocity was determined as micromoles of product formed over time. Red lines represent the fitted model as determined by the Michael-Menten equation modified by K. A. Johnson (2019).
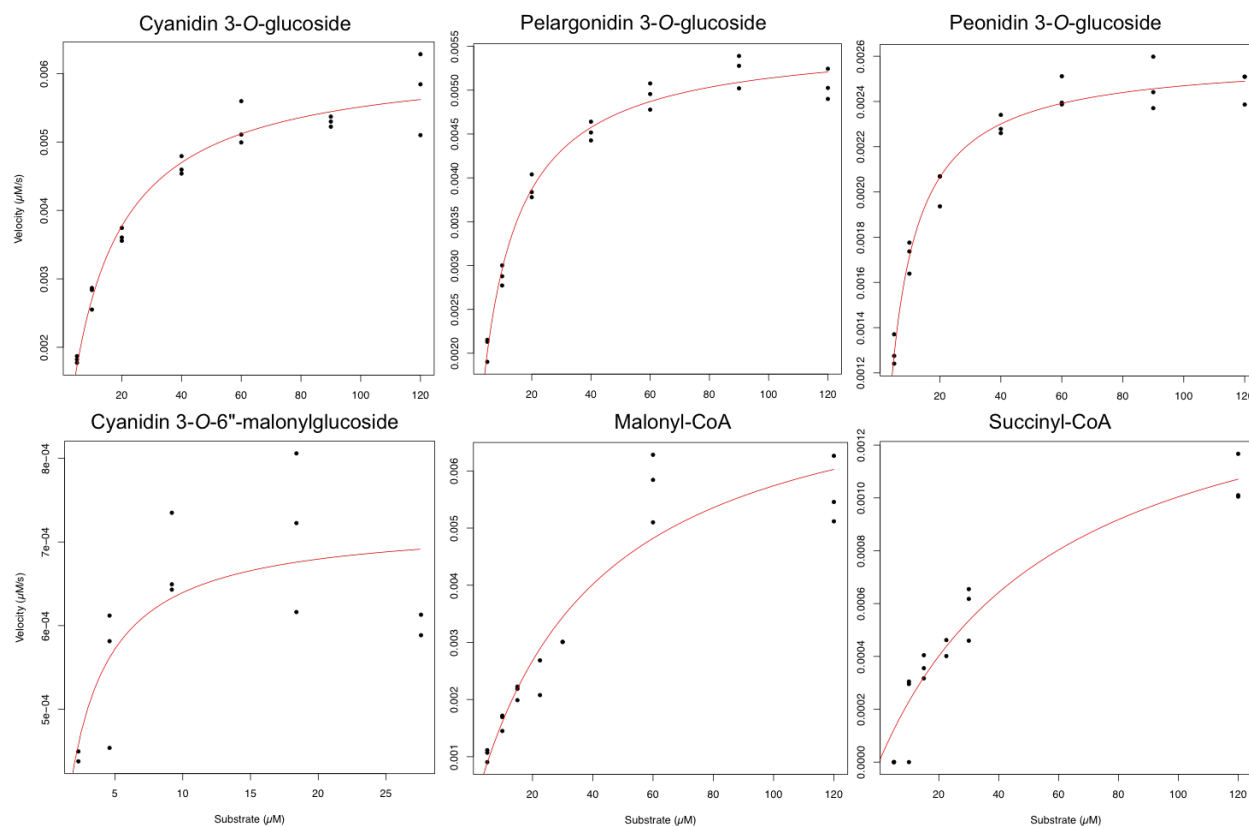
Table 2.1. Kinetic parameters and substrate specificity of Aat1

Compound	Relative Activity [a]	K_m	K_{cat}	K_{cat}/K_m
Units	%	µM	$s^{-1} \times 1000$	$µM^{-1}·s^{-1} \times 1000$
Acyl Acceptors [b]				
Cyanidin 3-*O*-Glucoside	100 ± 2.07	13.06 ± 0.23	6.233 ± 0.145	0.477 ± 0.034
Pelargonidin 3-*O*-Glucoside	89.8 ± 3.24	8.92 ± 0.41	5.594 ± 0.074	0.627 ± 0.030
Peonidin 3-*O*-Glucoside	43 ± 1.55	5.14 ± 0.93	2.596 ± 0.028	0.505 ± 0.026
Delphinidin 3-*O*-Glucoside	20.1 ± 0.74	ND	ND	ND
Cyanidin 3-*O*-6"-Malonylglucoside	ND	1.33 ± 3.97	0.725 ± 0.044	0.543 ± 0.173
Acyl Donors [c]				
Malonyl-CoA	100 ± 2.07	40.12 ± 0.04	8.043 ± 0.751	0.200 ± 0.026
Succinyl-CoA	18.5 ± 0.67	60.09 ± 0.02	1.607 ± 0.194	0.027 ± 0.004
Acetyl-CoA	<0.1	ND	ND	ND

[a] Relative activities were calculated with acyl donors in a final concentration of 60 µM and acyl acceptors in a final concentration of 120 µM. The reaction with malonyl-CoA and cyanidin 3-*O*-glucoside was taken to be 100%.
[b] Malonyl-CoA (60 µM) was the acyl donor.
[c] Cyanidin 3-*O*-glucoside (120 µM) was the acyl acceptor. ND = Not Determined; reported values are ± standard error.

2.8 References

Bąkowska-Barczak, A. (2005). Acylated anthocyanins as stable, natural food colorants. *Polish Journal of Food and Nutrition Sciences*, *55*(2), 107–116.

Cevallos-Casals, B. A., & Cisneros-Zevallos, L. (2003). Stoichiometric and kinetic studies of phenolic antioxidants from andean purple corn and red-fleshed sweetpotato. *Journal of Agricultural and Food Chemistry*, *51*(11), 3313–3319. https://doi.org/10.1021/jf034109c

Christensen, D. G., Orr, J. S., Rao, C. V., & Wolfe, A. J. (2017). Increasing growth yield and decreasing acetylation in escherichia coli by optimizing the carbon-to-magnesium ratio in peptide-based media. *Applied and Environmental Microbiology*, *83*(6). https://doi.org/10.1128/AEM.03034-16

D'Auria, J. C. (2006). Acyltransferases in plants: A good time to be BAHD. *Current Opinion in Plant Biology*, *9*(3), 331–340.

EMD Chemicals Inc. (2011). *pET System Manual 11th Edition*. https://www.emdmillipore.com/US/en/product/pET-30a+-DNA-Novagen,EMD_BIO-69909

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(4), 783–791. https://doi.org/10.2307/2408678

Hong, Y., Tang, X., Huang, H., Zhang, Y., & Dai, S. (2015). Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in chrysanthemum. *BMC Genomics*, *16*(1), 202. https://doi.org/10.1186/s12864-015-1428-1

Johnson, K. A. (2019). New standards for collecting and fitting steady state kinetic data. *Beilstein Journal of Organic Chemistry*, *15*, 16–29. https://doi.org/10.3762/bjoc.15.2

Kim, B.-G., Lee, Y., Hur, H.-G., Lim, Y., & Ahn, J.-H. (2006). Flavonoid 3′-O-methyltransferase from rice: CDNA cloning, characterization and functional expression. *Phytochemistry*, *67*(4), 387–394. https://doi.org/10.1016/j.phytochem.2005.11.022

Kim, D. H., Kim, S. K., Kim, J.-H., Kim, B.-G., & Ahn, J.-H. (2009). Molecular characterization of flavonoid malonyltransferase from Oryza sativa. *Plant Physiology and Biochemistry*, *47*(11–12), 991–997. https://doi.org/10.1016/j.plaphy.2009.08.004

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, msw054. https://doi.org/10.1093/molbev/msw054

51

Lao, F., Sigurdson, G. T., & Giusti, M. M. (2017). Health benefits of purple corn (Zea mays L.) phenolic compounds. *Comprehensive Reviews in Food Science and Food Safety*, *16*(2), 234–246. https://doi.org/10.1111/1541-4337.12249

Manjasetty, B. A., Yu, X.-H., Panjikar, S., Taguchi, G., Chance, M. R., & Liu, C.-J. (2012). Structural basis for modification of flavonol and naphthol glucoconjugates by Nicotiana tabacum malonyltransferase (NtMaT1). *Planta*, *236*(3), 781–793. https://doi.org/10.1007/s00425-012-1660-8

Paulsmeyer, M., Brown, P., & Juvik, J. (2018). Discovery of Anthocyanin Acyltransferase1 (AAT1) in maize using genotyping-by-sequencing (GBS). *G3 Genes|Genomes|Genetics*, *8*(11), 3669–3678. https://doi.org/10.1534/g3.118.200630

Paulsmeyer, M., Chatham, L., Becker, T., West, M., West, L., & Juvik, J. (2017). Survey of anthocyanin composition and concentration in diverse maize germplasms. *Journal of Agricultural and Food Chemistry*, *65*(21), 4341–4350. https://doi.org/10.1021/acs.jafc.7b00771

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Somavat, P., Li, Q., de Mejia, E. G., Liu, W., & Singh, V. (2016). Coproduct yield comparisons of purple, blue and yellow dent corn for various milling processes. *Industrial Crops and Products*, *87*, 266–272. https://doi.org/10.1016/j.indcrop.2016.04.062

Suzuki, H., Nakayama, T., & Nishino, T. (2003). Proposed Mechanism and Functional Amino Acid Residues of Malonyl-CoA:Anthocyanin 5-*O*-Glucoside-6'''-*O*-Malonyltransferase from Flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family. *Biochemistry*, *42*(6), 1764–1771. https://doi.org/10.1021/bi020618g

Suzuki, H., Nakayama, T., Yamaguchi, M., & Nishino, T. (2004). CDNA cloning and characterization of two Dendranthema × morifolium anthocyanin malonyltransferases with different functional activities. *Plant Science*, *166*(1), 89–96.

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Nakao, M., Tanaka, Y., Yamaguchi, M. -a., Kusumi, T., & Nishino, T. (2001). Malonyl-CoA:Anthocyanin 5-O-Glucoside-6'''-O-Malonyltransferase from scarlet sage (Salvia

52

splendens) Flowers: Enzyme purification, gene cloning, expression, and characterization. *Journal of Biological Chemistry*, *276*(52), 49013–49019. https://doi.org/10.1074/jbc.M108444200

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M., Tanaka, Y., Kusumi, T., & Nishino, T. (2002). CDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A:anthocyanidin 3-O-glucoside-6"-O-malonyltransferase from Dahlia flowers. *Plant Physiology*, *130*(4), 2142–2151. https://doi.org/10.1104/pp.010447

Suzuki, H., Sawada, S., Yonekura-Sakakibara, K., Nakayama, T., Yamaguchi, M., & Nishino, T. (2003). Identification of a cDNA encoding malonyl-coenzyme A: Anthocyanidin 3-O-Glucoside 6"-O-Malonyltransferase from Cineraria (Senecio cruentus) Flowers. *Plant Biotechnology*, *20*(3), 229–234. https://doi.org/10.5511/plantbiotechnology.20.229

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M., & Nakayama, T. (2007). Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *Journal of Biological Chemistry*, *282*(21), 15812–15822. https://doi.org/10.1074/jbc.M700638200

Zhao, C.-L., Yu, Y.-Q., Chen, Z.-J., Wen, G.-S., Wei, F.-G., Zheng, Q., Wang, C.-D., & Xiao, X.-L. (2017). Stability-increasing effects of anthocyanin glycosyl acylation. *Food Chemistry*, *214*, 119–128. https://doi.org/10.1016/j.foodchem.2016.07.073

Zhao, J., Huhman, D., Shadle, G., He, X.-Z., Sumner, L. W., Tang, Y., & Dixon, R. A. (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. *The Plant Cell*, *23*(4), 1536–1555. https://doi.org/10.1105/tpc.110.080804

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, *97*, 97–166.

CHAPTER 3: *ANTHOCYANIN3* (*A3*) IS A NEGATIVE REGULATOR OF ANTHOCYANIN SYNTHESIS IN MAIZE

3.1 Abstract

Anthocyanins are colorful molecules receiving recent attention due to their numerous health benefits and applications as natural colorants and nutraceuticals. Purple corn in particular is being investigated as a more economical source of natural colorants. *Anthocyanin3* (*A3*) is a known recessive intensifier of anthocyanin pigmentation in maize. In this study, anthocyanin content was elevated 100-fold in mutant *a3* plants. Two genetic approaches were used to pinpoint the gene involved with the *a3* intense purple plant phenotype. First, a large-scale transposon-tagging population was created with a *Dissociation* (*Ds*) insertion in the nearby *Anthocyanin1* gene. A *de novo a3-m1::Ds* mutant was generated and the transposon insertion was found to be located in *Mybr97*, which has homology to R3-Myb-like repressor *CAPRICE* in Arabidopsis. Second, a bulk segregant RNA-sequencing population found expression differences between pools of green *A3* plants and purple mutant *a3* plants. *Mybr97* was highly downregulated, confirming its role as the gene responsible for the *a3* phenotype. All characterized anthocyanin biosynthetic genes were upregulated in mutants along with several genes within the lignin pathway. Photosynthesis-related gene expression was reduced in mutant plants through an unknown mechanism. Novel transcription factors and biosynthetic genes were also upregulated and need further investigation. The mechanism of *A3* inhibition is most likely through the interaction with basic-helix-loop-helix transcription factors like master anthocyanin regulator *Booster1*. Overall, *A3* has a profound effect on the maize plant and reducing its expression has many favorable implications for crop protection, in human health, and as a source of natural colorants.

3.2 Introduction

Anthocyanins are the colorful molecules present in nearly every plant species. They are part of the diverse superfamily of molecules derived from phenylalanine in the phenylpropanoid pathway. Anthocyanins provide many valuable functions for the plant and are involved with a multitude of biotic and abiotic stress responses, including UV protection, herbivory defense, and reactive oxygen species (ROS) scavenging (Hatier & Gould, 2008; Lev-Yadun & Gould, 2008). Anthocyanins have many important human applications as well. The attractiveness of their hues and ease of extraction makes these pigments valuable as natural colorants. Purple corn (*Zea mays*

L.) in particular is an appealing source of natural colorants due to the economy of scale of corn production and density of pigment production (Chatham et al., 2019). In addition, purple corn has numerous known health benefits attributed to anthocyanins and other antioxidant compounds that make it beneficial as a potential nutraceutical (Lao et al., 2017). Anthocyanins are well-known health-promoting compounds and have been shown to reduce disease biomarkers associated with cancer, cardiovascular disease, diabetes, inflammation, and obesity (He & Giusti, 2010). Increasing dietary consumption of purple corn, whether through supplements or through replacing artificial dyes in foods and beverages, has a potential positive effect on human health. Given this, investigations into the genetics regulation of anthocyanin biosynthesis in maize are justified.

Anthocyanin synthesis in all plant species is regulated by a core set of transcription factors collectively referred to as the MBW complex for the Myb, bHLH, and WD-repeat proteins that must physically interact for anthocyanin-related genes to be transcribed (Hichri et al., 2011). In maize, the MBW complex is made of multiallelic gene families that coordinate the temporal and spatial expression of anthocyanin synthesis. In aleurone-pigmented tissue, Myb *Colored aleurone1* (*C1*), bHLH *Colored1* (*R1*), and WD-repeat *Pale aleurone color1* (*Pac1*) predominantly activate pigment synthesis (Cone, 2007). In mature leaf sheaths, husks, and tassels, anthocyanin synthesis is regulated by Myb *Purple plant1* (*Pl1*), bHLH *Booster1* (*B1*), and an unknown WD-repeat protein (Carey et al., 2004).

In addition to core regulators, two classic genes—*Anthocyanin3* (*A3*) and *Intensifier1* (*In1*)—are known to enhance the anthocyanin pathway when recessive. *Intensifer1* has homology to a bHLH transcription factor and inhibits pigment synthesis in the aleurone. It is thought that In1 physically interacts with R1 in order to inhibit pigment synthesis in aleurone (Burr et al., 1996). The other negative regulator, *A3*, has not been well-characterized, but the locus responsible has been known for decades (Lindstrom, 1934). The *a3* phenotype is very prevalent when the weak allele of *B1*, or *B1-b*, is present along with *Pl1*. Plants with this genetic makeup and *A3* are the normal condition of green but flush with color as the plants reach maturity and senesce. This is in contrast with *a3* plants that are intensely purple in leaf sheaths beginning after the ninth leaf stage (V9) and in husks and tassels (Figure 3.1) (Styles & Coe, 1986). The gene or element conferring the phenotype is thought to act at the level of transcription (Carey, 2002). The *A3* locus lies somewhere on Chromosome 3 in close proximity

55

to *Anthocyanin1* (*A1*) (Beckett et al., 1973), but there have been some discrepancies as to where in segment of the genome the gene or element lies. One study placed *A3* upstream of *A1* about 4.4 cM and this is reflected also on MaizeGDB (Andorf et al., 2010; Styles & Coe, 1986). One study linked intense leaf sheath pigmentation to *A3* using RFLP markers *umc63* and *csu25a*, which are both downstream of *A1* (Andorf et al., 2010; Lauter et al., 2004). The most recent calculation found that *A3* is 7.7 cM downstream of *A1* (Robinett et al., 1995). In this work, two genetics approaches are used to discover the gene responsible for the *a3* phenotype and to characterize the effect of this negative regulator on the anthocyanin and phenylpropanoid pathways in maize.

3.3 Materials and Methods

3.3.1 Plant materials

The reference allele for the *a3* mutant, referred to as *a3-ref*, was in genetic stock 320N and was provided by the Maize Genetics Cooperation Stock Center. The active transposon stock W22 *Ac-im*;*a1-m3::Ds* was kindly donated by Kevin Ahern. This transposon stock contains *Activator-immobile* (*Ac-im*) and a mobile *Dissociation* (*Ds*) insertion in *A1* (Conrad & Brutnell, 2005). To create the transposon-tagging population, reciprocal crosses of W22 *Ac-im;a1-m3::Ds* and 320N were made in the summer of 2018 at the University of Illinois Vegetable Crops Farm in Urbana, IL, USA. Progeny of the cross were grown in the same location the following year and monitored for the *a3* phenotype. Predicted *de novo* mutants were crossed to 320N and also selfed. To create the bulk segregant RNA-seq population, 320N was crossed to B73, the maize reference genome. In 2019, all F_2 plants with the *a3* phenotype and plants that appeared to have weak coloration due to *B1-b* (Figure 3.1) were selfed. Select $F_{2:3}$ families were grown in 2020 for bulk segregant RNA-seq analysis.

3.3.2 Nucleic acid extractions

DNA extractions for all plant tissues described below followed a CTAB procedure (Doyle & Doyle, 1990) with modifications. Freeze dried or frozen fresh tissue was ground to a fine powder and extracted in 2% (*m/v*) CTAB buffer containing 100 mM Tris (pH 7.5), 1.4 M NaCl, 20 mM EDTA, 1% (*m/v*) PVP-40, and 0.2% (*v/v*) β-mercaptoethanol. The extract was purified with an equal volume of 24 to 1 (*v/v*) chloroform to isoamyl alcohol and centrifuged. The top fraction was collected and precipitated with 1 volume of ice-cold isopropyl alcohol and precipitated overnight. The resulting pellet was centrifuged and washed with consecutive washes

56

of cold 70% ethanol and 95% ethanol before being resuspended and diluted in 10 mM Tris (pH 8.0). Quality and quantity of DNA was assessed for PCR suitability using a NanoDrop 2000 (Thermo Fisher Scientific, LLC). Tissue for RNA extraction was flash-frozen in liquid nitrogen during collection, freeze dried, ground into a fine powder, and stored at -20 ˚C before extraction. RNA extraction was performed using the RNeasy Plant Mini Kit (QIAGEN) with manufacturer's instructions. DNA was depleted during RNA extraction on the column using DNase (QIAGEN). RNA quality was assessed using a NanoDrop 2000 (Thermo Fisher Scientific, LLC) and a bleach gel (Aranda et al., 2012).

3.3.3 Inverse PCR of transposon-flanking sites

The general protocol for inverse PCR of *de novo Ds* transposon insertion events was previously described (Ahern et al., 2009). Here, three µg DNA from three separate leaf tissues was digested with *Pst*I-Hi-Fi (New England Biolabs) using the manufacturer's instructions, purified using the E.Z.N.A. DNA purification kit (Omega Bio-tek), and eluted in 50 µL 10 mM Tris (pH 8.0). Pure digested DNA was circularized with 1 U µL^{-1} T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) in a reaction volume of 100 µL for 16 hours at 16 ˚C and purified as described above. One µL of circularized DNA and primers LC18 and LC24 (Table 3.1) were used in the first round of PCR using the Platinum SuperFi DNA Polymerase (Invitrogen) kit. The thermocycler conditions were as follows: denature at 98 °C for 1 min followed by 35 cycles of 98 °C for 10 s, 62 °C for 15 s, and 72 °C for 1 min 30 s. The reaction ended with a 5 min extension at 72 °C. The second round of nested PCR (for stringent *Ds* selectivity) used 1 µL of the previous reaction diluted 1:500 in water and primers LC45 and JGp2 (Table 3.1). The thermocycler conditions of the second round were similar except the annealing temperature was 66 ˚C and cycle extension lasted 1 min 15 s. The final reaction was purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, LLC). The resulting transposon-flanking sites were cloned into DH5-α competent cells using the PCR Cloning Kit (New England Biolabs) according to the manufacturer's instructions. Resulting colonies were cultivated in 5 mL LB with ampicillin (100 µg/ml final concentration), pelleted, and extracted using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific, LLC) using manufacturer's instructions. The resulting plasmids were sequenced at the DNA Services laboratory of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign using the LC45 primer (Table 3.1).

57

3.3.4 Molecular marker analysis

The most recent mapping work determined that the *a3* locus was 7.7 cM downstream of *A1* (Robinett et al., 1995). For this reason, SSR markers downstream from *A1* were chosen from the Maize Genetics and Genomics Database to assess genotypes for *a3-ref* (Andorf et al., 2010). Marker *umc2008,* which is 3.65 Mb from *A1* in B73, was ultimately chosen based on efficient gel separation. Sequences for *umc2008* primers are provided is Table 3.1. The marker was amplified using *Taq* 5x Master Mix (New England Biolabs). The reaction consisted of 1.2 M betaine and 600 nm primers. The PCR conditions for *umc2008* are as follows: denature at 94 ˚C for 1 min followed by 35 cycles of 94 ˚C for 30 s, 53 ˚C for 30 s, and 72 ˚C for 30 s, ending with 3 min extension at 72 ˚C. Reactions were separated on a 4% SFR (VWR) agarose gel in TBE buffer for 45 min at 5.4 V cm^{-1} and scored visually based on parent band size.

3.3.5 RNA-seq library preparation

Sampling for bulk segregant RNA-seq analysis followed a similar procedure as previously described (Chayut et al., 2015). RNA from mature husk tissue (greater than 20 DAP) was chosen to assay for *A3*. In each $F_{2:3}$ family chosen from the B73 × 320N population, three individual plants were designated as biological replicates in either mutant (purple) or normal (green) pools and genotyped with *umc2008* to validate the phenotypes observed. Some families were segregating for *a3-ref* and had members representing both mutant and normal pools. One biological replicate from each family was pooled with their respective genotypes and collected all on the same day. In total there were 6 pooled samples: 3 biological replicates of 31 normal families and 3 biological replicates of 31 mutant families. RNA was extracted from three technical replicates of each biological replicate to alleviate tissue heterogeneity using the RNA extraction protocol above. Pure RNA was pooled in equal proportions according to Nanodrop 2000 concentrations. High-quality RNA from each biological replicate was submitted to the DNA Services laboratory of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign and quantified by Qubit (Life Technologies). Using the TruSeq Stranded RNA Sample Preparation Kit (Illumina San Diego, CA), PolyA+ RNA was selected from 1 µg of total RNA per sample. PolyA+ RNA was fragmented for 4 min at 94˚ C, then first-strand cDNA was synthesized with a random hexamer and SuperScript II (Life Technologies). Double-stranded DNA was blunt-ended, 3'-end A-tailed and ligated to unique dual-indexed adaptors. The adaptor-ligated double-stranded cDNA was amplified by PCR for 10 cycles with the Kapa HiFi

58

polymerase (Kapa Biosystems). The final libraries were quantified on Qubit and the average size determined on the AATI Fragment Analyzer (Advanced Analytics) and diluted to 5 nM final concentration. The 5 nM dilution was further quantitated by qPCR on a BioRad CFX Connect Real-Time System (Bio-Rad Laboratories, Inc.). The final stranded RNA-Seq library pool consisting of 6 libraries was sequenced on one lane of an Illumina NovaSeq 6000 SP flowcell as paired-reads with 150 nt length. Raw data was adaptor-trimmed and demultiplexed using bcl2fastq v2.20 Conversion Software (Illumina). Raw data has been deposited into the Short Read Archive (SRA) database under the project identity PRJNA764765.

3.3.6 Bulk segregant RNA-seq analysis

Raw adapter-trimmed transcriptome data was trimmed and quality filtered using Fastp (Chen et al., 2018). First, ten nucleotides from the beginning and three nucleotides from the end of each read were trimmed. Then, reads below a quality score of 30 and less than 50 nt were removed. Quality filtered reads were then aligned to the B73 RefGen_v4 genome using STAR 2.7.6 (Dobin & Gingeras, 2015; Jiao et al., 2017). Gene counts were calculated with featureCounts (Liao et al., 2014). The summary of the read counts from the RNA-seq filter pipeline is in Table 3.2. Differential gene expression was calculated in EdgeR (Robinson et al., 2010) using gene counts normalized to library size. The full normalized dataset with gene annotations and differential gene expression calculations is available at https://github.com/mpaulsmeyer/A3. Genes with an FDR-adjusted p-value less than 0.01 and an absolute value log_2-fold change greater than 1.2 were considered significant. GO-term enrichment was calculated using AgriGO v2 (Tian et al., 2017). To call SNPs from the RNA-seq libraries, STAR 2.7.6 two-pass mode (Dobin & Gingeras, 2015) was enabled, and the resulting BAM files were passed through SAMtools 1.11 (https://github.com/samtools/samtools/releases/1.11) and BCFtools 1.9 (https://github.com/samtools/bcftools/releases/tag/1.9) (Bonfield et al., 2021; Danecek et al., 2021). Genotypes were called using the multiallelic calling model using all compiled data from each biological replicate (Danecek et al., 2014). Only biallelic SNPs with a final depth greater than or equal to 150 and an average quality of greater than or equal to 30 were kept in the analysis. A custom script was used to accumulate the depth of each allele (either reference or alternate) at each position.

3.3.7 Quantitative PCR

Various time points of B73 development were sampled to determine *A3* expression in various tissues. All tissues were an average of 5 biological replicates. Roots and seedlings were collected from whole V1 samples grown in lab conditions. Leaf sheath tissue was collected from V5 and V9 tissue and R1 tissue was collected from husks just after silking from plants grown in the field. Pericarp samples were collected at 10, 15, and 20 DAP. For pericarp, whole self-pollinated ears were flash frozen, freeze dried, and pericarps from at least 5 kernels from each ear were peeled off to make a pericarp-enriched biological replicate. Validation of RNA-seq results was done using the same RNA as was sequenced. Additionally, husk tissue collected from three 320N and three B73 plants on the same day were compared for gene expression to the RNA-seq population. RNA was converted to cDNA using poly-T primers and the Protoscript M-MuLV First Strand cDNA Synthesis Kit (New England Biolabs). qPCR was performed using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, LLC). Reaction mixes and cycling conditions for qPCR and the melt curve analysis were as suggested by manufacturer's instructions. Expression values from the timepoint study were normalized to housekeeping gene *Elongation factor 1α,* while bulk-segregant RNA-seq validation used *Elongation factor 1α* and *β-Tubulin* as the controls as was proposed in a previous study (Lin et al., 2014). Genes and primers used for qPCR are in Table 3.1. Data analysis was done using QuantStudio 7.0 software (Thermo Fisher Scientific, LLC). Overall, there was high agreement with the data from qPCR and RNA-seq. The $\Delta\Delta C_T$ (relative change in cycle threshold) and $\log_2$-fold change data correlated very highly ($\rho = 0.97$) indicating the RNA-seq data is reliable and comparisons are generally accurate.

3.3.8 Cloning of maize anthocyanin transcription factors

Coding sequences of full *B1-b* and the Myb-interacting region (MIR) were amplified using cDNA from 320N husk tissue. Full-length *A3* cDNA was amplified using cDNA from B73 husk tissue. *B1-MIR* included amino acids 1–212, which include the MIR as predicted by a conserved domain search (Lu et al., 2020). Primers used introduced restriction enzyme sites *Eco*RI and *Hind*III. The 320N *a3-reference* allele was cloned from genomic DNA and introduced into the pMiniT 2.0 vector from the PCR cloning kit using manufacturer's instruction (New England Biolabs). Primers used for cloning are provided in Table 3.1. PCR was conducted using the Platinum SuperFi II PCR Master Mix with manufacturer's instructions (Thermo Fisher

60

Scientific, LLC). Restriction enzyme sites were used to insert *B1* and *B1-MIR* in-frame into vector pGSTag, which was kindly donated by D. Ron (Addgene plasmid #21877) (Ron & Dressler, 1992). The *A3* cDNA sequence was inserted into vector pET-30a(+) (MilliporeSigma), which contains a 6x-Herstidine tag. All vectors were cloned into DH5-α for sequence confirmation and Rosetta-gami2 pLysS (DE3) (MilliporeSigma) for protein expression. Sanger sequencing was performed at the DNA Services laboratory of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign using pGEX primers for pGSTag, T7 primers for pET-30a(+), and the Cloning Analysis primers for pMiniT 2.0 (Table 3.1).

3.3.9 GST-Pulldown assay

Cultures inoculated from glycerol stocks were cultivated overnight in LB media supplemented with 1 mM $MgSO_4$ (Christensen et al., 2017) and the appropriate antibiotic. The next day, one mL from GST-tagged strains was used to inoculate 50 mL of fresh media. For *A3*, two mL of culture was used to inoculate 100 mL fresh media. Cultures were grown to an OD600 of 0.6 to 0.8 before induction. Cultures containing *A3* and GST alone were induced with 0.5 mM IPTG for 3 hours at 37 ˚C. Cultures containing *B1* and *B1-MIR* were put on ice for 5 min before being induced with 0.1 mM IPTG overnight at 18 ˚C. Bacteria were pelleted by centrifugation and the media was decanted. Pellets were resuspended with 50 mM Tris (pH 8) + 150 mM NaCl + 1 mM EDTA + 0.1% (*v/v*) Triton-X100 + 10% glycerol (Buffer A) in one-tenth the volume of the original media. Whole cell cultures were sonicated with a Virtis Virsonic 475 20 kHz Sonicator probe (SP Industries) set to 50% power with five 10 s bursts with 20 s rest in between. For 6x-His tag purification, one mL of His-Pur Ni-NTA resin (ThermoFisher Scientific, LLC) was equilibrated with 10 resin-bed volumes (RV) of water and 20 RV Buffer A + 10 mM imidazole (Buffer B). Clarified A3 lysate was filtered through a 0.45 μm PES syringe filter (MilliporeSigma) and loaded onto the 6x-His tag column. Extract was incubated for 1 hr at 4 ˚C with constant mixing. Captured protein was washed with 30 RV of Buffer B before being eluted with Buffer A + 500 mM imidazole. Pure protein was desalted using a PD G-25 MidiTrap desalting column (Cytiva) according to manufacturer's instruction with 50 mM Tris (pH 8.0) + 150 mM NaCl + 10% glycerol as the equilibration buffer. GST was purified using 50 μL resin (GenScript). Resin was equilibrated using three 1 mL washes of Buffer A with centrifugation at 5000 *x g* between washes. Clarified lysate (500 μL) was equilibrated with the resin with constant mixing for 1 hour at 4 ˚C. Resin was washed three more times with Buffer A before being

61

incubated with 100 µL A3 (approximately 200 µg) + 400 µL Buffer A for 1 hr at 4 ˚C with constant mixing. The resin was washed three more times with Buffer A and mixed with 100 µL 2X SDS-PAGE Loading Buffer (Bio-Rad Laboratories). Samples were boiled for 10 min before being loaded on a 12% Mini-Protean TGX (Bio-Rad Laboratories) gel with 50 µL wells. About 5 µL A3, 20 µL GST controls, and 45 µL GST-tagged transcription factors were added to the gel.

3.3.10 Total anthocyanin and total phenolic assays

Bulk-segregant RNA-seq analysis tissue was used for anthocyanin extractions along with three biological replicates of B73 and 320N husks sampled on the same day. Triplicates of 25 mg were extracted with 1 mL 2% (*v/v*) formic acid for 1 hr at 50 ˚C for each biological replicate. Samples were centrifuged for 10 min and the supernatant was diluted in 25 mM potassium chloride (pH 1.0) and 0.4 M sodium acetate (pH 4.5) as per the AOAC standard pH differential method (Lee et al., 2005). Anthocyanins were quantified as cyanidin 3-*O*-glucoside equivalents using the molar extinction coefficient of 26,900 L M^{-1} cm^{-1} (Lee et al., 2005). Total phenolics were extracted in triplicate from 25 mg of tissue from each biological replicate in 1 mL 95% ethanol for 1 hr at 50 ˚C. Total phenolics were quantified using a method previously described (Singleton & Rossi, 1965), but scaled down 1:10 to be adapted for a 96-well plate. In this procedure, twenty µL diluted (1:5 or 1:10 in 95% ethanol) phenolics samples were mixed with 100 µL 0.2 N Folin-Ciocalteu's Regent and incubated for 5 to 6 min before the addition of 80 µL freshly made 7.5% (*w/v*) sodium carbonate. The plates were incubated for at least one hour in the dark and the absorbance was read at 765 nm. Total phenolics samples were quantified in gallic acid equivalents (GAEs) using a standard curve of 31.25 to 500 µg ml^{-1} on each plate.

3.4 Results

3.4.1 Transposon-tagging reveals the A3 gene

The exact location of *A3* within the genome was disputed. The gene either lies upstream or downstream of *A1* (Robinett et al., 1995; Styles & Coe, 1986). Regardless of this uncertainty, transposon-tagging could be employed utilizing a known active transposon stock with *A1* as the reporter gene in a W22 background (Conrad & Brutnell, 2005). W22 has an active *B1-b* allele, but has a recessive *Pl1* allele, so the plant is a normal green phenotype (Andorf et al., 2010). Crossing W22 *Ac-im;a1-m3::Ds* to 320N results in the *B1-b* weak pigmentation phenotype unless the *Ds* transposon is mobilized and randomly inserts into the *A3* gene (Figure 3.1). Approximately 25,000 plants were screened for the *a3* phenotype. One plant out of the

62

population was intensely purple and was a confirmed hybridization event as assayed by SSR marker *umc2008*. DNA was extracted from three different leaf tissues from this plant to be used for inverse PCR of the transposon-insertion site. In all three replicates, the *de novo a3-m1::Ds* plant was pinpointed to a G-box element in the promotor region of *Mybr97* (Zm00001d044194), a gene with homology to the Arabidopsis (*Arabidopsis thaliana*) *CAPRICE* (*AtCPC*) R3-Myb-like gene. This gene is a known negative regulator of anthocyanin synthesis in Arabidopsis (Zhu et al., 2009). The *Mybr97* insertion event was backcrossed to 320N, which resulted in 40 fully purple plants indicating it was a heritable insertion event.

3.4.2 Bulk-segregant RNA-seq confirms Mybr97 as the A3 gene

A segregating population was made by crossing B73, the maize reference genome, to genetic stock 320N with the *a3-ref* gene. B73 is a green normal type plant at maturity (Figure 3.1) with an inactive allele of *b1* and a light-inducible allele of *Pl1* (Andorf et al., 2010). 320N carries the *a3* intense purple phenotype (Figure 3.1) and has the weak *B1-b* allele with a functional *Pl1*. 320N also has a null allele of *r1* that does not confer pigmentation. B73 has anther and seedling pigmentation indicative of the *r1-r* allele (Ludwig & Wessler, 1990). RNA-seq of pools of mutant purple and normal type green husk tissues resulted in an average of 93.4 million 150 nt paired-end reads per pooled sample (Table 3.2). After filtering and mapping, an average of 64.9 million paired end reads per sample were used for downstream analyses. SNPs were called from each of the two combined pools and resulted in 78,815 high quality SNPs. The average distance between SNPs was 11.6 kb. Two measures were used to determine significant differences between mutant and normal pools. The first measurement, ΔSNP, took the difference of scored genotypes (0 = B73 allele, 1 = 320N allele, and 0.5 = heterozygote) for normal vs. mutant pools. Hypothetically, the ΔSNP window for the *A3* locus would be close to 1. The second calculated F_{st} given the allelic depth of each SNP, for which a sufficient threshold for differences is 0.25 (Hartl & Clark, 1997). Both values were smoothed in window size of 50 SNPs with a step of 10 SNPs to alleviate noise from the expression dataset. This resulted in an interval on Chromosome 3 from 219.8 to 225.0 Mb where ΔSNP was greater than 0.9 (Figure 3.2A) and an interval on Chromosome 3 from 219.3 to 225.5 Mb where F_{st} was greater than 0.25 (Figure 3.2B). No other intervals were significant in either analysis (data not shown). Both of these calculations confirm that *A3* is somewhere downstream of *A1* in accordance with the latest mapping studies (Lauter et al., 2004; Robinett et al., 1995). Within the F_{st} interval, there were 16

63

transcription factors and a total of 183 expressed genes. One of the genes within the interval was *A1* itself, which indicates the locus is too conservatively large. Recombination between markers was not enough in the $F_{2:3}$ generation to narrow down the interval.

Differential gene expression analysis found 1794 differentially expressed genes with a $\log_2$-fold change of 1.2 or greater and an FDR adjusted *p*-value of 0.01 or less. Within the ΔSNP interval, a total of 32 genes were differentially regulated with 16 of those being downregulated in mutant vs. normal samples (Table 3.3). Transcription factors downstream of *A1* include *bHLH25* (Zm00001d044242) and *Mybr97* with *Mybr97* being the most downregulated gene in the interval. Given the transcriptomic data along with the transposon-tagging results, it can reasonably be concluded that *Mybr97* is the gene responsible for the *a3* phenotype.

3.4.3 The entire phenylpropanoid pathway is downregulated with the expression of A3

Anthocyanin accumulation was markedly reduced in normal plants expressing *A3*. B73 produced undetectable levels of anthocyanins in mature husks. 320N and mutant pools expressed between 50- to 100-fold more anthocyanins than *A3* plants with *B1-b* expression (Figure 3.3). All canonical biosynthetic genes associated with anthocyanin production in maize were significantly upregulated in mutant plants in the RNA-seq dataset and confirmed with qPCR (Table 3.1, Table 3.4, and Figure 3.4). The anthocyanidin *O*-methyltransferase (AOMT) that creates peonidin has not been discovered in maize. Only three methyltransferases were upregulated in mutant plants. One gene, designated *O-methyltransferase4* (*Omt4*, Zm00001d052841), resides on chromosome 4 within a QTL for peonidin content from another study (Chatham & Juvik, 2021). Novel biosynthetic genes were also discovered that add to our understanding of the anthocyanin pathway in maize. Three genes are associated with chalcone synthase, which is the first committed step of anthocyanin production. It is known that *Colorless2* (*C2*) is the major gene involved with chalcone synthesis, but in husks it appears *White pollen1* (*Whp1*), and Zm00001d032662 are also involved. The full anthocyanin pathway with canonical and non-canonical genes is shown in Figure 3.4.

In addition to anthocyanin regulation, it appears that *A3* has a role in regulating the entire phenylpropanoid pathway (Table 3.4 and Figure 3.4). B73 produces around 7.2 mg g^{-1} total phenolics in gallic acid equivalents in mature husks. Normal *A3* pools produced slightly higher levels of phenolics in mature husks, most likely due to the activation of the flavonoid pathway by *B1*. 320N and mutant pools produced 3- to 5-fold more phenolics in mature husks than the

64

normal pools (Figure 3.3). The phenylpropanoid pathway is the quintessential secondary metabolite pathway in plants and has several branchpoints that include important compounds like coumarins, flavonoids, lignin, and stilbenes. These compounds are created from the manipulation of *p*-coumaryl-CoA, which is synthesized from phenylalanine in a series of enzymes consisting of phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumaroyl CoA-Ligase (4CL), in order (Boerjan et al., 2003). In husk tissue, *Pal1*, *Pal2*, and *Pal6* were all upregulated in mutant plants. The C4H and 4CL biosynthetic steps were also upregulated. Moreover, five hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferases involved with lignin quality and quantity (Chen & Dixon, 2007), were also significantly differentially regulated. Maysin biosynthetic gene *Salmon silks2* (*Sm2*) was upregulated in husks. *Pericarp1* (*P1*) activates the maysin pathway in silks (Casas et al., 2016) but was not expressed in husk tissue. This indicates *P1* is not sufficient to activate maysin biosynthesis. Overall, through the upregulation of the phenylpropanoid pathway in mutant *a3* plants, plant strength and defense compounds like lignin and maysin could be enhanced.

3.4.4 Vacuolar transport-related genes are differentially regulated in a3 mutants

Anthocyanin transport and storage is an important step in regulating pigment concentration. Anthocyanins are not stable at cytosolic pH, so they need to be transported into the vacuole to confer stable coloration. This effect is demonstrated in *Bronze2* (*Bz2*) mutants that encode a GST. In *bz2* mutants, anthocyanins are unable to enter the vacuole, conferring a titular bronze color to tissues due to oxidative breakdown products of anthocyanins (Marrs et al., 1995). Only one transporter has been characterized to this date that has demonstrated the ability to shuttle anthocyanins into the vacuole: multidrug resistance-associated protein (MRP) family gene *Mrpa3* (Goodman et al., 2004). Both *Bz2* and *Mrpa3* were found to be upregulated in mutant varieties confirming their roles in transport (Table 3.4). However, *Mrpa3* is not sufficient for anthocyanin transport as silencing the transporter does not affect anthocyanin content (Goodman et al., 2004). Several transporters were differentially regulated in mutant plants and might have redundant roles in anthocyanin transport (Table 3.4). Another MRP-like protein, designated Zm00001d045279, has 98% identity to *Mrpa3* (albeit with only 40% coverage) and is upregulated in mutants as well. ATP-binding cassette (ABC) transporters, *Abc1* and *Mrpa8* were both downregulated in mutants, so their association with *A3* is unclear. Multidrug and toxin extrusion (MATE) transporters, Zm00001d006494, Zm00001d017590, Zm00001d032971, and

65

Zm00001d049910 were upregulated and Zm00001d000196 was highly downregulated in mutants.

3.4.5 Expression of regulatory genes are impacted by the reduced A3 expression in mutants

Many transcription factors with predicted roles in anthocyanin regulation were differentially regulated in *a3* husks. A complete list of transcription factors significantly differentially regulated is included in Table 3.5. The central bHLH member expressed in husks, *B1*, was significantly upregulated in mutant plants (Table 3.5 and Figure 3.5B). Unexpectedly, transcript for *A3* could be detected in two of the mutant pools, albeit at a very low level (data not shown). Anthocyanin-related transcription factors typically only associated with aleurone pigmentation—*C1*, *In1*, *Pac1*, and *R1*—were also expressed in husks. *In1* was significantly upregulated in husks (Table 3.5 and Figure 3.5B). Many transcription factors upregulated in mutants were novel genes not known to be associated with anthocyanin synthesis. *Nactf25* (Zm00001d023294) and *Nactf44* (Zm00001d028999) are Nac transcription factors with homology to a peach (*Prunus persica*) transcription factor that confers red pigmentation to peach flesh (Zhou et al., 2015). *Wrky33* (Zm00001d024323) is a transcriptional regulator with homology to Arabidopsis *TRANSPARENT TESTA GLABRA2*, which affects late steps in the anthocyanin biosynthetic pathway (Gonzalez et al., 2016). Although *Pac1* transcript was found in husks, mutant *pac1* plants do differ in their vegetative anthocyanin content (Carey et al., 2004; Selinger & Chandler, 1999). This finding indicates a secondary WD-repeat protein is most likely responsible for gene activation. Genes Zm00001d002541, Zm00001d025699, and Zm00001d008742 were upregulated, but do not have homology to other anthocyanin WD-repeat proteins making them less than ideal candidates.

3.4.6 Photosynthesis-related gene expression is reduced in mutants

Numerous photosynthesis-related genes were differentially regulated in mutants. Sigma factors are nuclear-encoded RNA polymerases that activate synthesis of plastid- and chloroplast-related genes (Beardslee et al., 2002). In mutants, genes *Sig1a*, *Sig2b*, and *Sig8* were all downregulated indicating transcription of photosynthesis-related genes was reduced (Table 3.5). A master regulator of photosynthesis is *Elongated hypoctyl5* (*Hy5*) was downregulated in mutants (Table 3.5). Another transcription factor known to promote chloroplast development specifically in mesophyll tissue, *Golden2-like1* (*Glk1*), was downregulated in mutants (Table 3.5) (Wang et al., 2013). In addition, multiple subunits from Photosystems I and II, carbonic

66

anhydrase, PEP-carboxykinase, PEP-carboxylase, pyruvate orthophosphate dikinase, NADP-malic enzyme, and rubisco activase were downregulated. Reduced A3 expression seems to be associated with decreased expression of many photosynthesis-related genes.

3.4.7 A3 is involved with a range of biological processes related to stress response

Gene Ontology (GO) term enrichment for the 1794 differentially expressed genes found that *A3* is associated with a range of biological and cellular processes (Table 3.6). Secondary metabolic pathways, including the flavonoid and terpenoid pathways, were differentially regulated. In addition, photosynthesis-related genes, including genes from both Photosystem I and II, and photosynthetic regulatory genes were differentially regulated. GO-term enrichment found associations of *A3* with water, high light, osmotic, oxidative, salt, and UV stress. Not only is *A3* involved with abiotic stress, but genes involved with responses to bacteria, fungi, herbivores, and wounding were differentially regulated. GO-terms for systemic acquired resistance and the related salicylic acid pathway were also enriched in this gene expression dataset.

3.4.8 The expression of A3 is highest in leaf sheaths and husks

Quantitative PCR (qPCR) of *A3* across various developmental timepoints in B73 found that the gene is expressed in every tissue analyzed (Figure 3.5A). At the V5 stage, B73 and 320N plants are both green. When the plant matures to V9, 320N begins to flush with color until it is entirely purple except for the leaf blades (Figure 3.1), while B73 stays the normal condition of green. *A3* also follows this same pattern of expression and is more abundant in mature tissues outside of the grain (Figure 3.5A). Transcript was most abundant in husks at the silking stage (R1) followed by leaf sheath tissue at V9 and then V5. A detectable amount of transcript was detected in roots at V1 and pericarp 10 to 20 days after pollination (DAP), but the amount of transcript was minimal compared to mature leaf tissues. The expression of *A3* exceeded that of the housekeeping gene *Elongation factor 1α* for maturing leaf sheath and husk tissue indicating it is a potent transcriptional inhibitor.

3.4.9 A3 is a competitive inhibitor of the anthocyanin regulatory complex

The most probable mechanism for inhibition of the phenylpropanoid pathway is through the sequestration of the MBW complex. To determine the interaction of A3 with bHLH members from maize, a GST-pulldown assay was performed. Anthocyanin bHLH allele *B1-b* and the MIR of this gene (designated B1-MIR) were introduced into a GST fusion vector. Results indicate that

67

A3 is able to bind directly to B1-MIR, but not the GST vector alone (Figure 3.6). No interaction was detected with full-length B1. The interaction with B1-MIR implicates A3 as a competitive inhibitor of the MBW complex.

3.5 Discussion

Mybr97, or *A3*, is a strong negative regulator of the anthocyanin pathway in maize. Mutants display enhanced pigmentation in leaf sheath, husks, and tassels. Anthocyanin content in the husks is sometimes up to 100-fold greater than with a functional *A3*. Pigment content after one extraction yielded 4% pigment on a dry weight basis for 320N (Figure 3.3A). It is difficult to compare anthocyanin production across the literature since extraction protocols vary widely. In a nearly exhaustive anthocyanin extraction of purple husks from a breeding program, anthocyanin content reached up to 19% of total weight (Li et al., 2008). The population developed in this current study contains the weak *B1-b* allele, so pigment production was not optimized. Constitutively expressing *B1* or including naturally strong *B1* alleles with *a3* may be the most effective route to enhance pigment in husks, leaf sheaths, and tassels. Negative regulators *Myb11*, *Myb31*, *Myb42*, *Sro1*, and *In1* were expressed in husks, which indicates that reducing the expression of these genes might also help increase pigment yield in husks. *Myb11*, *Myb31*, and *Myb42* regulate lignin content while *Sro1* is a competitive inhibitor of *Pl1* (Agarwal et al., 2016; Qin et al., 2021; Vélez-Bermúdez et al., 2015). *In1* affects aleurone pigmentation and is not known to be associated with husks, so this is novel regulation for this gene (Burr et al., 1996). The expression of *A3* in pericarp (Figure 3.5A) also indicates that reducing *A3* expression may help enhance purple corn anthocyanins, which are increasingly being used as sources of natural colorants in foods and beverages (Chatham et al., 2019). Increasing anthocyanin content in maize has implications on human health as these pigments are well-characterized antioxidants and have a range of health-promoting benefits (Lao et al., 2017).

Two methods for pinpointing the *A3* gene were utilized in this study. Transposon-tagging has long been used to discover genes since the system is capable of producing knock-outs with visible phenotypes. In this study, a modified *Activator* transposon that encodes transposase, but cannot excise itself (Conrad & Butnell, 2005). This is important so that the *Ds* element is the only mobile element causing phenotypic changes. The transposon-tagging population was developed with the expectation that 4.5% of the plants will have heritable excision events (Conrad & Butnell, 2005). In a pilot experiment utilizing this system, it was estimated that 5000

individuals would need to be generated to insert into an average sized gene 4 cM away (Ahern et al., 2009). Since *Ds* prefers closely linked (4 cM or less) sites (Ahern et al., 2009), then it was necessary to produce a large population for a site expected to be 7.7 cM downstream of *a1-m3::Ds* (Robinett et al., 1995). The visible phenotype assisted in pinpointing *A3* as well. The second method for pinpointing *A3* involved using SNP data from RNA-seq experiments. Bulk segregant RNA-seq analysis is an important tool for discovering genes and is beneficial as a form of low-representation sequencing method for complex genomes like maize (Liu, et al. 2012). The two methods for determining significance in this study left conservatively wide confidence intervals. The inclusion of some wild-type individuals in the mutant pool may have reduced the power to detect significant differences. Despite this, the two methods for detecting significant intervals in this study were able to pinpoint only a handful of genes downregulated in the mutant pools. This demonstrates the effectiveness of bulk segregant RNA-seq analysis to provide a narrow list of candidate genes that traditional genotypic methods are unable to do.

Molecular marker *umc2008* seems to be an appropriate marker for *A3*. The marker is only 30.2 kb away from *A3* according to the B73 RefGen_v4 reference genome and has high variability in simple repeat number among common varieties tested. However, in the W22 v2 genome, the distance is as far as 1.85 Mb (Springer et al., 2018). This large variability may account for the accidental inclusion of normal type *A3* plants in the mutant pools of samples. All samples were checked with *umc2008* and visually for the *a3* phenotype before sampling. Interestingly, some intensely purple plants contained the B73 marker allele. These plants were excluded from the RNA-seq population because they were outliers and their genotypes were uncertain. Conversely, some intensely purple plants had *A3* transcript, indicating that there might be alternative factors involved with intensifying anthocyanins in this population. More evidence for alternative intensification factors was visible in the pericarps of the segregating population. B73 has yellow kernels and a clear pericarp, while 320N contains white kernels with a clear pericarp. The F_1 segregates for yellow and white kernels as expected. However, after selfing, red and purple kernels are fairly common. The source of the cryptic genetic variation shown in this population is not currently understood. It may be due to the interaction of transcription factor alleles from the two parents in the population.

Not only is the involvement of *A3* with the anthocyanin pathway important, but also the upregulation of the phenylpropanoid pathway has implications on structural stability; insect,

69

disease, and wounding defense; and forage quality in maize. Three PAL genes, along with C4H and 4CL were upregulated in *a3* mutants (Table 3.4 and Figure 3.4). In maize, there appears to be multiple copies of PAL with various tissue-specific regulation (Morohashi et al., 2012; Yuan et al., 2019). Since these genes are at the base of the phenylpropanoid pathway, then increased expression implies the potential for increased lignin content and increased structural support (Boerjan et al., 2003; Zheng et al., 2017). These genes have been also associated with increased insect and disease defense through the increased production of salicylic acid (Yuan et al., 2019). Production of anthocyanins is very costly for the plant, so it is no wonder that many modern maize varieties have a functional *A3* and do not produce appreciable amounts of anthocyanins. Knocking out anthocyanin biosynthetic genes along with *A3* might pull vital precursors away from anthocyanin synthesis into, say, lignin, maysin, or salicylic acid synthesis, where they might be more beneficial for crop protection.

The anthocyanin pathway is important in numerous plant species and has its roots as far back as the bryophytes (Markham, 1988). It is reasonable to hypothesize then that regulatory control of the anthocyanin pathway may have also been phylogenetically conserved among angiosperms. Indeed, anthocyanin R3-Myb-like repressor genes appear to be functionally conserved among angiosperms. This study is the first reported characterization of an R3-Myb-like repressor in monocots, but others have been characterized in Arabidopsis, gentian (*Gentiana trifolia*), grape hyacinth (*Muscari spp.*), *Iochroma loxense*, monkeyflower (*Mimulus lewisii*), petunia (*Petunia × hybrida*), poplar (*Populus spp.*), and tomato (*Solanum lycopersicum*) (Chen et al., 2019; Zhang et al., 2020). Utilizing the wealth of knowledge from other plant species could decipher clues into additional regulators of the anthocyanin pathway in maize. In this study, Nac and WRKY transcriptional regulators with homology to eudicot anthocyanin regulators were upregulated (Table 3.5). All the canonical biosynthetic genes involved with anthocyanin production were upregulated as was expected (Table 3.4 and Figure 3.4). The AOMT in maize has not been characterized to this date, but *Omt4* appears to be a good candidate.

The negative association of *A3* with photosynthesis is concerning in terms of plant productivity. However, if one's goal is to create an intensely purple plant, then *A3* is an ideal candidate. The mechanism for reduced photosynthetic capacity in mutants is currently unknown. Previous studies have implicated *Mybr97*, or *A3*, in shade avoidance syndrome (SAS) (Shi et al., 2019; Wang et al., 2016). In these studies, *Gt1*, *Hb53*, and *Mybr97* are positively associated with

70

SAS, while *Hb66* was positively associated with red light. The positive regulators were downregulated and *vice versa* in mutant plants (Table 3.5). In mutants, photosynthesis regulator genes *Hy5*, *Sig1a*, *Sig2b*, and *Sig8* were all downregulated. *Hy5* positively regulates photosynthesis and is known to bind to *C2* and promote anthocyanin synthesis in light (Huai et al., 2020). However, this gene was downregulated in mutants, indicating *A3* might not be regulated by this gene. In Arabidopsis, AtCPC and other CPC-like genes have a range of inhibitory effects and influence trichome, stomata, and flowering development (Zhu et al., 2009). No other phenotypes aside from anthocyanin production were scored in mutant plants in this study. It is possible that A3 associates with other transcription factors that have an impact on photosynthesis and shade sensing.

The activation of *A3* in the absence of the MBW complex in B73 and other normal green varieties implies that *A3* is involved with a broader activation response mechanism independent of anthocyanins. *A3* seems to be involved with many stress response pathways in maize (Table 3.6). A previous study also implicated *A3* in cold, heat, salt, and UV stress (Makarevitch *et al.*, 2015). The main function of anthocyanins besides providing coloration is to respond to ROS. Anthocyanin molecules are antioxidants and accumulate during ROS stress (Naing & Kim, 2021). In order for proper anthocyanin expression, *A3* needs to be downregulated so as to not interfere with transcriptional activation. Promoter elements may infer probable mechanisms for the control of *A3*. As mentioned before, two G-box elements (CACGTG) are in the promoter of *A3*. These elements are known binding sites for anthocyanin bHLH proteins and photosynthetic regulator *Hy5* (Ang et al., 1998; Kong et al., 2012). It is unlikely that *Hy5* is involved since *A3* appears to be activated by shade (Shi et al., 2019; Wang et al., 2016). There is evidence that *B1/R1* may bind to these G-Box elements. In grape hyacinth, the R3-Myb-repressor gene is auto-regulated by *MabBhlh1*, which is the functional equivalent of *B1/R1* (Zhang et al., 2020). It was found that *A3* is typically downregulated in the light, however, while G-Box elements are typically positively associated with light (Ang et al., 1998; Wang et al., 2016). Using the PROMO tool, predicted binding sites for light-associated transcription factors *Gt1*, *Phytochrome interacting factor1*, and *Thioredoxin m1* were found (Farre, 2003; Gao et al., 2015; Xu et al., 2001). Hormonal control of *A3* might be another plausible mechanism of activation. Abscisic acid-responsive cis-element TACGTG and auxin responsive elements TGGTTT and TGTCTC

71

are also present in the promoter of *A3* (Guilfoyle & Hagen, 2007; Song et al., 2018). Future work needs to investigate the role of these binding sites and hormone levels on the activation of *A3*.

The association of A3 with B1-MIR implicates A3 in the role of competitive inhibition of the MBW complex in maize. No DNA-binding experiments were performed, but R3-Myb-like genes are not predicted to have DNA-binding capabilities because they lack the R2-domain typically found in most Myb proteins (Dubos et al., 2010). It is not currently understood why full-length B1 protein could not bind to A3. *In planta*, anthocyanin bHLH proteins are capable of homodimerizing, which may interfere with the association of A3 to B1 (Kong et al., 2012). Future experiments will be aimed at discovering other targets for A3 protein. A previous study showed that AtCPC is able to bind to the MIR of *R1* allele *Leaf color1* (*Lc1*) as demonstrated in a yeast-two hybrid and GST-pulldown assay (Tominaga-Wada et al., 2012). Future experiments should test the MIR of this protein and see if is able to inhibit anthocyanins in either the pericarp or aleurone layers of the grain. Furthermore, protein interaction assay should be performed with photosynthetic bHLH members to see if A3 physically interacts with regulators of photosynthesis.

3.6 Conclusions

A3 was determined to be *Mybr97*, an R3-Myb-like repressor gene, via a transposon-tagging population and bulk segregant RNA-seq analysis. Mutants display enhanced pigmentation in leaf sheath, husks, and tassels. Anthocyanin content in the husks is up to 100-fold greater without a functional *A3*. Transcriptomic analysis found that the entire phenylpropanoid pathway was upregulated in *a3* mutants. Novel transcriptional regulators were discovered that may have associations with the anthocyanin pathway. Numerous stress response related genes were upregulated in mutants which indicates *A3* has implications on plant health. There was a negative association with photosynthesis-related genes, which needs to be investigated further. Future work will be focused on finding other targets for *A3* besides *B1* and determining what regulates its expression. Overall *A3* has a profound effect on maize and breeding for reduced *A3* expression has implications on crop protection and human health.

3.7 Figures and Tables

Figure 3.1: The *A3* locus confers intense anthocyanin accumulation in maize when recessive.

A) B73, the normal *A3* parent of the bulk-segregant RNA-seq population. B) Genetic stock 320N from the Maize Genetics Cooperation Stock Center (Urbana, IL, USA) with the *a3-ref* and active *B1-b* and *Pl1* alleles. C) 320N × W22 *Ac-im;a1-m3::Ds* normal *A3* ear expressing the *B1-b* weak pigmentation phenotype in the husk. D) 320N × W22 normal *A3* condition expressing the *B1-b* weak pigmentation phenotype. E) *De novo* insertion in 320N × W22 *Ac-im;a3-m1::Ds*. F) Husk during the silking stage of the 320N genetic stock demonstrating the *a3* intense purple phenotype. Photo Credit A–C UI Public Affairs: L. Brian Stauffer
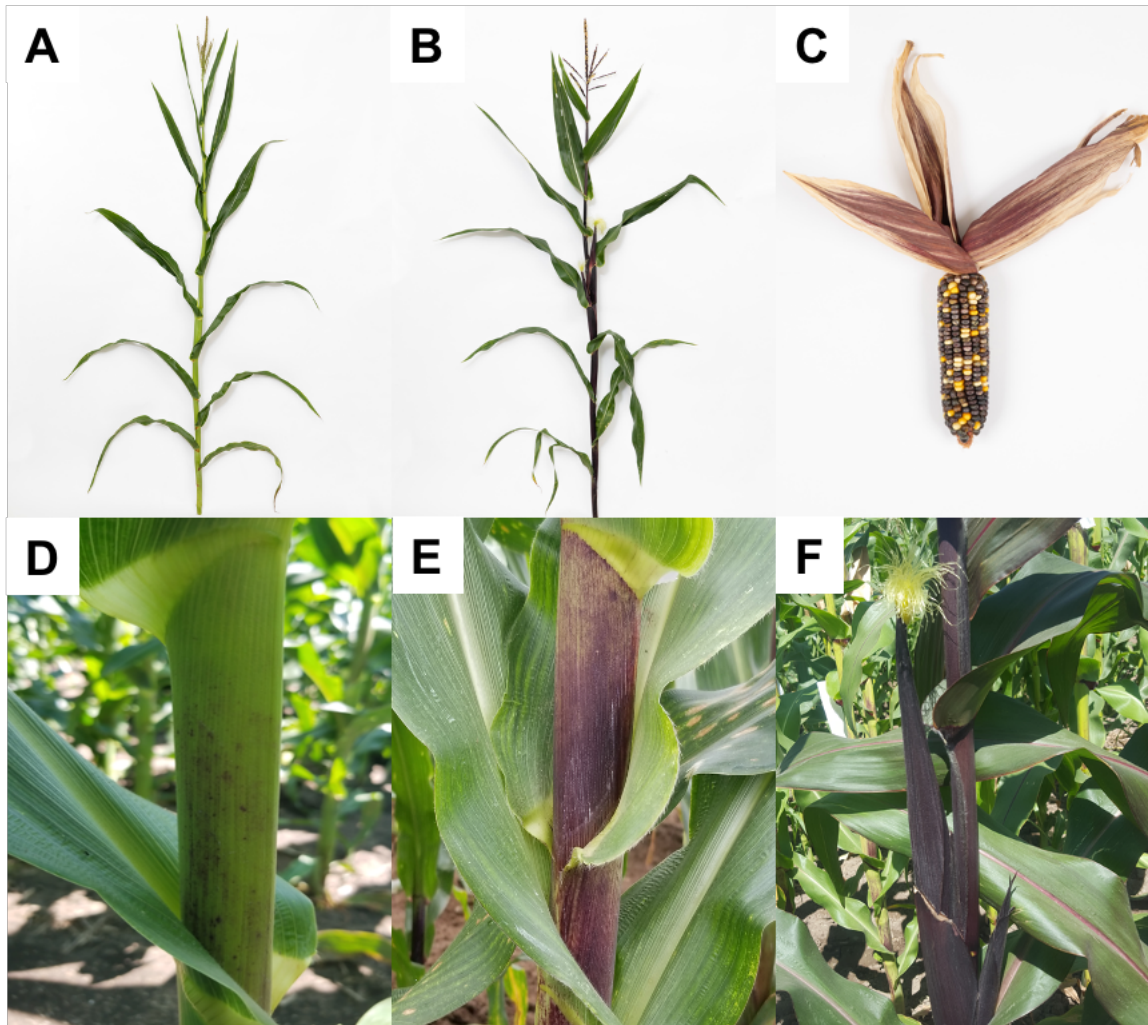
Figure 3.2: Bulk-segregant RNA-seq analysis results

A) The ΔSNP calculation was the difference between scored genotypes where 0 equals the reference B73 allele and 1 equals the 320N allele. A score of 0.5 was a heterozygote. Theoretically the value will approach 1 as a locus becomes fixated in the population. B) The F_{st} index is a measure of the heterozygosity within the mutant pools and the wild type pools compared to total heterozygosity at a SNP. A common threshold for this value is 0.25 for populations becoming fixated for an allele (Hartl & Clark, 1997). Each measure was smoothed within a window of 50 SNPs and step size of 10 SNPs. Lines indicate the Loess smoothing parameter calculated in the R package "ggplot2" (Wickham, 2009).
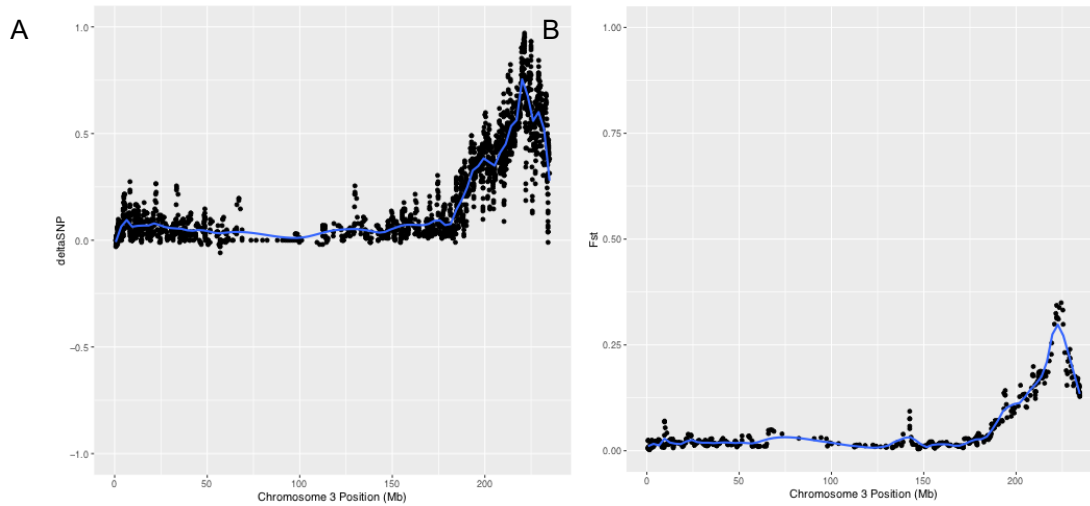
Figure 3.3: Anthocyanin and phenolic content of normal and mutant *A3* genetic stocks and their progeny.

A) Total anthocyanin content of the bulk-segregant RNA-seq population mutant and normal pools and their parental stocks B73 and 320N. Values are expressed in mg of cyanidin 3-glucoside (C3G) equivalents per g dry weight. B) Total phenolic content of the samples in A expressed in gallic acid equivalents (GAE) per g of dry weight. Error bars represent standard errors of three biological replicates.
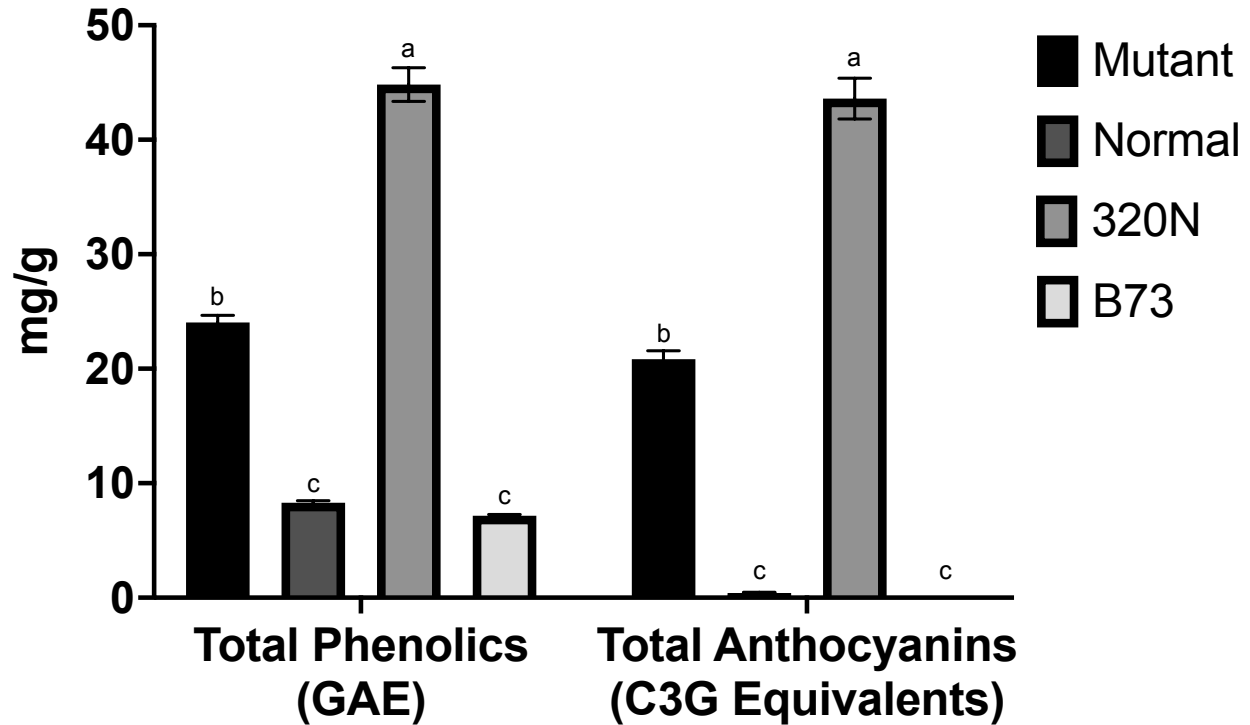
Figure 3.4: The anthocyanin pathway in maize with RNA-seq expression values.

Shown is the anthocyanin biosynthetic pathway in maize with gene functions on the top of arrows and gene models or symbols of genes predicted to have those functions. Predicted genes are colored according to their $\log_2$-fold expression values of mutant vs. normal pools of samples. Abbreviations: 4CL, 4-Coumarate:CoA Ligase; AAT, Anthocyanin Acyltransferase; ANS, Anthocyanidin Synthase; AOMT, Anthocyanidin O-Methyltransferase; C4H, Cinnamate-4-Hydroxylase; CHI, Chalcone Isomerase; CHS, Chalcone Synthase; DFR, Dihydroflavonol 4-Reductase; F3'H, Flavonoid 3'-Hydroxylase; F3H, Flavanone 3-Hydroxylase; GST, Glutathione S-Transferase; PAL, Phenylalanine Ammonia Lyase; UGT, UDP-Glucosyltransferase.
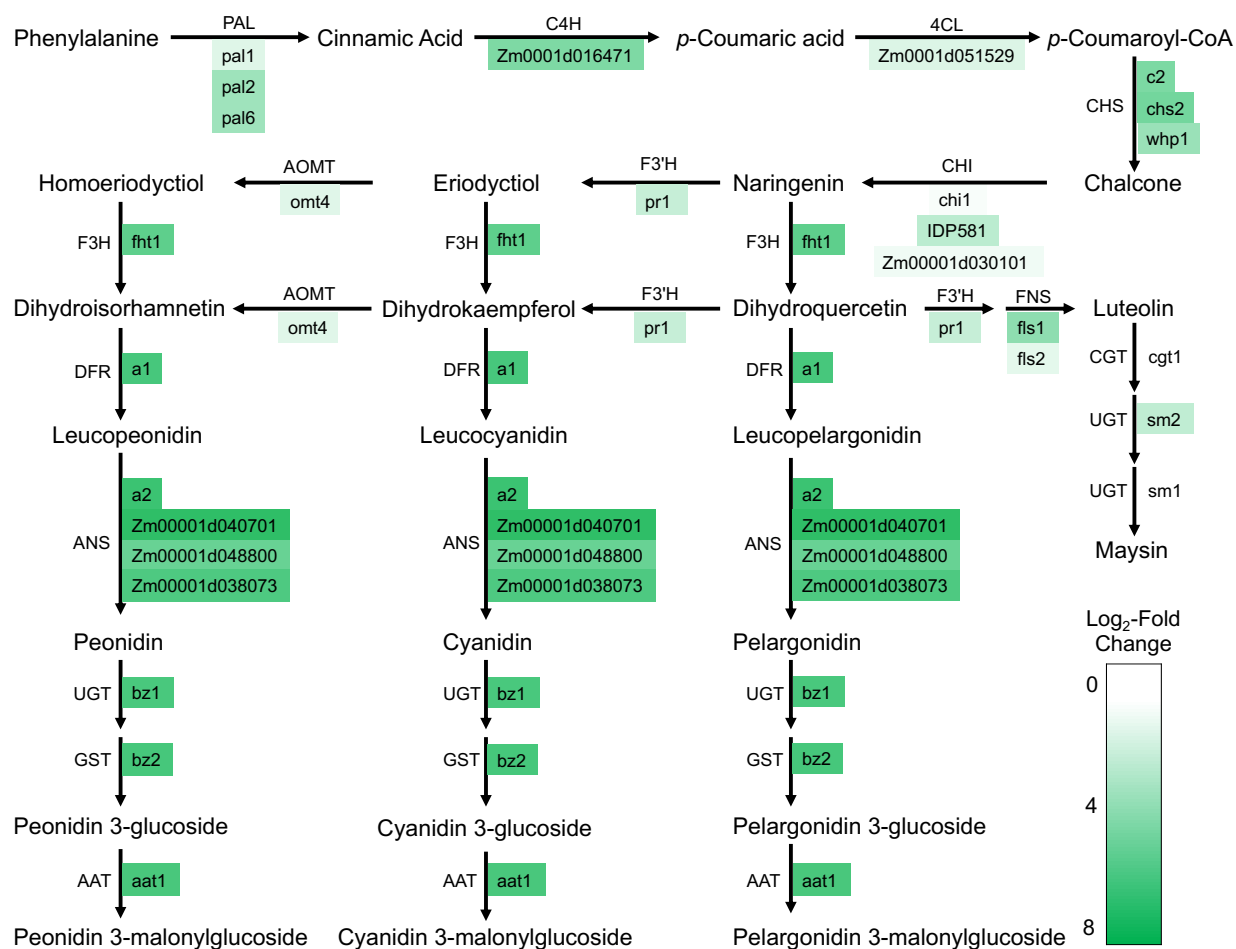
Figure 3.5: Relative expression of *A3* across development and validation of important differentially expressed genes from bulk-segregant RNA-seq analysis.

A) Expression of *A3* in various timepoints of development normalized to maize *Elongation factor 1α*. B) Relative expression values of differentially expressed genes determined by qPCR. All values are relative to the normal *A3* pool from the bulk-segregant RNA-seq population and standardized to *Elongation factor 1α* and *β-Tubulin*. Error bars represent 95% confidence intervals around the mean.
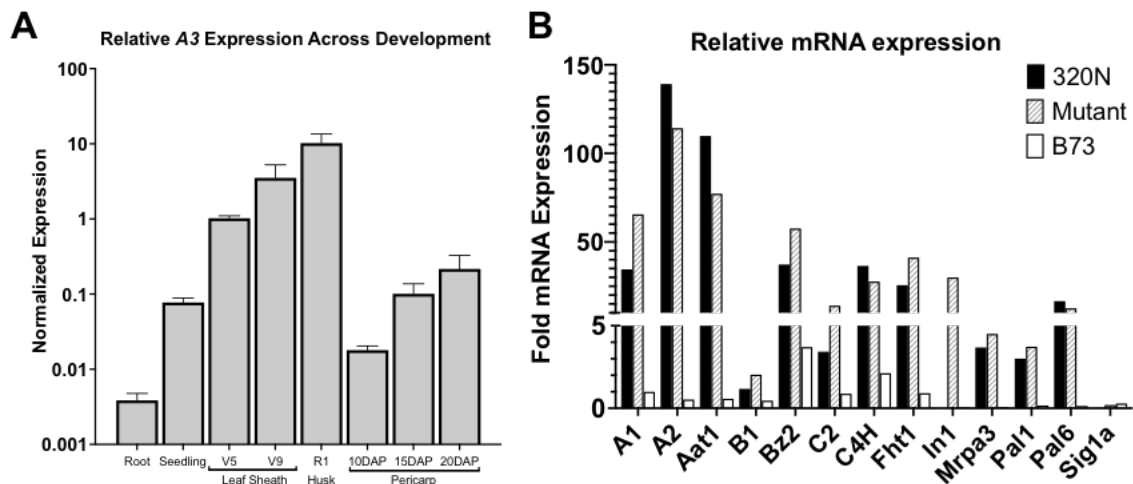
Figure 3.6: GST-pulldown assay.

Affinity tag purification of maize transcription factors. Prey A3 protein was purified with His-Pure Ni-NTA resin (ThermoFisher Scientific, LLC). Remaining bait proteins were purified using GST resin (Genscript). Lanes 2, 3, and 5 on each gel are the isolated proteins. Lanes 4 and 6 were equilibrated with A3 protein. Lane 1 contains Precision Plus Protein Dual Color Standards (Bio-Rad Laboratories, Inc.). GST = Glutathione S-Transferase, MIR = Myb-interacting Region.
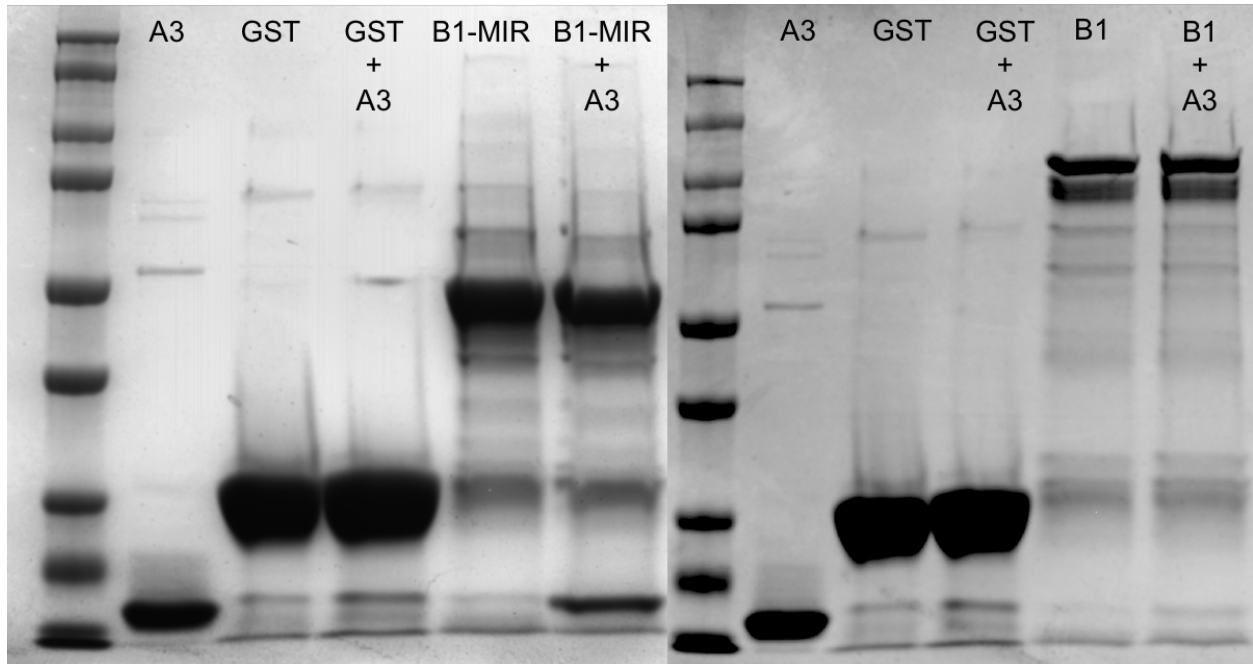
Table 3.1: Primer Sets

PRIMER	TARGET	STRAND	SEQUENCE	SOURCE
MP-A1-F	A1	+	CCAAAGACCCTGAGAATGAGGTA	This study
MP-A1-R	A1	-	ACGTCGGTCCAGCTTTCCT	This study
MP-A2-F	A2	+	GCACAAAGCAAAAGCAGAGGAA	This study
MP-A2-R	A2	-	TCGTAGTAACGTAGGAGCATGAA	This study
MP-A3-F	A3	+	AGACAATGATCGCCCTGAAG	This study
MP-A3-R	A3	-	AAACCGGTAACATCGTAGGC	This study
MP-A3-END	A3	-	tatcccaagcttTCATCTTTCCTGGTATTTTTTGGACCA	This study
MP-A3-START	A3	+	accttagaattcATGGATAGCAGCAGTGGTAGCCAG	This study
MP-AAT1-F	Aat1	+	GCGTCTACGACCTCGACTTT	This study
MP-AAT1-R	Aat1	-	AGAGAGATACCCACCTCCACG	This study
MP-B1-F	B1	+	CCTCGTTCTCAAGTCGTTGGT	This study
MP-B1-R	B1	-	GGAGCCCACACAGACTTTCT	This study
MP-B1-END	B1	-	tatcccaagcttTCATCGCTTCCCTATAGCTTTACGA	This study
MP-B1-MIR-END	B1	-	tcattgaagcttAGCTCGGCTCTTTCGAGTATG	This study
MP-B1-START	B1	+	acgttagaattcgtATGGCGCTCTCAGTCTCAGCTTC	This study
MP-BZ2-F	Bz2	+	GCCTTCATCGACGACAAGTTT	This study
MP-BZ2-R	Bz2	-	GTCGCCACCGGAGAAGAAAG	This study
MP-C2-F	C2	+	CACCAAGAGCGAGCACCT	This study
MP-C2-R	C2	-	CCAGGAACTCCTCCGTCAG	This study
MP-C4H-F	C4H	+	CTCGGCGGCTACACCATC	This study
MP-C4H-R	C4H	-	ACGGCAGGAAGCGGAAGT	This study
JGP2	Ds	+	CCGGTTCCCGTCCGATTTCG	Ahern, et al. (2009)
LC18	Ds	+	CCTTGTTTTGATTGGCTGCTA	Ahern, et al. (2009)
LC24	Ds	-	TTGTTGCAGCAGCAATAACAGCAT	Ahern, et al. (2009)
LC45	Ds	-	GTGCTGTACTGCTGTGACTTGTG	Ahern, et al. (2009)
EF1A	Elfa9	+	ACATACCCACGCTTCAGATCCT	Lin, et al. (2014)
EF1A	Elfa9	-	TGGGCCTACTGGTCTTACTACTGA	Lin, et al. (2014)
MP-FHT1-F	Fht1	+	GGACCTGGATCACCGTGC	This study
MP-FHT1-R	Fht1	-	TTCTTGAACCTGCCGTTGCT	This study
ZMMRP3	Mrpa3	+	CATCTCTGTGGAAAGGGTTC	Goodman, et al. (2004)
ZMMRP3	Mrpa3	-	CTGTTCGCCCAACTATACCG	Goodman, et al. (2004)
ZMPAL-QRT- ZM00001D017274	Pal1	+	AAGGAGAAGAGGAGGGAGGG	Yuan, et al. (2019)
ZMPAL-QRT- ZM00001D017274	Pal1	-	GAAGAAAGAGCAACGCCACA	Yuan, et al. (2019)
ZMPAL-QRT- ZM00001D003015	Pal6	+	GGGAGATGAAGTGAAGTGAAGACA	Yuan, et al. (2019)
ZMPAL-QRT- ZM00001D003015	Pal6	-	GACATAAAACAAAGGAACAGTGAGC	Yuan, et al. (2019)

79

Table 3.1 (Cont.)

PRIMER	TARGET	STRAND	SEQUENCE	SOURCE
T7-pro	pET-30a(+)	+	TAATACGACTCACTATAGGG	MilliporeSigma
T7-term	pET-30a(+)	-	GCTAGTTATTGCTCAGCGG	MilliporeSigma
pGEX5	pGSTag	+	GGGCTGGCAAGCCACGTTTGGTG	Ron, D. & Dressler, H. (1992)
pGEX3	pGSTag	-	CCGGGAGCTGCATGTGTCAGAGG	Ron, D. & Dressler, H. (1992)
Cloning Analysis Forward	pMiniT 2.0	+	ACCTGCCAACCAAAGCGAGAAC	NEB
Cloning Analysis Reverse	pMiniT 2.0	-	TCAGGGTTATTGTCTCATGAGCG	NEB
MP-Sig1a-F	*Sig1a*	+	GCACCAAAGATCCAAGCTGA	This study
MP-Sig1a-R	*Sig1a*	-	TCCCTCGCAAGAAACGACTC	This study
β-TUB	*Tub4*	+	GTCACACACACTCGACTTCACG	Lin, et al. (2014)
β-TUB	*Tub4*	-	CTACCTCACGGCATCTGCTATGT	Lin, et al. (2014)
umc2008	*umc2008*	+	GTGGACTACTCTCGCTTCGCTTT	Andorf, et al. (2010)
umc2008	*umc2008*	-	CGTGGACGTACTCGATTAGTTTGTT	Andorf, et al. (2010)

Table 3.2: RNA-seq Read Summary Statistics

Pool	Rep	Raw Reads	Clean Reads	% Clean Reads	Uniquely Mapped	% Uniquely Mapped	Counted	% Counted	% of Raw
Normal	3	86,793,741	86,601,516	99.78%	59,714,680	68.95%	53,985,443	90.41%	62.20%
Normal	2	119,172,038	118,986,923	99.84%	83,088,241	69.83%	78,275,663	94.21%	65.68%
Normal	1	95,069,301	94,967,015	99.89%	66,770,764	70.31%	60,954,564	91.29%	64.12%
Mutant	3	85,262,985	85,135,257	99.85%	58,668,962	68.91%	54,384,344	92.70%	63.78%
Mutant	2	87,808,953	87,732,031	99.91%	61,400,777	69.99%	58,789,904	95.75%	66.95%
Mutant	1	86,211,621	86,118,861	99.89%	59,554,659	69.15%	55,160,872	92.62%	63.98%

Table 3.3: Genes Within the ΔSNP A3 Significance Interval

GENE SYMBOL	LOG$_2$-FC	FDR *P*-VALUE	ARABIDOPSIS ANNOTATION
mybr97	-6.519	5.52E-03	CAPRICE-like MYB3
mybr41	-3.564	6.35E-19	RAD-like 6
Zm00001d044186	-3.173	1.98E-109	cysteine-rich RLK (RECEPTOR-like protein kinase) 8
csu845	-2.890	1.81E-06	phy rapidly regulated 2
Zm00001d044156	-2.737	2.61E-04	cytochrome P450, family 72, subfamily A, polypeptide 14
Zm00001d044105	-2.515	1.60E-17	
Zm00001d044202	-2.391	3.94E-06	alpha/beta-Hydrolases superfamily protein
Zm00001d044119	-2.278	4.37E-04	Cystathionine beta-synthase (CBS) family protein
Zm00001d044183	-2.239	1.72E-18	Stabilizer of iron transporter SufD / Polynucleotidyl transferase
Zm00001d044109	-2.231	2.85E-14	tonoplast dicarboxylate transporter
Zm00001d044146	-1.894	2.49E-05	cytochrome P450, family 72, subfamily A, polypeptide 14
cah1	-1.853	9.77E-39	
bhlh25	-1.671	9.51E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
cyp11	-1.471	1.05E-18	cytochrome P450, family 72, subfamily A, polypeptide 15
AY110540	-1.363	1.06E-21	Ribosomal protein L31
Zm00001d044310	-1.323	9.30E-27	Protein of unknown function (DUF1118)
Zm00001d044226	-1.318	1.37E-17	
Zm00001d044103	1.290	1.93E-13	adenylate kinase 1
Zm00001d044230	1.293	4.04E-08	Protein of unknown function, DUF538
hsftf23	1.305	7.76E-03	heat shock transcription factor C1
Zm00001d044244	1.323	2.51E-10	Protein of unknown function (DUF3511)
Zm00001d044321	1.353	8.85E-04	Uroporphyrinogen decarboxylase
gar1	1.439	1.49E-09	
Zm00001d044320	1.527	2.54E-06	Protein of unknown function (DUF594)
Zm00001d044219	1.570	1.83E-11	
Zm00001d044187	1.911	5.00E-14	domains rearranged methylase 1
Zm00001d044262	2.136	1.36E-39	
yz1	2.443	6.52E-08	SNF1-related protein kinase regulatory subunit gamma 1
Zm00001d044318	2.466	5.39E-29	voltage dependent anion channel 1
cbl10	2.780	5.61E-28	
Zm00001d044131	2.860	3.87E-05	
a1	6.276	1.46E-35	dihydroflavonol 4-reductase

Table 3.4: Differentially Expressed Phenylpropanoid Pathway Genes

GENE SYMBOL	LOG$_2$-FC	FDR P-VALUE	ARABIDOPSIS ANNOTATION
a1	6.28	1.46E-35	dihydroflavonol 4-reductase
a2	6.68	7.02E-23	leucoanthocyanidin dioxygenase
aat1	6.24	1.25E-20	HXXXD-type acyl-transferase family protein
abc1	-1.55	4.57E-39	Protein kinase superfamily protein
bz1	6.22	3.79E-22	UDP-Glycosyltransferase superfamily protein
bz2	6.31	3.27E-24	Glutathione S-transferase family protein
c2	4.71	1.44E-110	Chalcone and stilbene synthase family protein
cgt1	-1.52	4.19E-07	UDP-Glycosyltransferase superfamily protein
chi1	1.21	5.66E-11	Chalcone-flavanone isomerase family protein
chls2	4.96	3.35E-07	Chalcone and stilbene synthase family protein
cl4921_1	-1.31	7.62E-03	cinnamyl alcohol dehydrogenase 9
fht1	5.61	2.31E-14	flavanone 3-hydroxylase
fls1	4.20	8.23E-17	flavonol synthase 1
fls2	1.78	7.53E-16	flavonol synthase 1
hct3	-2.53	7.73E-05	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
hct4	2.32	2.87E-10	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
hct7	-1.84	1.50E-25	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
IDP158	2.89	3.96E-14	Chalcone-flavanone isomerase family protein
IDP582	-3.82	3.66E-42	flavonol synthase 1
mrpa3	2.35	9.24E-57	multidrug resistance-associated protein 14
mrpa8	-3.04	0.00259	multidrug resistance-associated protein 3
omt4	1.90	3.05E-20	O-methyltransferase
pal1	1.92	8.81E-25	PHE ammonia lyase 1
pal2	3.73	7.84E-51	phenylalanine ammonia-lyase 2
pal6	3.70	1.93E-21	PHE ammonia lyase 1
pr1	2.62	6.31E-26	Cytochrome P450 superfamily protein
sm2	2.78	3.73E-13	UDP-Glycosyltransferase superfamily protein
whp1	3.85	8.49E-13	Chalcone and stilbene synthase family protein
Zm00001d004443	-1.40	1.53E-10	cinnamyl alcohol dehydrogenase 9
Zm00001d006494	1.57	9.85E-08	MATE efflux family protein
Zm00001d016471	4.67	2.83E-12	cinnamate-4-hydroxylase
Zm00001d017590	1.88	0.000213	MATE efflux family protein
Zm00001d026109	5.24	7.77E-68	polyphenol oxidase putative
Zm00001d026336	3.27	1.72E-17	polyphenol oxidase
Zm00001d030101	1.42	4.81E-03	Chalcone-flavanone isomerase family protein
Zm00001d032971	1.24	2.53E-10	MATE efflux family protein
Zm00001d038073	6.08	1.18E-15	leucoanthocyanidin dioxygenase

Table 3.4 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR *P*-VALUE	ARABIDOPSIS ANNOTATION
Zm00001d040701	7.12	1.79E-49	leucoanthocyanidin dioxygenase
Zm00001d045091	-1.31	1.93E-03	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
Zm00001d045279	2.00	4.56E-16	multidrug resistance-associated protein 14
Zm00001d048800	5.31	1.91E-19	leucoanthocyanidin dioxygenase
Zm00001d049704	2.64	6.59E-16	HXXXD-type caffeoyl-CoA:shikimate O-(hydroxycinnamoyl)transferase
Zm00001d049910	1.42	3.10E-24	MATE efflux family protein
Zm00001d051529	2.01	1.24E-21	4-coumarate:CoA ligase 3

Table 3.5: Differentially Expressed Transcription Factors

GENE SYMBOL	LOG$_2$-FC	FDR *P*-VALUE	ARABIDOPSIS ANNOTATION
AW330564	-1.2453	3.98E-18	Phototropic-responsive NPH3 family protein
b1	1.2032	5.66E-10	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh101	4.9230	4.88E-09	FER-like regulator of iron uptake
bhlh109	-1.5490	8.14E-06	Basic helix-loop-helix (bHLH) DNA-binding family protein
bhlh116	1.8695	1.45E-07	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh117	-1.3726	9.85E-13	phytochrome interacting factor 3
bhlh135	-1.2985	7.14E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh153	-1.5028	5.84E-09	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh155	2.7547	3.36E-06	basic helix-loop-helix (bHLH) DNA-binding family protein
bhlh156	1.6099	6.89E-03	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh161	3.0311	1.23E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh25	-1.6706	9.51E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh44	-1.2230	8.32E-03	basic helix-loop-helix (bHLH) DNA-binding family protein
bhlh57	2.0877	5.99E-20	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh69	1.4195	3.43E-03	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bhlh83	-1.2605	5.06E-08	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
bzip109	1.7586	1.59E-05	bZIP transcription factor family protein
bzip111	-1.6216	2.95E-18	basic leucine zipper 9
bzip13	4.4544	1.54E-11	basic leucine-zipper 42
bzip26	1.7600	2.89E-03	basic leucine-zipper 58
bzip27	1.8613	1.70E-09	TGACG motif-binding factor 4
bzip50	3.1669	1.89E-04	ABA-responsive element binding protein 3
bzip75	1.7192	3.68E-04	abscisic acid responsive element-binding factor 1
bzip79	1.9157	1.02E-08	bZIP transcription factor family protein
bzip95	1.2072	1.64E-15	G-box binding factor 4
ca5p11	-1.3355	7.53E-15	nuclear factor Y, subunit C4
ca5p12	-2.5303	6.41E-08	nuclear factor Y, subunit C4
cca1	-1.4925	6.28E-11	Homeodomain-like superfamily protein
cchh25	1.4246	1.09E-04	C2H2-like zinc finger protein
cl31479_3	-7.9579	6.27E-03	Zinc-binding dehydrogenase family protein
col6	1.2245	1.35E-03	B-box type zinc finger protein with CCT domain
crr1	-1.2102	1.24E-08	response regulator 6
dbb1	-1.4069	1.08E-23	B-box zinc finger family protein
dbb6	-1.2016	1.71E-19	salt tolerance homologue
dbp4	1.7634	8.42E-04	dehydration response element B1A
dof10	1.6939	7.67E-03	Dof-type zinc finger DNA-binding family protein

Table 3.5 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR P-VALUE	ARABIDOPSIS ANNOTATION
ereb117	-2.2995	1.73E-04	Integrase-type DNA-binding superfamily protein
ereb134	-1.6017	2.86E-03	ethylene responsive element binding factor 1
ereb181	1.9126	1.67E-09	Integrase-type DNA-binding superfamily protein
ereb190	-1.3910	3.49E-04	Integrase-type DNA-binding superfamily protein
ereb34	1.4327	1.53E-07	related to AP2 4
ereb50	-1.2407	8.67E-03	cytokinin response factor 4
ereb7	1.2334	3.94E-06	ethylene-responsive element binding protein
fha9	-2.9092	1.60E-53	SMAD/FHA domain-containing protein
flz18	-1.8814	1.05E-14	Protein of unknown function (DUF581)
flz19	-1.6310	8.69E-03	Protein of unknown function (DUF581)
gata15	-2.5206	6.44E-05	GATA type zinc finger transcription factor family protein
gata29	-1.3810	1.66E-14	GATA type zinc finger transcription factor family protein
gata32	1.4117	6.15E-05	GATA transcription factor 12
glk1	-2.3223	2.78E-09	GOLDEN2-like 2
glk43	-1.5599	7.10E-07	myb-related protein 1
glk44	1.9356	2.53E-25	Homeodomain-like superfamily protein
glk8	1.4499	1.70E-28	Homeodomain-like superfamily protein
gras73	2.3856	1.64E-03	GRAS family transcription factor
grftf10	1.7750	1.39E-06	
grftf6	-1.7566	3.17E-05	
GRMZM2G154580	-1.4381	4.60E-11	CONSTANS-like 4
GRMZM2G416184	1.8196	3.61E-09	
gt1	1.6020	9.79E-09	
hb24	-3.3076	2.49E-03	Homeobox-leucine zipper protein family
hb26	-1.5530	1.48E-14	Tetratricopeptide repeat (TPR)-like superfamily protein
hb38	-1.2184	3.70E-15	POX (plant homeobox) family protein
hb42	1.9685	2.37E-03	Homeobox-leucine zipper protein family
hb46	-2.0712	3.15E-05	homeobox-leucine zipper protein 17
hb53	-1.2624	5.05E-09	homeobox-leucine zipper protein 3
hb66	2.0248	7.08E-23	homeobox 7
hb98	1.4043	1.10E-06	Homeobox-leucine zipper protein family
hsftf23	1.3045	7.76E-03	heat shock transcription factor C1
HY5/bzip61	-1.3738	2.76E-14	Basic-leucine zipper (bZIP) transcription factor family protein
iaa2	-2.1245	7.40E-16	AUX/IAA transcriptional regulator family protein
iaa44	1.4427	1.60E-15	indole-3-acetic acid inducible 33
IDP1459	-1.5641	2.46E-11	chloroplast import apparatus 2

Table 3.5 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR P-VALUE	ARABIDOPSIS ANNOTATION
IDP80	-1.8177	5.67E-51	one helix protein
in1	7.0569	1.62E-16	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
iqd11	1.5121	3.29E-11	IQ-domain 31
kn1	3.9733	1.28E-04	KNOTTED-like from Arabidopsis thaliana
lbd20	1.3709	8.83E-04	LOB domain-containing protein 13
lbd3	1.4388	7.00E-04	LOB domain-containing protein 18
lbd33	1.3093	3.96E-05	LOB domain-containing protein 41
lhy1	-1.6964	3.80E-12	Homeodomain-like superfamily protein
mads75	-2.3683	5.65E-07	AGAMOUS-like 61
myb105	2.0060	2.29E-06	MYB-like 102
myb137	4.2901	4.82E-19	myb domain protein 62
myb139	1.2745	1.08E-09	myb domain protein 70
myb155	-1.3230	1.10E-03	myb domain protein 98
myb64	2.4200	3.19E-08	myb domain protein 4
myb67	1.4658	2.32E-06	myb domain protein 4
myb85	1.6726	6.42E-07	myb domain protein 4
myb86	2.7391	5.10E-07	myb domain protein 79
mybr20	1.5935	5.60E-04	
mybr41	-3.5643	6.35E-19	RAD-like 6
mybr47	-5.6649	2.20E-09	RAD-like 1
mybr48	-1.8711	1.81E-07	Homeodomain-like superfamily protein
mybr55	1.4097	1.06E-08	
mybr78	-3.3220	1.03E-14	RAD-like 6
mybr87	-4.1320	2.52E-19	RAD-like 1
mybr89	-6.2744	2.39E-05	RAD-like 1
mybr97	-6.5192	5.52E-03	CAPRICE-like MYB3
myo4	-3.0415	3.84E-15	Myosin family protein with Dil domain
nactf108	1.3012	1.84E-15	NAC domain containing protein 1
nactf122	1.2894	1.70E-10	NAC-like, activated by AP3/PI
nactf131	1.4897	1.18E-05	NAC 007
nactf20	1.2963	5.68E-14	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
nactf24	1.5146	9.57E-04	NAC domain containing protein 38
nactf25	1.9307	2.86E-32	NAC domain containing protein 47
nactf27	2.6385	4.55E-03	NAC 007
nactf36	1.7516	3.04E-05	NAC domain containing protein 74
nactf43	1.6061	7.72E-06	NAC domain containing protein 71

Table 3.5 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR *P*-VALUE	ARABIDOPSIS ANNOTATION
nactf44	2.0091	4.57E-39	NAC domain containing protein 47
nactf51	1.4092	2.22E-11	NAC domain containing protein 74
nactf63	1.7461	6.98E-11	NAC domain containing protein 41
nactf65	2.4635	4.02E-04	NAC domain containing protein 74
nactf76	1.2850	1.84E-10	xylem NAC domain 1
nactf96	3.2946	4.84E-12	NAC domain containing protein 42
nlp15	2.2868	2.45E-12	Plant regulator RWP-RK family protein
npy5	-1.5479	8.72E-26	Phototropic-responsive NPH3 family protein
obf3	1.7881	2.46E-18	TGACG motif-binding factor 6
ocl2	1.3135	9.93E-03	homeodomain GLABROUS 1
ofp2	1.9622	2.25E-06	ovate family protein 12
ofp23	2.6335	7.79E-16	Ovate family protein
ofp27	-2.6793	1.06E-12	ovate family protein 13
ofp6	5.2427	1.01E-14	ovate family protein 14
ofp8	-3.0111	9.65E-04	ovate family protein 8
ohp3	-1.7959	2.02E-25	one-helix protein 2
pco098770(707)	-1.2426	6.12E-11	CCCH-type zinc fingerfamily protein with RNA-binding domain
pco147721a	1.3046	3.38E-20	GroES-like zinc-binding dehydrogenase family protein
prh11	1.4173	8.40E-06	Protein phosphatase 2C family protein
prh20	2.5647	7.40E-13	highly ABA-induced PP2C gene 3
prh6	1.3918	2.85E-08	highly ABA-induced PP2C gene 2
prh7	1.2404	1.38E-04	highly ABA-induced PP2C gene 3
prh8	3.0476	2.88E-15	highly ABA-induced PP2C gene 3
rcc4	-1.4235	1.59E-08	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
sig2B	-1.6807	9.77E-29	RNApolymerase sigma subunit 2
sig8	-1.8135	7.00E-39	sigma factor E
srs1	1.6544	5.78E-05	Lateral root primordium (LRP) protein-related
tcptf23	-1.5847	2.23E-05	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 5
thx12	1.4173	1.50E-10	
thx19	2.3420	4.42E-14	Duplicated homeodomain-like superfamily protein
thx21	-3.1384	2.86E-06	Duplicated homeodomain-like superfamily protein
thx27	1.4387	9.92E-09	Fasciclin-like arabinogalactan family protein
umc1028	2.3974	3.05E-10	zinc induced facilitator-like 1
umc1029	1.6159	5.48E-18	Phototropic-responsive NPH3 family protein
vq23	1.5910	2.56E-03	VQ motif-containing protein
vq31	1.4707	1.75E-07	VQ motif-containing protein

Table 3.5 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR P-VALUE	ARABIDOPSIS ANNOTATION
vq51	2.2508	3.63E-04	MAP kinase substrate 1
vq60	3.1255	1.82E-04	VQ motif-containing protein
wri1	2.3525	7.92E-09	Integrase-type DNA-binding superfamily protein
wrky119	4.2185	5.24E-21	WRKY DNA-binding protein 56
wrky19	1.4556	7.22E-03	WRKY DNA-binding protein 64
wrky33	4.6176	1.31E-37	WRKY family transcription factor family protein
wrky40	-2.1688	5.13E-08	
wrky63	-1.8027	1.06E-03	WRKY DNA-binding protein 70
wrky91	-1.7216	1.43E-03	WRKY family transcription factor
wrky96	-4.2411	5.48E-03	WRKY DNA-binding protein 49
yab10	2.6898	2.94E-06	Plant-specific transcription factor YABBY family protein
yab5	1.2245	1.00E-10	Plant-specific transcription factor YABBY family protein
zhd10	1.2988	9.92E-03	mini zinc finger 2
zhd13	1.5430	7.74E-06	homeobox protein 28
zhd21	-1.6702	2.17E-05	homeobox protein 31
zhd5	-1.3167	5.61E-08	zinc finger homeodomain 1
zhd9	2.2033	1.11E-10	mini zinc finger 2
zim28	-1.2245	2.91E-05	
Zm00001d002544	1.4590	2.14E-04	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
Zm00001d003297	1.6918	8.49E-03	Phototropic-responsive NPH3 family protein
Zm00001d005837	-1.3236	1.29E-07	zinc finger protein 8
Zm00001d009022	-2.0102	2.60E-04	NmrA-like negative transcriptional regulator family protein
Zm00001d009877	-1.3533	2.44E-19	plastid transcriptionally active 16
Zm00001d010380	1.3816	1.27E-17	Zinc finger C-x8-C-x5-C-x3-H type family protein
Zm00001d011501	1.2614	5.09E-03	E2F/DP family winged-helix DNA-binding domain
Zm00001d012016	1.3140	1.69E-13	HCP-like superfamily protein with MYND-type zinc finger
Zm00001d013118	-2.5388	2.79E-05	C2H2-like zinc finger protein
Zm00001d014515	-1.6644	1.81E-05	DHHC-type zinc finger family protein
Zm00001d015268	-1.7419	7.04E-21	chloroplast import apparatus 2
Zm00001d015517	-1.3125	4.19E-06	WRKY family transcription factor
Zm00001d016898	-1.7139	6.32E-14	Phototropic-responsive NPH3 family protein
Zm00001d018203	3.7081	5.80E-19	WRKY DNA-binding protein 9
Zm00001d018247	-1.8119	1.64E-09	zinc finger protein 8
Zm00001d021025	-1.4771	3.84E-15	Oxidoreductase, zinc-binding dehydrogenase family protein
Zm00001d022552	-1.2769	7.28E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Zm00001d028615	1.6431	4.56E-04	pol-like 5

Table 3.5 (Cont.)

GENE SYMBOL	LOG$_2$-FC	FDR *P*-VALUE	ARABIDOPSIS ANNOTATION
Zm00001d028664	2.0881	2.91E-04	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Zm00001d029898	-1.3969	2.98E-14	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Zm00001d030063	-1.5617	5.24E-07	heat shock transcription factor A2
Zm00001d030347	1.3730	4.82E-16	GroES-like zinc-binding dehydrogenase family protein
Zm00001d030958	-1.5959	7.59E-16	TRAF-like superfamily protein
Zm00001d032208	1.7936	5.79E-04	NAC domain containing protein 100
Zm00001d034359	-1.3658	4.40E-07	zinc-finger protein 1
Zm00001d034699	-1.4290	1.23E-11	plastid transcriptionally active 13
Zm00001d035195	-2.5330	5.19E-03	LSD1 zinc finger family protein
Zm00001d035434	1.6924	8.17E-05	Zinc finger (C3HC4-type RING finger) family protein
Zm00001d041960	1.2127	2.36E-03	Homeodomain-like superfamily protein
Zm00001d042779	4.0149	4.39E-13	Basic-leucine zipper (bZIP) transcription factor family protein
Zm00001d042781	1.7929	6.70E-10	
Zm00001d043159	-1.3933	5.13E-07	Calmodulin-binding transcription activator protein with CG-1 and Ankyrin domains
Zm00001d043707	-1.2935	1.90E-14	plastid transcriptionally active7
Zm00001d044525	-1.4429	2.85E-22	thylakoid-associated phosphatase 38
Zm00001d044526	2.2683	7.74E-03	Ran BP2/NZF zinc finger-like superfamily protein
Zm00001d051119	-1.7924	2.40E-03	Zinc finger (C3HC4-type RING finger) family protein
Zm00001d052118	-1.7065	3.31E-06	zinc finger protein 8
znf5	4.6928	1.29E-11	C2H2-type zinc finger family protein

Table 3.6: GO-term Enrichment Analysis Results

GO-TERM	ONTOLOGY*	DESCRIPTION	NO. DEG	NO. BACKGROUND	FDR *P*-VALUE
GO:0005576	C	extracellular region	215	2695	2.80E-14
GO:0005737	C	cytoplasm	990	19628	1.90E-07
GO:0009507	C	CHLOROPLAST	528	5439	7.90E-65
GO:0009536	C	plastid	543	5796	3.30E-62
GO:0000302	P	response to reactive oxygen species	74	765	2.10E-07
GO:0006950	P	response to stress	619	12007	0.00026
GO:0006952	P	defense response	206	3346	0.00011
GO:0006970	P	response to osmotic stress	132	1843	9.40E-06
GO:0009266	P	response to temperature stimulus	143	1840	3.00E-08
GO:0009314	P	response to radiation	240	2302	2.20E-29
GO:0009409	P	response to cold	105	1197	2.10E-08
GO:0009411	P	response to UV	55	457	2.70E-08
GO:0009415	P	response to water	79	1025	0.00022
GO:0009416	P	response to light stimulus	240	2186	2.30E-32
GO:0009605	P	response to external stimulus	275	4265	3.70E-08
GO:0009607	P	response to biotic stimulus	204	3201	1.50E-05
GO:0009611	P	response to wounding	67	883	0.0015
GO:0009617	P	response to bacterium	98	1243	7.80E-06
GO:0009627	P	systemic acquired resistance	56	676	0.00076
GO:0009628	P	response to abiotic stimulus	446	5674	4.70E-30
GO:0009644	P	response to high light intensity	64	375	1.20E-15
GO:0009651	P	response to salt stress	121	1708	4.80E-05
GO:0009697	P	salicylic acid biosynthetic process	23	254	0.05
GO:0009699	P	phenylpropanoid biosynthetic process	55	456	2.60E-08
GO:0009813	P	flavonoid biosynthetic process	57	438	1.00E-09
GO:0015979	P	photosynthesis	189	585	4.40E-85
GO:0016114	P	terpenoid biosynthetic process	77	446	3.70E-19
GO:0016117	P	carotenoid biosynthetic process	40	183	1.30E-12
GO:0050832	P	defense response to fungus	57	737	0.0033
GO:0080027	P	response to herbivore	6	10	0.0013

*C = Cellular Process, P = Biological Process

3.8 References

Agarwal, T., Grotewold, E., Doseff, A. I., & Gray, J. (2016). MYB31/MYB42 syntelogs exhibit divergent regulation of phenylpropanoid genes in maize, sorghum and rice. *Scientific Reports*, *6*, 28502. https://doi.org/10.1038/srep28502

Ahern, K. R., Deewatthanawong, P., Schares, J., Muszynski, M., Weeks, R., Vollbrecht, E., Duvick, J., Brendel, V. P., & Brutnell, T. P. (2009). Regional mutagenesis using Dissociation in maize. *Methods*, *49*(3), 248–254. https://doi.org/10.1016/j.ymeth.2009.04.009

Andorf, C. M., Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Campbell, D. A., & Sen, T. Z. (2010). The Locus Lookup tool at MaizeGDB: Identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics*, *26*(3), 434–436. https://doi.org/10.1093/bioinformatics/btp556

Ang, L.-H., Chattopadhyay, S., Wei, N., Oyama, T., Okada, K., Batschauer, A., & Deng, X.-W. (1998). Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of Arabidopsis development. *Molecular Cell*, *1*(2), 213–222. https://doi.org/10.1016/S1097-2765(00)80022-2

Aranda, P. S., LaJoie, D. M., & Jorcyk, C. L. (2012). Bleach gel: A simple agarose gel for analyzing RNA quality. *Electrophoresis*, *33*(2), 366–369. https://doi.org/10.1002/elps.201100335

Beardslee, T. A., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S., Stern, D. B., & Allison, L. A. (2002). A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. *The Plant Journal*, *31*(2), 199–209. https://doi.org/10.1046/j.1365-313X.2002.01344.x

Beckett, J. B., Neuffer, M. G., & Coe Jr., E. H. (1973). A chlorophyll mutant associated with *a3* located on chromosome 3. *Maize Genetics Cooperation Newsletter*, *47*, 147–148.

Boerjan, W., Ralph, J., & Baucher, M. (2003). Lignin biosynthesis. *Annual Review of Plant Biology*, *54*(1), 519–546. https://doi.org/10.1146/annurev.arplant.54.031902.134938

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab007

Burr, F. A., Burr, B., Scheffler, B. E., Blewitt, M., Wienand, U., & Matz, E. C. (1996). The maize repressor-like gene Intensifier1 shares homology with the R1/B1 multigene family of transcription factors and exhibits missplicing. *The Plant Cell*, *8*(8), 1249–1259.

Carey, C. C. (2002). *Three Regulators of the Maize Anthocyanin Pathway* [Dissertation]. University of Oregon.

Carey, C. C., Strahle, J. T., Selinger, D. A., & Chandler, V. L. (2004). Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 Gene in Arabidopsis thaliana. *The Plant Cell*, *16*(2), 450–464. https://doi.org/10.1105/tpc.018796

Casas, M. I., Falcone-Ferreyra, M. L., Jiang, N., Mejía-Guerra, M. K., Rodríguez, E., Wilson, T., Engelmeier, J., Casati, P., & Grotewold, E. (2016). Identification and characterization of maize salmon silks genes involved in insecticidal maysin biosynthesis. *The Plant Cell*, *28*(6), 1297–1309. https://doi.org/10.1105/tpc.16.00003

Chatham, L. A., & Juvik, J. A. (2021). Linking anthocyanin diversity, hue, and genetics in purple corn. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa062

Chatham, L., Paulsmeyer, M., & Juvik, J. (2019). Prospects for economical natural colorants: Insights from maize. *Theoretical and Applied Genetics*, *132*(11), 2927–2946. https://doi.org/10.1007/s00122-019-03414-0

Chayut, N., Yuan, H., Ohali, S., Meir, A., Yeselson, Y., Portnoy, V., Zheng, Y., Fei, Z., Lewinsohn, E., Katzir, N., Schaffer, A. A., Gepstein, S., Burger, J., Li, L., & Tadmor, Y. (2015). A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit. *BMC Plant Biology*, *15*(1), 274. https://doi.org/10.1186/s12870-015-0661-8

Chen, C., Zhang, K., Khurshid, M., Li, J., He, M., Georgiev, M. I., Zhang, X., & Zhou, M. (2019). MYB transcription repressors regulate plant secondary metabolism. *Critical Reviews in Plant Sciences*, *38*(3), 159–170. https://doi.org/10.1080/07352689.2019.1632542

Chen, F., & Dixon, R. A. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology*, *25*(7), 759–761. https://doi.org/10.1038/nbt1316

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ
preprocessor. *Bioinformatics*, *34*(17), i884–i890.
https://doi.org/10.1093/bioinformatics/bty560

Christensen, D. G., Orr, J. S., Rao, C. V., & Wolfe, A. J. (2017). Increasing growth yield and
decreasing acetylation in Escherichia coli by optimizing the carbon-to-magnesium ratio
in peptide-based media. *Applied and Environmental Microbiology*, *83*(6).
https://doi.org/10.1128/AEM.03034-16

Cone, K. C. (2007). Anthocyanin synthesis in maize aleurone tissue. In O.-A. Olsen (Ed.),
Endosperm (Vol. 8, pp. 121–139). Springer Berlin Heidelberg.
http://link.springer.com/10.1007/7089_2007_117

Conrad, L. J., & Brutnell, T. P. (2005). Ac-Immobilized, a stable source of Activator transposase
that mediates sporophytic and gametophytic excision of dissociation elements in maize.
Genetics, *171*(4), 1999–2012. https://doi.org/10.1534/genetics.105.046623

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A.,
Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools
and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Danecek, P., Schiffels, S., & Durbin, R. (2014). *Multiallelic calling model in bcftools (-m)*.
http://samtools.github.io/bcftools/call-m.pdf

Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Current Protocols in
Bioinformatics*, *51*(1), 11.14.1-11.14.19. https://doi.org/10.1002/0471250953.bi1114s51

Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13–15.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB
transcription factors in Arabidopsis. *Trends in Plant Science*, *15*(10), 573–581.
https://doi.org/10.1016/j.tplants.2010.06.005

Farre, D. (2003). Identification of patterns in biological sequences at the ALGGEN server:
PROMO and MALGEN. *Nucleic Acids Research*, *31*(13), 3651–3653.
https://doi.org/10.1093/nar/gkg605

Gao, Y., Jiang, W., Dai, Y., Xiao, N., Zhang, C., Li, H., Lu, Y., Wu, M., Tao, X., Deng, D., &
Chen, J. (2015). A maize phytochrome-interacting factor 3 improves drought and salt
stress tolerance in rice. *Plant Molecular Biology*, *87*(4–5), 413–428.
https://doi.org/10.1007/s11103-015-0288-z

Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., Moore, J., Montes, D., Mosley, T., Resendez, J., Nguyen, H., Wilson, L., Campbell, A., Sudarshan, D., & Lloyd, A. (2016). TTG2 controls the developmental regulation of seed coat tannins in Arabidopsis by regulating vacuolar transport steps in the proanthocyanidin pathway. *Developmental Biology*, *419*(1), 54–63. https://doi.org/10.1016/j.ydbio.2016.03.031

Goodman, C. D., Casati, P., & Walbot, V. (2004). A multidrug resistance–associated protein involved in anthocyanin transport in Zea mays. *The Plant Cell*, *16*(7), 1812–1826. https://doi.org/10.1105/tpc.022574

Guilfoyle, T. J., & Hagen, G. (2007). Auxin response factors. *Current Opinion in Plant Biology*, *10*(5), 453–460. https://doi.org/10.1016/j.pbi.2007.08.014

Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (3rd ed). Sinauer Associates.

Hatier, J.-H. B., & Gould, K. S. (2008). Anthocyanin function in vegetative organs. In C. Winefield, K. Davies, & K. Gould (Eds.), *Anthocyanins* (pp. 1–19). Springer New York. http://link.springer.com/10.1007/978-0-387-77335-3_1

He, J., & Giusti, M. M. (2010). Anthocyanins: Natural colorants with health-promoting properties. *Annual Review of Food Science and Technology*, *1*(1), 163–187. https://doi.org/10.1146/annurev.food.080708.100754

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., & Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany*, *62*(8), 2465–2483. https://doi.org/10.1093/jxb/erq442

Huai, J., Jing, Y., & Lin, R. (2020). Functional analysis of ZmCOP1 and ZmHY5 reveals conserved light signaling mechanism in maize and Arabidopsis. *Physiologia Plantarum*, *169*(3), 369–379. https://doi.org/10.1111/ppl.13099

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Kong, Q., Pattanaik, S., Feller, A., Werkman, J. R., Chai, C., Wang, Y., Grotewold, E., & Yuan, L. (2012). Regulatory switch enforced by basic helix-loop-helix and ACT-domain

mediated dimerizations of the maize transcription factor R. *Proceedings of the National Academy of Sciences*, *109*(30), E2091–E2097. https://doi.org/10.1073/pnas.1205513109

Lago, C. (2014). *Study of Maize Genotypes Rich in Anthocyanins for Human and Animal Nutrition* [Dissertation, Università Degli Studi Di Milano]. http://air.unimi.it/handle/2434/230010

Lao, F., Sigurdson, G. T., & Giusti, M. M. (2017). Health benefits of purple corn (Zea mays L.) phenolic compounds. *Comprehensive Reviews in Food Science and Food Safety*, *16*(2), 234–246. https://doi.org/10.1111/1541-4337.12249

Lauter, N., Gustus, C., Westerbergh, A., & Doebly, J. (2004). the inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics*, *167*(4), 1949–1959. https://doi.org/10.1534/genetics.104.026997

Lee, J., Durst, R. W., & Wrolstad, R. E. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *Journal of AOAC International*, *88*(5), 1269–1278.

Lev-Yadun, S., & Gould, K. S. (2008). Role of anthocyanins in plant defence. In C. Winefield, K. Davies, & K. Gould (Eds.), *Anthocyanins* (pp. 22–28). Springer New York. http://link.springer.com/10.1007/978-0-387-77335-3_2

Li, C.-Y., Kim, H.-W., Won, S., Min, H.-K., Park, K.-J., Park, J.-Y., Ahn, M.-S., & Rhee, H.-I. (2008). Corn husk as a potential source of anthocyanins. *Journal of Agricultural and Food Chemistry*, *56*(23), 11413–11416. https://doi.org/10.1021/jf802201c

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J., Cao, M., Pan, G., Rong, T., & Zhang, S. (2014). Validation of potential reference genes for qpcr in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0095445

Lindstrom, L. W. (1934). News items from ames, iowa. *Maize Genetics Cooperation Newsletter*, *8*, 10.

Liu, S., Yeh, C.-T., Tang, H. M., Nettleton, D., & Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS ONE*, *7*(5), e36406. https://doi.org/10.1371/journal.pone.0036406

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Ludwig, S. R., & Wessler, S. R. (1990). Maize R gene family: Tissue-specific helix-loop-helix proteins. *Cell*, *62*(5), 849–851. https://doi.org/10.1016/0092-8674(90)90259-H

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., & Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLOS Genetics*, *11*(1), e1004915. https://doi.org/10.1371/journal.pgen.1004915

Markham, K. R. (1988). Distribution of flavonoids in the lower plants and its evolutionary significance. In J. B. Harborne (Ed.), *The Flavonoids: Advances in Research since 1980* (pp. 427–468). Springer US. https://doi.org/10.1007/978-1-4899-2913-6_12

Marrs, K. A., Alfenito, M. R., Lloyd, A. M., & Walbot, V. (1995). A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature*, *375*(6530), 397–400. https://doi.org/10.1038/375397a0

Morohashi, K., Casas, M. I., Falcone Ferreyra, M. L., Mejia-Guerra, M. K., Pourcel, L., Yilmaz, A., Feller, A., Carvalho, B., Emiliani, J., Rodriguez, E., Pellegrinet, S., McMullen, M., Casati, P., & Grotewold, E. (2012). A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *The Plant Cell*, *24*(7), 2745–2764. https://doi.org/10.1105/tpc.112.098004

Naing, A. H., & Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiologia Plantarum*. https://doi.org/10.1111/ppl.13373

Qin, L., Sun, L., Wei, L., Yuan, J., Kong, F., Zhang, Y., Miao, X., Xia, G., & Liu, S. (2021). Maize SRO1e represses anthocyanin synthesis through regulating the MBW complex in response to abiotic stress. *The Plant Journal*, *105*(4), 1010–1025. https://doi.org/10.1111/tpj.15083

97

Robinett, D., Coe, E. H., & Cone, K. C. (1995). Map location of anthocyanin3. *Maize Genetics Cooperation Newsletter*, *69*, 46.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ron, D., & Dressler, H. (1992). pGSTag—A versatile bacterial expression plasmid for enzymatic labeling of recombinant proteins. *BioTechniques*, *13*(6), 866–869.

Selinger, D. A., & Chandler, V. L. (1999). A mutation in the pale aleurone color1 gene identifies a novel regulator of the maize anthocyanin pathway. *The Plant Cell*, *11*(1), 5–14. https://doi.org/10.1105/tpc.11.1.5

Shi, Q., Kong, F., Zhang, H., Jiang, Y., Heng, S., Liang, R., Ma, L., Liu, J., Lu, X., Li, P., & Li, G. (2019). Molecular mechanisms governing shade responses in maize. *Biochemical and Biophysical Research Communications*, *516*(1), 112–119. https://doi.org/10.1016/j.bbrc.2019.05.142

Singleton, V. L., & Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *American Journal of Enology and Viticulture*, *16*(3), 144–158.

Song, N., Xu, Z., Wang, J., Qin, Q., Jiang, H., Si, W., & Li, X. (2018). Genome-wide analysis of maize CONSTANS-LIKE gene family and expression profiling under light/dark and abscisic acid treatment. *Gene*, *673*, 1–11. https://doi.org/10.1016/j.gene.2018.06.032

Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., Barbazuk, W. B., Bass, H. W., Baruch, K., Ben-Zvi, G., Buckler, E. S., Bukowski, R., Campbell, M. S., Cannon, E. K. S., Chomet, P., Dawe, R. K., Davenport, R., Dooner, H. K., Du, L. H., … Brutnell, T. P. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*, *50*(9), 1282–1288. https://doi.org/10.1038/s41588-018-0158-0

Styles, E. D., & Coe, E. H., Jr. (1986). Unstable expression of an R allele with a3 in maize: A recessive intensifier of plant color. *Journal of Heredity*, *77*(6), 389–393. https://doi.org/10.1093/oxfordjournals.jhered.a110267

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, *45*(W1), W122–W129. https://doi.org/10.1093/nar/gkx382

Tominaga-Wada, R., Iwata, M., Nukumizu, Y., Sano, R., & Wada, T. (2012). A full-length R-like basic-helix-loop-helix transcription factor is required for anthocyanin upregulation whereas the N-terminal region regulates epidermal hair formation. *Plant Science*, *183*, 115–122. https://doi.org/10.1016/j.plantsci.2011.11.010

Vélez-Bermúdez, I.-C., Salazar-Henao, J. E., Fornalé, S., López-Vidriero, I., Franco-Zorrilla, J.-M., Grotewold, E., Gray, J., Solano, R., Schmidt, W., Pagés, M., Riera, M., & Caparros-Ruiz, D. (2015). A MYB/ZML complex regulates wound-induced lignin genes in maize. *The Plant Cell*, *27*(11), 3245–3259. https://doi.org/10.1105/tpc.15.00545

Wang, H., Wu, G., Zhao, B., Wang, B., Lang, Z., Zhang, C., & Wang, H. (2016). Regulatory modules controlling early shade avoidance response in maize seedlings. *BMC Genomics*, *17*, 269. https://doi.org/10.1186/s12864-016-2593-6

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M. K., Westhoff, P., Slamet-Loedin, I. H., Quick, W. P., Hibberd, J. M., & Langdale, J. A. (2013). Evolution of GOLDEN2-LIKE gene function in C3 and C4 plants. *Planta*, *237*(2), 481–495. https://doi.org/10.1007/s00425-012-1754-3

Xu, T., Purcell, M., Zucchi, P., Helentjaris, T., & Bogorad, L. (2001). TRM1, a YY1-like suppressor of rbcS-m3 expression in maize mesophyll cells. *Proceedings of the National Academy of Sciences*, *98*(5), 2295–2300. https://doi.org/10.1073/pnas.041610098

Yuan, W., Jiang, T., Du, K., Chen, H., Cao, Y., Xie, J., Li, M., Carr, J. P., Wu, B., Fan, Z., & Zhou, T. (2019). Maize phenylalanine ammonia-lyases contribute to resistance to Sugarcane mosaic virus infection, most likely through positive regulation of salicylic acid accumulation. *Molecular Plant Pathology*, *20*(10), 1365–1378. https://doi.org/10.1111/mpp.12817

Zhang, H., Gong, J., Chen, K., Yao, W., Zhang, B., Wang, J., Tian, S., Liu, H., Wang, Y., Liu, Y., & Du, L. (2020). A novel R3 MYB transcriptional repressor, MaMYBx, finely regulates anthocyanin biosynthesis in grape hyacinth. *Plant Science*, *298*, 110588. https://doi.org/10.1016/j.plantsci.2020.110588

Zheng, M., Chen, J., Shi, Y., Li, Y., Yin, Y., Yang, D., Luo, Y., Pang, D., Xu, X., Li, W., Ni, J., Wang, Y., Wang, Z., & Li, Y. (2017). Manipulation of lignin metabolism by plant densities and its relationship with lodging resistance in wheat. *Scientific Reports*, *7*(1), 41805. https://doi.org/10.1038/srep41805

Zhou, H., Lin-Wang, K., Wang, H., Gu, C., Dare, A. P., Espley, R. V., He, H., Allan, A. C., & Han, Y. (2015). Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *The Plant Journal*, *82*(1), 105–121. https://doi.org/10.1111/tpj.12792

Zhu, H.-F., Fitzsimmons, K., Khandelwal, A., & Kranz, R. G. (2009). CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Molecular Plant*, *2*(4), 790–802. https://doi.org/10.1093/mp/ssp030

CHAPTER 4: INCREASED ALEURONE AND PERICARP YIELD FOR ENHANCED ANTHOCYANIN AND PHYTONUTRIENT CONTENT

4.1 Abstract

Anthocyanins are desired by the natural colorant industry and are valuable for their health-promoting benefits. In maize, anthocyanins are mainly produced in the aleurone and pericarp layers of the grain. In addition to producing anthocyanins, these layers are nutritionally beneficial fractions of the kernel. Increasing pericarp and aleurone yield, therefore, has implications for biofortification of maize. Since quantification of these two layers is difficult, an additional goal of this study was to develop molecular markers for pericarp and aleurone yield. Two populations with various characteristics were genotyped using genotyping-by-sequencing. The first was a yellow corn population with contrasting pericarp thicknesses. The second was a blue corn population segregating for *Intensifier1* alleles. Both populations segregated for the multiple aleurone layer (MAL) trait that is known to increase aleurone yield. In this study, it was found that MALs are mostly determined by a locus on chromosome 8, but several minor loci are also involved. The inheritance of MALs was complex and seemingly more additive than dominant. In the blue corn population, anthocyanin content increased 30 to 40% with the addition of MALs demonstrating its effectiveness at increasing aleurone yield. Elemental analysis was performed on MAL lines and indicated a possible role of MALs in increasing iron content in the grain. QTL analyses are presented in this study on many pericarp, aleurone, and grain quality traits. Molecular markers were also tested for the MAL locus on chromosome 8 and candidate genes are discussed. Results of this study may assist breeders enhancing anthocyanin content and other beneficial phytonutrients in maize.

4.2 Introduction

The increasing consumer preference for natural ingredients is driving a need for more economical sources of natural colorants. Anthocyanins found in plants are a widely used source of natural colorants for several reasons. They are water soluble and can be used for many food and beverage applications. The color is similar to FD&C Red No. 40, the most common synthetic dye on the market (Center for Food Safety and Applied Nutrition, 2020). Additionally, anthocyanins are well-known health-promoting compounds. Aside from natural colorants, anthocyanins are useful as nutraceuticals. Studies have shown that anthocyanins reduce biomarkers for inflammation, cancer, and cardiovascular disease (He & Giusti, 2010). Finally,

101

anthocyanins are produced by nearly every plant species. Maize is an attractive source for natural colorants due to the abundant production of anthocyanins in the grain, and due to the economy of scale of maize production. In the grain, anthocyanins can be produced in two major tissues: pericarp and aleurone. Pericarp is the outermost layer of the kernel consisting of two to twenty cell layers and is maternally derived tissue (Tracy et al., 1978). Aleurone is an epidermal layer of the endosperm and contains many of the components necessary to break the kernel from dormancy and to sustain the development of the seedling (Becraft & Yi, 2011). Aleurone is genetically triploid, containing one paternal and two maternal genomes as a result of double fertilization.

Aside from producing anthocyanins, pericarp and aleurone are nutritionally beneficial fractions of the maize kernel. Pericarp is a good source of dietary fiber and may be a valuable supplement to human diets (Singh et al., 2012). Aleurone is sometimes viewed as a storage organ that accumulates many important phytonutrients. Quality protein and mineral reserves are stored in aleurone (Wolf et al., 1972). In barley aleurone, for example, a majority of the calcium, magnesium, phosphate, and potassium reserves in the endosperm are contained within the aleurone (Becraft, 2007). In many resource-scarce societies, staple crops which can be low in certain micronutrients and vitamins tend to be the main source of daily caloric intake. In Sub-Saharan Africa, maize accounts for as high as 25% of daily caloric intake (FAO, 2018). Biofortification by breeding for increased nutrition is one of the most effective strategies of reducing incidences of nutrient deficiencies (Bouis et al., 2011). Therefore, investigations into tools for breeders to enhance anthocyanin and nutrient accumulation in these nutritively rich tissues is warranted.

Increasing anthocyanin production in the grain for a more economical source of natural colorants can be accomplished in several ways. One approach is by combining genetic factors known to enhance pigmentation in maize. Factors such as *Colored1* (*R1*) and *Colorless aleurone1* (*C1*) are transcription factors primarily involved with activation of anthocyanin biosynthesis in the aleurone of blue corn. The gene, *Intensifier1* (*In1*), is a dominant inhibitor of anthocyanin production in the aleurone and when homozygous for the recessive allele at this locus results in a greater than two-fold increase in anthocyanin content (Paulsmeyer et al., 2018). Purple corn contains anthocyanins in the pericarp layer and is generally the most productive fraction of the maize kernel in terms of anthocyanin yield (Paulsmeyer et al., 2017). Purple

102

pericarp is activated by *Leaf color1*, *Hopi1*, or *Booster1* loci in combination with *Plant color1* (Ludwig & Wessler, 1990; Petroni et al., 2000).

A second conceptually simple approach to increasing anthocyanin content in the grain is to increase the yield of anthocyanin-producing tissue per unit of grain. The process by which this can be accomplished is not so simple. This approach requires the tedious separation of kernel tissue layers. Pericarp is easily separated by boiling kernels or with the dry-milling process. Pericarp yield can be represented in two ways: proportion of pericarp of total weight, or weight of pericarp per kernel. Previous studies have investigated pericarp thickness mostly as a proxy for eating quality traits in sweet corn, waxy corn, and popcorn and not from a yield perspective (Choe & Rocheford, 2012; Mohamed et al., 1993; Wu et al., 2020).

Aleurone is more difficult to fractionate than pericarp since it is bound to the endosperm. Typically, aleurone is a single layer of cells around the periphery of the endosperm. However, some tropical landraces are known to produce two to nine cellular layers in what is referred to as the multiple aleurone layer (MAL) trait (Welch et al., 1993). Previous studies and this study establish that the MAL trait is dominant or partially dominant (de Miranda, 1980; Duangploy et al., 1976; Nelson & Chang, 1974; Wolf et al., 1972). The number of genes involved with the trait is currently unknown but was hypothesized to include at least three (de Miranda, 1980). The effect of this trait is that it can increase the proportion of aleurone per kernel from around 2% on average to up to 4% in the varieties tested in a previous study (Wolf et al., 1972). MALs would be beneficial in increasing anthocyanins in blue corn and also the beneficial phytonutrients stored in this layer. The tedious nature of quantifying pericarp yield and viewing aleurone layers necessitates the need for molecular markers associated with increasing pericarp and aleurone yield in maize to enhance anthocyanin content and other beneficial phytonutrients contained within those tissues.

4.3 Materials and Methods

4.3.1 Plant materials

Multiple aleurone layer landraces San Martin 105 and San Martin 119 were provided by the North Central Regional Plant Introduction Station in Ames, IA, USA. Genetic stocks 707G, 805B, 805C, 805CA, and 5708E were provided by the Maize Genetics Cooperation Stock Center in Urbana, IL, USA. Two populations were developed for the purpose of determining the inheritance of MALs and pericarp yield (Table 4.1). Most MAL landraces assayed were

103

segregating for the ability to produce MALs, and therefore heterozygosity may be an issue. The first population, designated MAL1, was a yellow corn population from a cross between San Martin 105 (PI 515087) and Mo17. San Martin 105 has the ability to produce five to six aleurone layers and consistently produces MALs in every kernel. These two parents contrasted in the thickness of their pericarps as well. Most MAL landraces have thin pericarps (personal observation), while temperate yellow dent lines tend to have thicker pericarps (Ito & Brewbaker, 1991). The second population was a cross between MAL landrace San Martin 119 (PI 571553) and genetic stock 707G from the Maize Genetics Cooperation Stock Center and is designated MAL2. This MAL landrace had the capability to produce three aleurone layers consistently. The genetic stock 707G was chosen because it contains blue aleurone pigmentation along with a recessive *intensifier1* (*in1*) gene, which is known to enhance aleurone pigmentation. Both populations were backcrossed once with their respective non-MAL parent since there is deleterious genetic load associated with MAL landraces that make them difficult to grow in the Midwest USA. MAL2 segregated for the blue aleurone trait, so only blue kernel samples were chosen after each generation of selfing. MAL1 was phenotyped and genotyped for MALs and pericarp content in the BC_1 stage and so the aleurone in each kernel is genetically BC_1F_1 (Table 4.1). MAL2 was selfed to the $BC_1F_{2:3}$ stage and the families were genotyped and phenotyped (Table 4.1). To bulk more seed for subsequent years of analysis, plants within families were sib-pollinated to maintain the family structure.

4.3.2 Phenotypic measurements

The number of aleurone layers were assayed by rough sectioning of mature kernels with a razor blade. Kernels were hydrated for at least 16 hours in water at room temperature prior to analysis. Hydrated kernels were cut into 1 to 2 mm sections longitudinally in the center of the kernel for most analyses or sagitally along the crown if the purpose was to germinate the kernel later. The resulting sections were viewed within 3 to 5 min to prevent drying on an Olympus BX-66 Fluorescent microscope (Olympus Corporation) equipped with a triple band DAPI-FITC-TRITC fluorescence filter and 10X magnification. Sections could also be stained using 0.1% (*w/v*) toluidine blue for 2 min prior to viewing with white light illumination from above and 10X magnification. The number of aleurone layers per kernel was determined as the consensus number of layers as viewed around the entire circumference of the endosperm. These values were averaged among at least five kernels for MAL1 and at least ten kernels for MAL2. The

104

maximum number of aleurone layers among the kernels assayed was also recorded. Commonly, the number of aleurone layers was heterogeneous. The maximum number of aleurone layers possible within a kernel was recorded and then averaged for all kernels assayed to get an inflated average. Grain quality traits, including moisture, protein, oil, and density were measured for the MAL1 population using the Perten DA 7200 Near Infrared (NIR) Analyzer. Two replicates of 10 kernels were weighed and adjusted to 0% moisture using the NIR values. These 10-kernel lots were boiled for 15 min in 20 mM NaOH to soften the kernels and loosen pericarps. Pericarp and kernel fractions were separated using forceps into separate aluminum weigh boats. Fractions were heated at 103 ˚C for 72 hours according to AACC approved method 44-40.01 (AACC Approved Methods of Analysis, 2009) to remove moisture and weighed immediately. The dry pericarp weight was divided by the moisture-adjusted ten-kernel weight to calculate proportion of pericarp per kernel. Average variability of measurements within ear were within 8% CV (standard deviation divided by mean) for proportion of pericarp, so the values are acceptably reproducible. Variability in measurements among ears tended to be higher than within ear measurements (data not shown). Micronutrient analysis was performed using ICP-OES (Inductively Coupled Plasma-Optical Emission Spectrometry) at Brookside Laboratories Inc. in New Bremen, OH, USA. A subset of samples (n=19) from MAL1 were selfed twice and analyzed for their aleurone content. Two 35-gram replicates of single and MAL samples along with two 35-gram replicates and two ears of Mo17 were run for elemental analysis.

4.3.3 GBS library construction

Genomic DNA for this study was extracted using a CTAB procedure modified for 96-well plates (Doyle & Doyle, 1990). A double restriction enzyme genotyping-by sequencing approach was used to generate genetic markers associated with MAL formation (Poland et al., 2012). The total number of taxa in MAL1 and MAL2 was 165 and 150, respectively. Additional samples from external experiments were included within the libraries to bring the total to 192 samples in each library. Libraries generated were sequenced at the DNA Services laboratory of the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign. Before sequencing, libraries were cleaned one additional time using a 1:1 ratio with AxyPrep Mag PCR Cleanup beads (Axygen, Inc.) to ensure removal of primer and adaptor dimers, then quantitated on Qubit (Life Technologies) and evaluated on AATI Fragment Analyzer (Advanced Analytics). The pools were diluted to 5 nM concentration and further quantified by qPCR on a

105

BioRad CFX Connect Real-Time System (Bio-Rad Laboratories, Inc.). The two final pools were loaded onto separate lanes and then sequenced on an Illumina HiSeq4000 with version 1 SBS sequencing reagents from one end of the molecules for a total read length of 100 nt. Raw data was adapter-trimmed and converted to fastq format using bcl2fastq v2.20 Conversion Software (Illumina). Four samples were removed from MAL2 after it was found that they were a result of pollination contamination in the BC_1F_1 stage. This brings the total number of samples to 165 for MAL1 and 146 for MAL2.

4.3.4 SNP discovery pipeline

Raw reads were de-multiplexed and the barcodes were removed using Stacks 2.54 (Catchen et al., 2013). Reads with a quality score of less than 30 were removed. Taxa were aligned using BWA-MEM using the Mo17 genome for MAL1 and the B73 RefGen_v5 genome for MAL2 (Li & Durbin, 2010; Sun et al., 2018). SNP discovery used the SAMtools 1.11 (https://github.com/samtools/samtools/releases/1.11) and BCFtools 1.9 (https://github.com/samtools/bcftools/releases/1.9) pipelines (Bonfield et al., 2021; Danecek et al., 2021; Hufford et al., 2021). Genotypes were called using the multiallelic calling model (Danecek et al., 2014). SNPs with a read depth greater than 6 and a minimum quality score of 30 were kept for analysis. Taxa with less than 100,000 reads were removed after this step, which removed one sample from MAL1 and two samples from MAL2. The total number of samples for the final dataset was 164 for MAL1 and 144 for MAL2. This dataset was then fed into TASSEL 5.0 where the dataset was filtered to remove missing information and any sites that had a minor allele frequency and heterozygosity frequency less than 0.05 or greater than 0.9 (Bradbury et al., 2007). The cleaned dataset was thinned so that markers were spaced at least 10,000 base pairs apart since closely linked markers do not add additional information. This resulted in 3871 SNPs for MAL1 and 5403 SNPs for MAL2. GBS data is filled with errors in heterozygote calling due to low-depth sequencing. In addition, the maize genome is repetitive sequences and duplications that make alignment difficult (Jiao et al., 2017). A custom script was developed in R to smooth the dataset and correct the errors similar to the bin-mapping method from a previous study analyzing GBS data (Chen et al., 2014; R Core Team, 2019). MAL1 was smoothed with a window of 20 SNPs with a step size of 2 SNPs. MAL2 was smoothed with a window of 40 SNPs with a step size of 2 SNPs. If a window was 75% homozygous for either the reference or alternate allele, then the middle position of that window was reported as homozygous. If not,

106

then it was considered heterozygous. This resulted in 1835 high quality SNPs in MAL1 and 2498 high quality SNPs in MAL2.

4.3.5 SSR marker analysis

Several PCR-based markers were used in this study. Many of these markers were publicly available from the Maize Genetics and Genomics Database (https://maizegdb.org/data_center/ssr) (Andorf et al., 2010). Others were custom designed based on polymorphic sequence content between genomic sequences of B73 and Mo17 (Hufford et al., 2021; Sun et al., 2018). Eight markers chosen were spaced across the chromosome 8 QTL from 19.1 Mb to 27.1 Mb in Mo17 (Table 4.2). Most markers were validated in the MAL1 population and are known to be polymorphic. *MPMAL1*, *MPMAL2*, and *MPMAL3* custom markers fully aligned to each maize genome available on the Maize Genomics and Genetics Database (Cannon et al., 2011). The reaction consisted of 1.2 M betaine, 600 nm of each primer, and 1 to 100 ng DNA in Taq 5X Master Mix (New England Biolabs). The thermal cycler program was the same for each marker except for the annealing temperature listed in Table 4.2. The cycles used were 1 min initial denaturation at 94 ˚C followed by 35 cycles of 30 s at 94 ˚C, 30 s at the annealing temperature, and 30 s at 72 ˚C; and ending with 3 min at 72 ˚C. Additionally, custom RFLP marker *MPMAL4* was designed based on genotype reads. A restriction enzyme assay consisted of 20% (v/v) PCR product, 1X CutSmart Buffer (New England Biolabs), and 0.5 units *Eae*I run for 1 hour at 37 ˚C. The PCR products were separated on a 4% SFR agarose gel (VWR) while restriction enzyme products were separated on a 2% gel in ice-cold TBE at 5.4 V cm[-1]. Running in the gel for 1.5 to 2 hours was sufficient for all SSR markers, while 45 min was sufficient to separate the restriction enzyme marker.

4.3.6 Statistical analyses and QTL mapping

Elemental analysis data was averaged among the replicates and elemental composition between the single versus multiple aleurone layer lines was compared using the Tukey's Honest Significance Difference method in R (R Core Team, 2019). Phenotypes for MAL2 were a result of multi-year trial, so variance components and Best Linear Unbiased Predictions (BLUPs) could be calculated. Mixed models were run using the *lme4* package with defaults and the model in Equation 1 (Bates et al., 2015, p. 4). In this equation, term y is the phenotypic values, μ is the grand mean, α is the random effect of year, β is the random effect of year, γ is the random effect of genotype, αγ is the random year × genotype interaction term, and ε is the residual.

107

Equation 1: $y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \varepsilon_{(ijk)l}$

Estimated variance components were used in Equation 2 to calculate broad-sense heritability (H^2). In Equation 2, e is the number of years ($e = 2$) and r is the average number of replications per year ($r = 1.5$).

Equation 2: $H^2 = \dfrac{\sigma_g^2}{\sigma_g^2 + \dfrac{\sigma_{g \times e}^2}{e} + \dfrac{\sigma_e^2}{r*e}}$

A genetic map was estimated using the Carter-Falconer method in the "qtl" package in R (Broman et al., 2003; Carter & Falconer, 1951). Conditional genotype probabilities were calculated with fixed step width of 0.1 using "calc.genoprob". QTL mapping was performed using the "Expectation-Maximization" method in "scanone" and the 1.5-LOD interval was calculated using "lodinterval". Thresholds for each trait were calculated using 1000 permutations (Broman et al., 2003). The most significant markers that were significantly above the LOD threshold were run in a linear model in "fitqtl" to calculate "Type III" sums of squares, additive and dominance estimates, and the partial R^2 values (Broman et al., 2003). Estimated effects of each marker were calculated using a linear model in R (R Core Team, 2019).

4.4. Results

4.4.1 Developing an efficient protocol for viewing multiple aleurone layers

The aleurone layer is a relatively thin layer on the periphery of the endosperm. In most cases, the aleurone is the same color as the endosperm and indistinguishable by sight when using rough sectioning (Figure 4.1A and B). In blue corn, the aleurone is pigmented and the layer is visible. However, in blue corn with MALs, the dark pigment obscures the aleurone layers. Maize kernels can be sectioned using a cryotome or rotary microtome, but this requires special equipment, technical expertise, and processing samples is laborious and tedious. Maize kernels are also difficult to section due to their hardness and starch content. Even rough sectioning with a scalpel or razor blade is difficult without prior rehydration. Boiling kernels for 2 hours rehydrates the kernels enough for sectioning with at least a razor blade but inactivates the kernel and can destroy the integrity of the layers. Soaking kernels for at least 16 hours in water at room temperature is the preferred method for rehydration followed by rough sectioning with a razor blade to 1 to 2 mm thickness. Toluidine blue is a common stain used for plant tissues as it tends to stain polysaccharides in cell walls. Aleurone layers in yellow corn samples can be efficiently viewed with 0.1% (*w/v*) toluidine blue staining of rough sections for 2 min (Figure 4.1D, F, and

108

H). The only equipment necessary are razorblades and a standard microscope with 10X magnification and top lighting. This staining method contrasts the aleurone cell walls from the cell bodies allowing for efficient viewing of the aleurone layers. This method does not work with blue corn, however, since there is already pigmentation within the aleurone cells. Staining also adds an additional 2 min to the protocol which becomes a bottleneck when assessing thousands of samples, say, for a breeding program. The most efficient method for viewing aleurone layers is rough sectioning with a razor blade followed immediately (within 3 to 5 min) by viewing with a fluorescent microscope with a triple band DAPI-FITC-TRITC fluorescence filter. The DAPI filter allows for viewing of yellow corn and weakly pigmented blue aleurone samples (Figure 4.1D, E, G, and I), while the contrast with the red excitation of the TRITC filter contrast blue aleurone cell bodies from aleurone cell walls (Figure 4.1J). Sectioning longitudinally gives an accurate cross-section of the kernel and allows for multiple 1 to 2 mm sections to be made from a single kernel easily. Additionally, horizontal sectioning on the crown of the kernel can be done if the goal is to germinate the kernels after analysis. In this method, one is attempting to cut the crown of the kernel off without damaging the coleoptile. However, this sectioning is less accurate because the whole kernel is not being viewed and only the crown.

4.4.2 Multiple genes confer the MAL trait

The number of genes conferring MALs differed depending on the MAL landrace used for the two populations. MAL1 was phenotyped and genotyped at the BC_1 generation and so frequencies of MAL are expected to be 50% for a single gene or 75% for two genes assuming dominance and no epistasis. MAL2 was phenotyped and genotyped at the $BC_1F_{2:3}$ stage and so the frequency of MAL individuals is expected to be 28.1% for a one-gene system, 48.3% for a two-gene system, and 60.2% for a three-gene system, also assuming no epistasis and complete dominance. Given the ratio of MAL samples shown in Table 4.3, it can be concluded that at least two genes are segregating MAL1 (χ^2 p=0.87), with three possibly segregating in the MAL2 population (χ^2 p=1).

Families from MAL2 were replicated across two years to calculate broad-sense heritability (H^2). The MAL trait is moderately heritable in this population with a relatively large contribution of genotype × environment interactions (Table 4.3). MAL counts were determined based on the consensus number of aleurone layers. Often a kernel would have a heterogeneous distribution of aleurone layers across the crown to the tip. In these cases, the maximum layers

109

possible was also recorded. Adding a correction for this value resulted in a large increase in genotypic variance and an increase in H^2 from 0.68 to 0.86 (Table 4.3). The MAL trait has a strong genetic component but can be influenced by the environment or other unknown factors to produce more or less aleurone layers.

4.4.3 Reciprocal crosses indicate the MAL trait is mainly additive

To determine if there was a dosage effect with MAL alleles, reciprocal crosses were made between single layer and MAL lines. Aleurone is triploid, receiving one paternal and two maternal genomes, so it is expected that aleurone layers will differ if the inheritance is more additive. To reduce genetic variability, members from MAL1 were backcrossed two or three additional times to Mo17, the recurrent temperate yellow dent inbred (Table 4.1). In order to assay the aleurone layers in this population, the crown of the kernel was viewed under the microscope as described above so the kernels of interest could be germinated. Ears were designated females if they received pollen from the single aleurone parent or males if they donated pollen to the single aleurone parent. Several female ears with up to four aleurone layers and MALs in 75% of kernels were found indicating that at least two MAL genes were still segregating even after three backcrosses in this population. These females with four layers were reciprocal crossed to Mo17 once again to test the inheritance of MAL formation. An average of 47% of female kernels (n=240) and 16% of the male MAL kernels produced MALs (n=77). These testcrosses had significantly fewer MALs than expected for a two dominant gene system (χ^2 p<0.0001). The average number of layers was 1.62 for female ears (n=9) and 1.13 for male ears (n=7). This was statistically significantly different in a paired t-test (p=0.016). Two individuals from MAL1 and MAL2 that consistently produced MALs for at least two generations of selfing or sib-crossing were also used to reciprocal cross to Mo17 and 707G (Table 4.1). These kernels were assessed by slicing the kernel longitudinally. Surprisingly, these individuals began losing the MAL phenotype with only around 65% of kernels able to produce MALs that season (MAL1, n=50; MAL2, n=49). There was a significant difference between female versus male parents for MAL2 (pooled t-test; p=0.0004), but not for MAL1 (pooled t-test; p=0.32). The average layers for females and males were 1.81 and 1.12 in MAL2 and 1.65 and 1.49 for MAL1. Genetic stock 5708E from the Maize Genetics Cooperation Center was found to contain two to three aleurone layers consistently. Reciprocal crosses (n=5) to Mo17 and 707G also did not produce a significant difference in the dosage of MAL genes (pooled t-test; p=0.73). The

110

reciprocal cross experiment determines that MAL dosage is additive and that increasing the number of MAL alleles increases aleurone layer number.

4.4.4 Pericarp yield

Pericarp is the most productive anthocyanin-producing tissue in the maize grain (Paulsmeyer et al., 2017). Although most work has been done on pericarp thickness, this does not necessarily translate to pericarp yield on a per-kernel or per-weight basis. Two separate measurements were made to calculate pericarp yield: pericarp yield in terms of weight per kernel and in terms of proportion of total weight. MAL landrace San Martin 105 had an average of 14.18 mg pericarp per kernel, which was a proportion of 4.97% by weight. Mo17 had an average of 7.87 mg pericarp per kernel, or 5.56%. Proportion of pericarp by weight was not significantly different between the two parents with three replicates. The kernels were larger for San Martin 105 than Mo17 in the season they were sampled, which is the reason for the higher pericarp yield per kernel. A larger kernel volume (calculated as 10-kernel weight divided by density) and kernel weight was highly correlated with pericarp weight per kernel, but density was only loosely correlated with these traits (Table 4.4). In this study, pericarp weight per kernel and proportion of pericarp by weight were normally distributed despite only one round of recombination through backcrossing before measurements. The amount of variability was relatively high considering the parents differed only slightly in their proportion of pericarp by weight. There was a range of 4.62 to 18.78 mg pericarp per kernel and 2.94 to 7.22% pericarp by weight in MAL1 (Table 4.3). Pericarp volume and pericarp weight were moderately positively correlated in this population (Table 4.4), which suggests these traits could be controlled by different factors that can be improved simultaneously.

4.4.5 Investigating the effect of MALs on nutritional quality of maize

The distribution of anthocyanin content was complex. The data appears to be trimodal and heavily skewed to the right (Figure 4.2A). Many values were low indicating *in1* was segregating. The second distribution was of *in1* families and the third were those containing MALs with *in1*. Anthocyanin content and the log-transformation of anthocyanin content were both significantly positively correlated with the average number of MALs and maximum number of aleurone layers. As the maximum aleurone layer number increased, so did anthocyanin content (Figure 4.2B). Overall, there was a 30 to 40% increase in anthocyanin production just due to the presence of MALs in recessive and dominant *In1* lines, respectively (Table 4.3). This

111

difference was significant in a *t*-test (*In1* p=0.0418, *in1* p= 0.0035). The highest anthocyanin content in MAL2 was in a family with *in1* and a maximum of three layers. The predicted anthocyanin content over the two years studied was 457 mg anthocyanins per kg kernel weight. Anthocyanin content was also heavily controlled by genetic factors in this population with nearly all variation due to *in1* and MALs (Table 4.3).

Most other quality traits did not increase with the presence of MALs. There was a small but significant increase in oil content in MAL1 (Table 4.4). To assess the micronutrient content, elemental analysis with ICP-OES of single aleurone layer and MAL lines was conducted. The population chosen for study was the MAL1 yellow corn population. Individual selfed ears from the $BC_1F_{2:3}$ generation were assessed for their aleurone layer content (Table 4.1). Eight single aleurone layer and eleven MAL ears were represented. Data is shown in Table 4.5. Iron content was significantly increased in MAL lines, indicating a potential role of MALs in increasing micronutrient content.

4.4.6 QTL mapping in the MAL1 population

A major effect QTL on chromosome 8 was found for increased number of aleurone layers in this population (Figure 4.3). This locus encompasses a 1.5-LOD significance interval of 6.9 Mb centered around 21.3 Mb and accounts for 19.39% to 22.01% of the variation in MALs depending on the aleurone measurement (Table 4.6). It is predicted that there are at least two loci conferring MALs in this population based on the segregation pattern. Many MAL measurements only had a single significant locus, but average aleurone layers had an additional QTL on chromosome 1 (Figure 4.3). This SNP marker was located around 219.7 Mb and accounted for 5.98% of the total variation in average aleurone layers. As indicated above, it appears the MAL trait is additive. Having MAL alleles from both QTL increased the average aleurone layer number more than just having one of each allele (Figure 4.4A).

Pericarp traits volume and weight were controlled by different loci, which is logical since they were only moderately correlated (Table 4.4). Kernel volume, pericarp weight per kernel, and kernel weight all had a significant locus at the start of chromosome 6 (Table 4.6). Proportion of pericarp by weight and kernel weight had a QTL on chromosome 5 around 88.9 Mb that accounted for 8.42% to 14.25% of the variation in proportion of pericarp (Table 4.6). Additional gain quality traits assessed by NIR spectroscopy were also mapped and included in Table 4.6.

112

4.4.7 QTL mapping in the MAL2 population

The MAL2 blue corn population was theoretically segregating for at least three genes according to the proportion of taxa with MALs. The highest effect locus for MAL formation was due to the chromosome 8 QTL shared with MAL1 (Figure 4.5). In most cases, this was the only significant QTL (Table 4.7). When considering imprinting and calculating the average number of aleurone layers on the basis of the maximum number of layers possible in the kernel, three loci were significantly greater than the LOD threshold (Figure 4.5). These loci contributed 0.21% to 1.17% of the variation in adjusted aleurone layers, which was not significant in a linear model. The effect of these loci can be visualized in Figure 4.4B. The contribution of the minor loci to average aleurone counts was minimal. As mentioned in the reciprocal cross study, the MAL trait is most likely additive more than dominant. The ratio of additive to dominance deviations for MAL alleles was calculated between 6.83 to 30.83 supporting the alternative hypothesis that MALs is mostly additive (Table 4.7).

Anthocyanin content was measured across two years. The distribution of predicted anthocyanin content was skewed (Figure 4.2A), so the log-transformation of anthocyanin content was mapped. Four loci could be detected (Figure 4.7). The highest effect QTL contained the gene *in1*, which is known to have a large effect on anthocyanin content (Paulsmeyer et al., 2018). The second highest effect locus was within the same interval as the QTL for MAL formation. Anthocyanin content increased significantly as aleurone layers increased, so this finding demonstrates that anthocyanin content can be increased with the inclusion of this major effect locus. The other two loci were relatively minor but assisted in increasing anthocyanin content (Table 4.7). These did not overlap with any canonical anthocyanin biosynthetic or regulatory genes.

4.4.8 Fine-mapping the major MAL QTL

In an attempt to narrow down the significance interval for the chromosome 8 QTL for MAL formation, a fine-mapping approach was used. Several PCR-based markers were developed to increase the resolution of the QTL. Many of these markers were validated using the MAL1 population (Figure 4.6). Several individuals from MAL1 were selfed an additional time and the $BC_1F_{2:3}$ plants were grown. Of these ears, 190 individuals were chosen based on variation in aleurone layer number. Fifteen seedlings from each ear were pooled and genotyped using SSR markers *umc1974*, *umc1530*, and *umc1778*. These markers span the entire

113

chromosome 8 QTL region (Table 4.2). Results of the fine-mapping experiment show that the interval between *umc1974* to *umc1530* was the most significant. The LOD scores ranged from 9.3 to 22.5 and were all highly linked to the MAL phenotype. The most significant locus is 4.5 cM after *umc1974* or 4.6 cM before *umc1530*. The 1.5-LOD significance interval is 2.3 to 6.8 cM after *umc1974*. If assuming equal recombination and high correlation between the physical and genetic map, the chromosome 8 QTL region in MAL1 can be narrowed down to a 1.5-LOD interval of 20.4 Mb to 23.0 Mb in Mo17. Within this interval are 42 annotated genes (Sun et al., 2018).

4.5 Discussion

   Increasing the yield of pericarp and aleurone tissues in the maize grain has never been a direct goal for breeding programs until now. The initial goal of this study was to increase anthocyanin content in the grain by altering morphological features of the kernel. However, it is known that aleurone is a storage organ for dietarily important nutrients like calcium, magnesium, phosphate, potassium, and quality protein, while pericarp is a source of dietary fiber (Becraft, 2007; Singh et al., 2012; Wolf et al., 1972). The implications of increased pericarp and aleurone yield are more far-reaching than anthocyanin content. Biofortification is a tool breeders are utilizing now to reduce the incidence of "hidden hunger" or micronutrient deficiencies in staple crops (Bouis et al., 2011). The addition of MALs is a promising tool for breeders trying to elevate nutritional content of maize. Breeding for increased pericarp and aleurone yield is difficult because these tissues are difficult to quantify individually. The methods created here to quantify tissue yield have been used to assay thousands of samples efficiently. In the future, information from this study could be used to develop molecular markers for these traits so that breeding for these important morphological traits can be even more efficient.

   The inheritance of MALs in this study was complex. The first report of MALs was from a recurrent-selection study for high-amylose starch utilizing diverse lines (Wolf et al., 1969). The line they used was designated "Peru 442" and was later found to be derived from the Coroico landrace, whose region extends from Ecuador into Peru, and to Bolivia and parts of Brazil (Wolf et al., 1972). The San Martin landraces used here originated in Peru. The maximum aleurone layers reported for any MAL landrace is nine layers (Duangploy et al., 1976). This current study saw a maximum of six layers in the original San Martin 119 parent, but a maximum of four layers in the populations developed. The first published study for MAL inheritance indicated

114

through reciprocal crosses that doubled layers were dominantly inherited (Wolf et al., 1969). According to the reciprocal cross study here, having one parental copy was not always sufficient to confer MALs. Only 16% of male MAL1 kernels produced MALs when 75% was expected. Some minor effect loci were also apparently not sufficient enough to confer MALs in some samples. The inheritance of MALs appears to be more additive than dominant, with the average number of average layers increasing with more MAL alleles (Figure 4.4, Table 4.7). The inconsistency of aleurone layer production from a genotype perspective is also in part due to recombination between SNP markers and the causative genes. Rarely is a SNP marker within the causative trait, especially with low coverage sequencing methods like GBS. Furthermore, unknown epigenetic or environmental factors may be involved with modulating aleurone layer number. As MAL lines were selfed, the number of offspring with MALs appeared to decrease. This inconsistency will be investigated in future studies. Given these inconsistencies, it would be advantageous for breeders to have inbred parents from both heterotic pools contributing MALs in a hybrid so the maximum layers could be combined.

Anthocyanin content exceeded expectations in the MAL2 population (Figure 4.2). The highest anthocyanin content family had over two times the concentration of the *in1* parental stock 707G (Paulsmeyer et al., 2017). The aleurone layer is only approximately 2% of the kernel by weight (Wolf et al., 1972), but was still able to produce an appreciable amount of anthocyanins in MAL2. There was a 30 to 40% increase in anthocyanin content on average with additional aleurone layers (Table 4.3). This increase in anthocyanin content is similar to the increase in aleurone yield from a previous study investigating MAL effect on aleurone yield (Wolf et al., 1972). Aleurone yield measurements were not directly measured in this population, but the stark increase in anthocyanin content is indicative of an overall increase in aleurone tissue yield. Because of the implications of anthocyanins and aleurone tissue with human health, additional health benefits of the MAL trait were also assessed. There was no association with total protein content and MAL formation in MAL1 (Table 4.4). Additional experiments should be focused on analyzing individual amino acid contents for quality of protein in MAL lines. There was a small significant increase in oil content of about 3.6% in MAL lines. This increase in oil was interesting since it is known that aleurone accumulates oils for use by the developing seedling (Geisler-Lee & Gallie, 2005). Elemental analysis found a significant increase in iron content (Table 4.5). There was a large range in many of the micronutrients measured, indicating

115

that the inability to parse out significant differences may have been due to high genetic variability among samples. Environmental variability does not seem to have a major effect on zinc and iron content as the heritability of these micronutrients has been shown to be very high (Hindu et al., 2018). Future work will be aimed at utilizing near-isogenic lines with MALs to remove genetic variability not associated with MALs. The HarvestPlus (https://www.harvestplus.org/) goals for micronutrients in maize for zinc and iron are 38 and 60 µg g^{-1}, respectively (Bouis et al., 2011). The ranges found in this population did not meet those goals (Table 4.5). Mo17 appears to be low in micronutrient content, which brought the total micronutrient potential down. Plant breeders should introgress MALs into varieties already improved for micronutrient content to further increase content.

The mechanism for MAL formation in this study is currently unknown, but much work has been done to investigate aleurone layer development in maize. The MAL trait in the tropical landraces studied here is unique in that endosperm development is abnormal, but germination rates and plant vigor are seemingly unaffected (personal observation). Most of the known endosperm mutants producing MALs characterized to this date result in inviable kernels (Becraft & Yi, 2011). Exceptions are *Extra cell layer1* (*Xcl1*) and certain alleles of *Supernumerary aleurone layer1* (*Sal1*) that add an additional layer to the aleurone (Kessler et al., 2002; Shen et al., 2003).

Normal aleurone layer development depends on several factors. First, differentiation depends on positional cues. During endosperm development, the peripheral cells must detect their position in relation to the whole endosperm to differentiate into starchy endosperm or aleurone cells. Furthermore, hormones auxin and cytokinin appear to be the major signals for sensing positional cues. For example, *PIN-formed protein1* (*Pin1*) is a polar auxin efflux transporter that determines the direction of auxin transport in the developing maize kernel (Forestan et al., 2010). In an experiment determining *Pin1* function, auxin transport inhibitor *N*-1-naphthylphthalamic acid (NPA) was added to kernels developing on media. The result was the hyperaccumulation of indole-3-acetic acid in the aleurone and the differentiation of four aleurone layers in the treated kernels (Forestan et al., 2010). Therefore, auxin appears to be a positive regulator of aleurone fate. Additionally, cytokinin is involved with cell division, but may also be involved with cell fate. An experiment overexpressing cytokinins in the developing embryo found mosaic patches of aleurone and starchy endosperm on the crown, suggesting that

116

enhancing cytokinin levels inhibits aleurone differentiation (Geisler-Lee & Gallie, 2005). In some cases, only the crown accumulated MALs, while the sides of the kernel did not (personal observation). This indicates misregulation of cytokinin levels may interact with MAL formation in the populations studied here.

In addition to positional cues, an additional factor involved with proper aleurone layer development is the plane of cell division. In the later steps of endosperm development, the aleurone must expand to accompany the growing kernel. Cell division during these steps is restricted to either the anticlinal or periclinal plane by a pre-prophase band of microtubules. This is in contrast with the starchy endosperm cells which are more flexible in their division orientation (Becraft, 2007). Examples of maize mutants with disorganized planes of division are *xcl1* and *disorganized aleurone layer1* and *2* (Brown & Lemmon, 2007; Lid et al., 2004). Most of the MAL kernels in this study demonstrated this sort of oblique division that is characteristic of cell division mutants (Figure 4.1). Since multiple loci were involved with MAL formation in the two populations, it may be that there are multiple mechanisms for MAL formation in these tropical landraces.

The gene or genes underlying the QTL for MAL formation are currently unknown, but given previous knowledge of MAL formation in maize, candidates can be found. Several minor loci were also associated with modulating aleurone layer numbers (Table 4.6, Table 4.7). A previous study found MALs were linked with chromosome 4 and 9 (de Miranda, 1980). In addition, a membrane vesicle trafficking gene associated with MAL formation, *Supernumerary aleurone layer1* (*Sal1*), is also on chromosome 9. The two chromosome 9 loci from previous studies are proximal to the QTL from chromosome 9 in this study and do not overlap. No gene candidates could be found for the chromosome 6 or chromosome 9 QTL in MAL2 at this time. Near the QTL on chromosome 1 from the MAL1 population are *Globby1* (*Glo1*) and cytokinin oxidase ortholog *Ckx10*. Both of these are cell division-related loci (Costa et al., 2003; Gu et al., 2010)

The largest effect QTL in both populations was on chromosome 8. Using a conservatively wide QTL, there are 138 annotated genes for Mo17 in the MAL1 average aleurone layers QTL using a 1.5-LOD score confidence interval. The MAL2 population had a much narrower confidence interval that was separate from the interval and distal. The 1.5-LOD confidence interval using adjusted number of aleurone layers in MAL2 contains only 27 genes. It

117

may be possible that there are two highly linked genes on chromosome 8 that contribute to MAL formation, but a larger population would need to be assessed to see a significant difference. No obvious gene candidates could be found in the narrow MAL2 QTL. However, this QTL did contain a cytokinin response factor gene designated *AP2-EREBP-transcription factor 96.*

Within the MAL1 QTL, numerous gene candidates with associations with hormones or cell division could be found. Molecular marker *umc1590* from the fine-mapping study was significantly associated with MAL formation, but was not the highest confidence marker. Interestingly though, *umc1590* is within a cytokinin oxidase ortholog *Cko5*. Another gene associated with cell division within the chromosome 8 QTL in MAL1 is Zm00014a029257. This gene has homology to *Arabidopsis thaliana Aurora2*, which has a role in establishing mitotic division planes (Van Damme et al., 2011). A defect in a gene such as this in aleurone may allow cells to divide randomly instead of in an ordered plane, which is similar to what is seen in this study (Figure 4.1).

Furthermore, within the chromosome 8 interval of MAL1 is *Gigantea1* (*Gi1*), a circadian clock-associated negative regulator of flowering time in maize (Bendix et al., 2013). Many tropical MAL landraces are photoperiod sensitive and selection pressure for lines adapted to the Midwest could have incidentally caused variation at this locus. To test whether artificial selection was the cause of the chromosome 8 MAL QTL, average pollination dates were recorded for all MAL2 families in 2019. There was a moderate correlation between pollination date and MAL content ($\rho=0.36$). No MAL loci overlapped with any QTL for pollination date, however. There was a QTL for pollination date that overlapped with the well-known flowering time locus *Vegetative to generative transition1* (*Vgt1*) on the opposite arm of chromosome 8 from the MAL QTL (Castelletti et al., 2014). The recombination frequency between the most significant *Vgt1* marker and the most significant MAL marker in MAL2 is 0.29. The correlation between pollination date and MAL formation could be an artefact of a small population with large linkage blocks.

Another candidate gene for the MAL phenotype on chromosome 8 in MAL1 is *Barren inflorescence1* (*Bif1*). This gene shows defects in organogenesis and results in barren stalks and ears when mutated. The primary mechanism for this gene appears to be defects in auxin transport (Barazesh & McSteen, 2008). Treatment with auxin transport inhibitor NPA resulted in MALs in tissue cultures, so it is logical that a mutant with impaired auxin transport like *Bif1* could also

118

result in MALs. Genetic stocks 805B, 805C, and 805CA were assayed for their aleurone layer content. These stocks are testcrosses, so 50% will be *Bif1* dominant. Out of 26 kernels, none exhibited the MAL phenotype (χ^2 *p*=0.373), indicating that it is most likely not involved with MAL formation. Another auxin-related gene, *Vanishing tassel2* and meristem-related protein *clavata3/esr-related26* are also within the chromosome 8 QTL in MAL1 and could theoretically be involved with MAL formation (Je et al., 2016; Phillips et al., 2011).

An additional goal of this study was to determine markers associated with increased pericarp yield. An earlier study determined that pericarp thickness is partially dominant for thinness and could be controlled by less than six genes (Ito & Brewbaker, 1991). Another study conducted in dent corn found a more quantitative distribution (Helm & Zuber, 1972). Most studies affirm that there is a strong genetic component to pericarp thickness. Broad sense heritability ranged from 0.7 to 0.82 for Korean waxy corn (Choe & Rocheford, 2012), 0.8 to 0.82 for Midwestern inbred lines, 0.92 for a mapping population (Hoenisch & Davis, 1994), 0.595 for a subset of hybrids (Ito & Brewbaker, 1991), and 0.67 for sweetcorn varieties (Wu et al., 2020). Pericarp thickness does not necessarily translate to pericarp yield. This study proves that volume of pericarp and weight of pericarp per kernel are controlled by different factors and are only moderately correlated (Table 4.4, Table 4.6). Pericarp weight and kernel weight were controlled by a QTL on chromosome 5. This QTL is near a QTL for grain weight and yield using Mo17 as a parent in a previous study (Beavis et al., 1994; Veldboom & Lee, 1996). The MAL landrace allele decreased pericarp yield at this locus. Kernel volume, pericarp weight per kernel, and kernel weight was controlled by a QTL at the start of chromosome 6 (Figure 4.3). In contrast to the chromosome 5 QTL, the MAL parent actually increased pericarp yield at this locus (Table 4.6). No gene candidates could be found for these loci at this time. Developing markers for these significant loci should help assist in breeding. Future work should integrate these loci into other backgrounds to see if they have a similar effect. An unexpected finding of this study was that density of the kernel was not strongly correlated with kernel mass nor the pericarp traits (Table 4.4). This low association with pericarp traits allows for the measurement of pericarp traits without kernel hardness confounding the results. The MAL landraces tend to have floury type endosperms, but the gene conferring this endosperm type was unknown. In the QTL analysis, there was a large-effect QTL for density on chromosome 2 where the *Floury endosperm1* gene lies (Table 4.6) (Holding et al., 2007).

119

4.6 Conclusions

Enhancing the yield of aleurone and pericarp tissue has implications for increasing anthocyanin content in the grain, but also has wider implications on human health. The methods developed here efficiently assay pericarp and aleurone yield. Results from this study show there is genetic variation for these traits. The MAL trait increased anthocyanin production 30 to 40%, due to an increase in aleurone yield. Iron content was significantly increased in MAL lines, but future studies will have to test micronutrient levels in near-isogenic MAL lines to reduce genetic variability. The inheritance of MALs seems to be mostly genetically inherited, but multi-year trials and reciprocal-cross experiments demonstrate that MAL formation is complex with multiple loci altering aleurone layer content. A major effect QTL was found on chromosome 8 that warrants further investigation. Genetic markers for this locus were generated that will assist breeders in making selections for new MAL varieties. Pericarp yield traits were also measured in this population. It appears that in this population few factors were involved in increasing pericarp yield. Overall, information from this study will help breeders trying to increase nutrient levels and anthocyanin levels in maize.

4.7 Acknowledgements

4.8 Figures and Tables

Figure 4.1 Microscope images of aleurone phenotypes in this study

A, B) Single aleurone layer kernels from Mo17 (A) and 707G (B) viewed with white light. C, D) Kernels with three aleurone layers from San Martin 119 as viewed with the DAPI filter (C) and Toluidine Blue staining (D). E, F) Kernels with three to four aleurone layers from San Martin 119 as viewed with the DAPI filter (E) and Toluidine Blue staining (F). G, H) Kernels with three to four aleurone layers from the MAL1 yellow corn population as viewed with the DAPI filter (G) and Toluidine Blue staining (H). I) Kernel with heterogeneous layers as viewed with the DAPI filter. J) Member of the MAL2 blue corn population viewed with the TRITC filter. All images were viewed with a BX-66 Fluorescent microscope (Olympus Corporation) and 10X magnification.
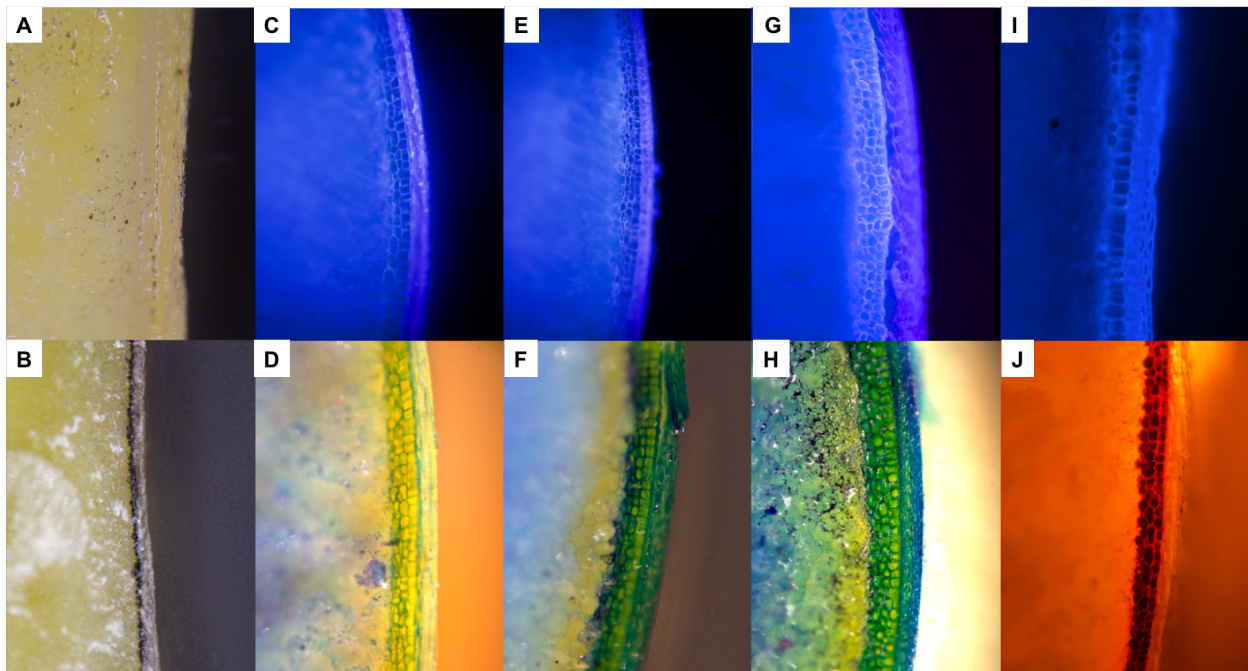
Figure 4.2 Anthocyanin content increases with the number of aleurone layers

A) Histogram of anthocyanin concentration among families in MAL2 in terms of mg anthocyanins per kg weight of kernel. B) The effect of *Intensifier1* alleles and aleurone layer number on anthocyanin content among families in MAL2. Aleurone layers refer to the maximum number of layers possible in each family.
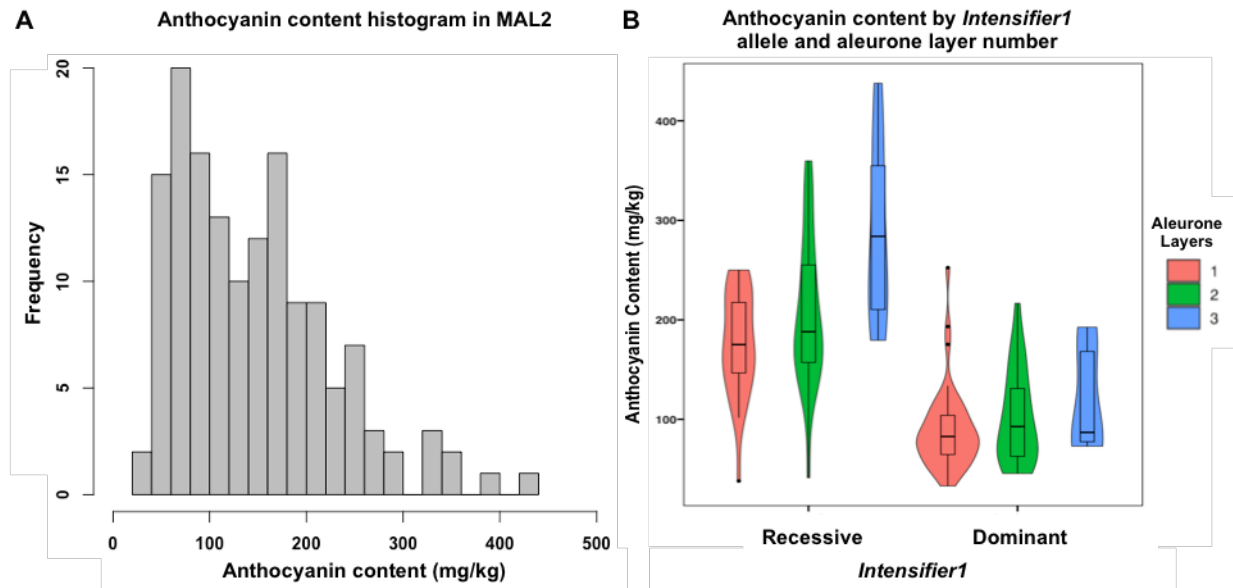
Figure 4.3 Marker-trait associations in the MAL1 yellow corn population

ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers. KWT = kernel weight. MAX = maximum aleurone layers per taxa. PCT = percent pericarp per kernel. PWT = pericarp weight per kernel.
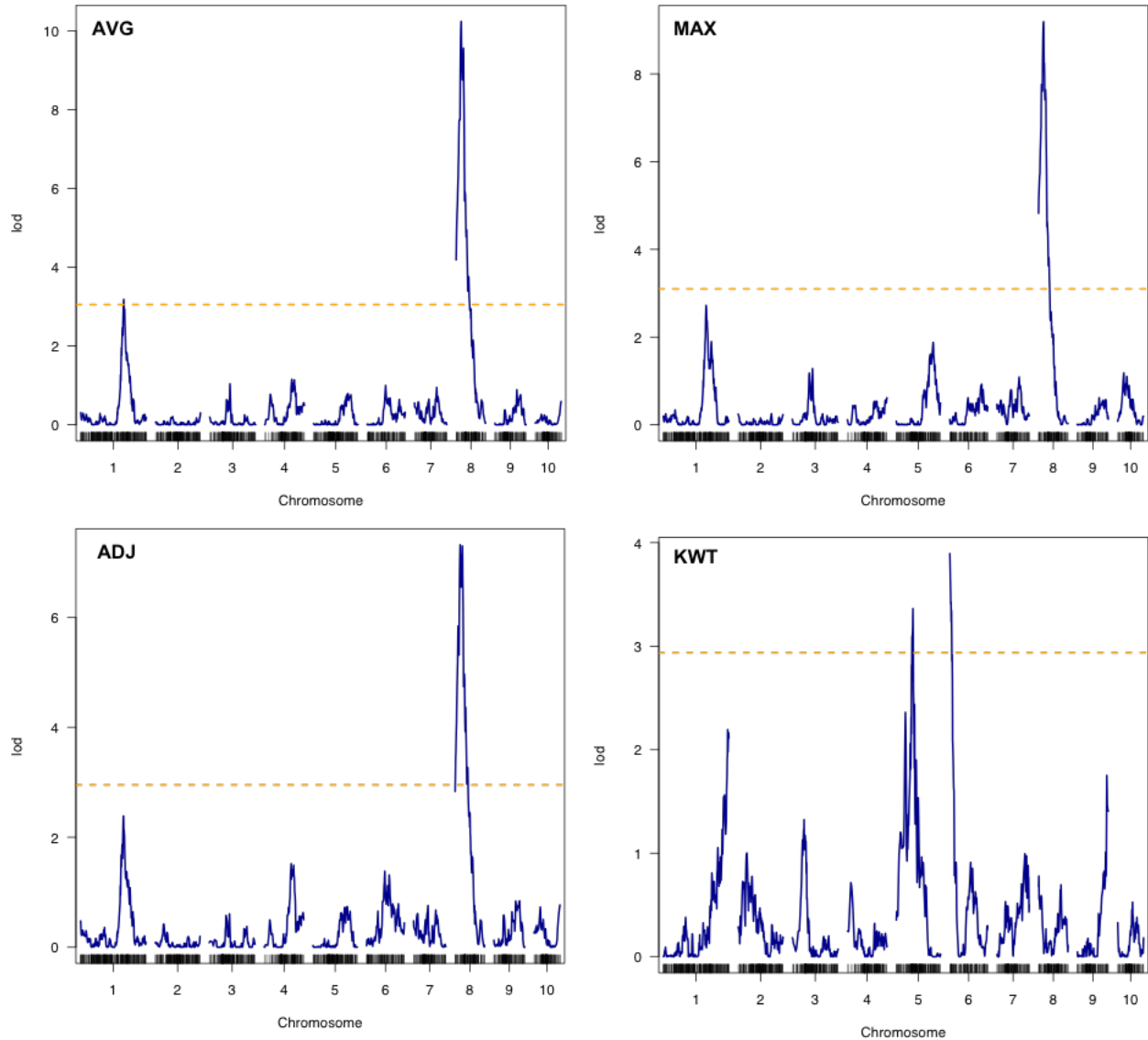
Figure 4.3 (Cont.)
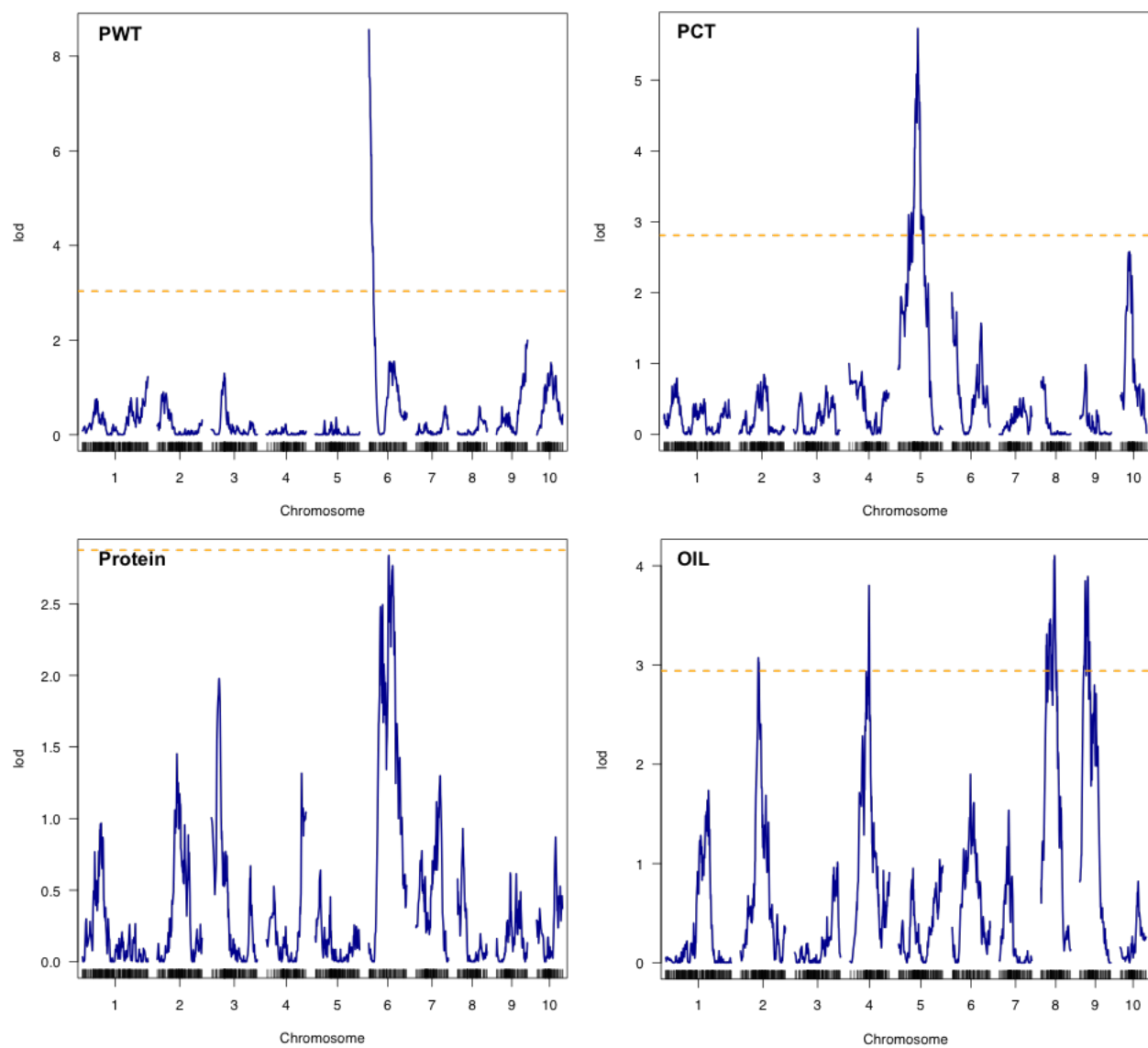
Figure 4.3 (Cont.)
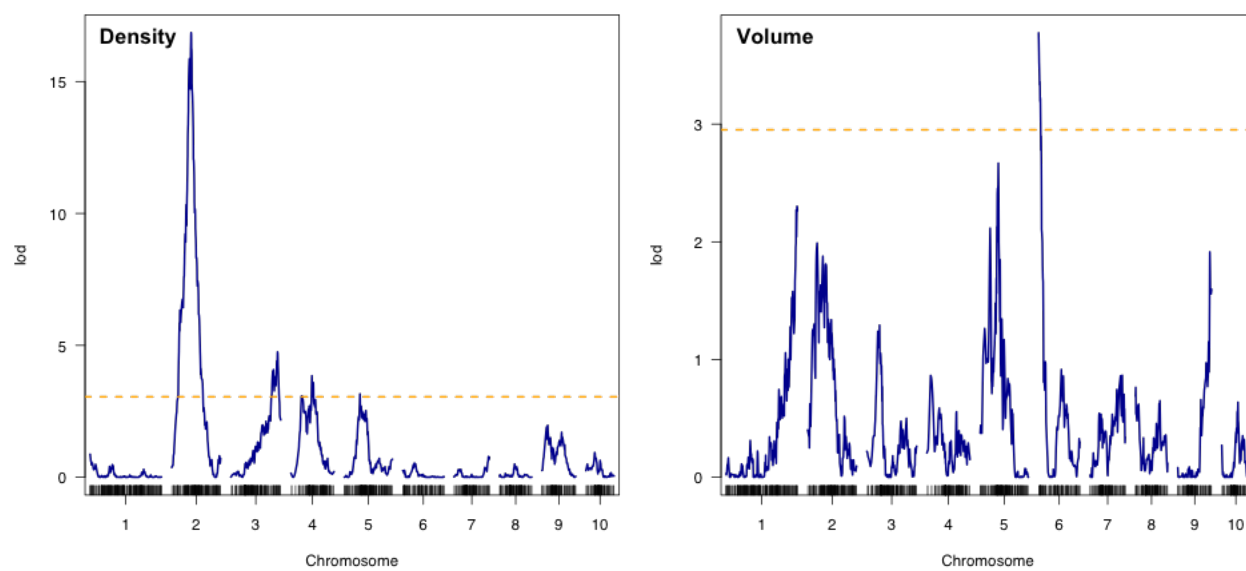
Figure 4.4 Dosage of significant MAL QTL in the MAL1 and MAL2 populations

Shown are boxplots for the average number of aleurone layers per taxa separated by alleles at each significant QTL. Genotypes are represented as +/– for presence or absence of the MAL QTL. A) Genotypes in MAL1 are for markers 1:219,711,643 | 8:21,306,878. B) Genotypes in MAL2 are for markers 6:174,199,500 | 8:27,568,159 | 9:156,275,580.
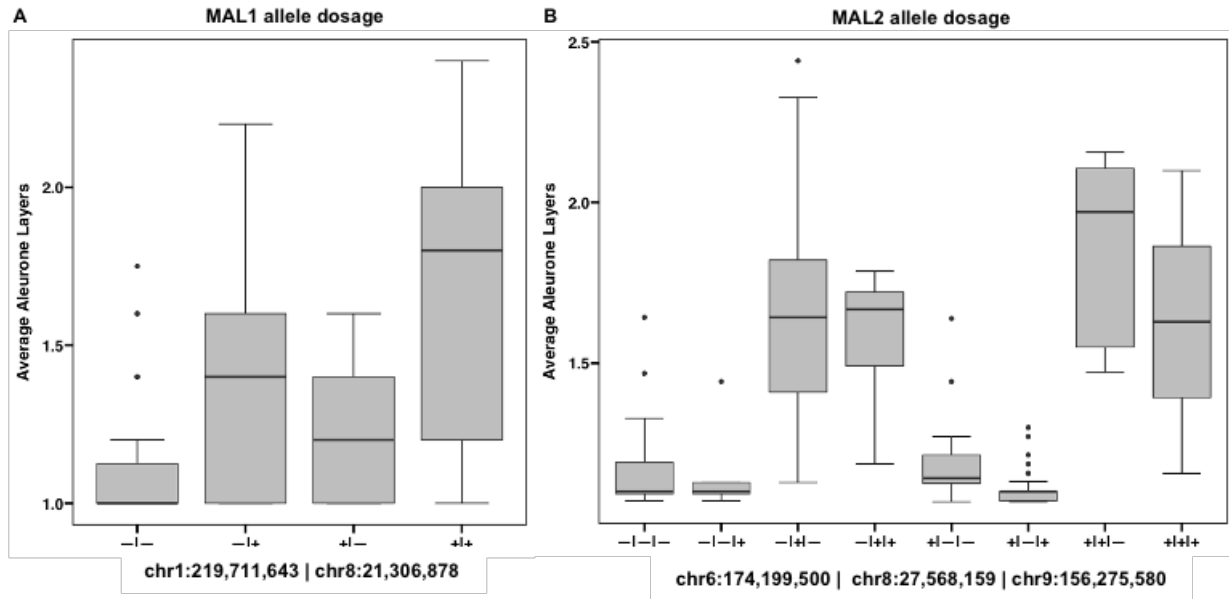
Figure 4.5 Marker-trait associations in the MAL2 blue corn population

ACN = anthocyanin content, ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers. FWT = flowering time measured as average deviation from the average. LogACN = natural log transformation of anthocyanin content. MAX = maximum aleurone layers per taxa.
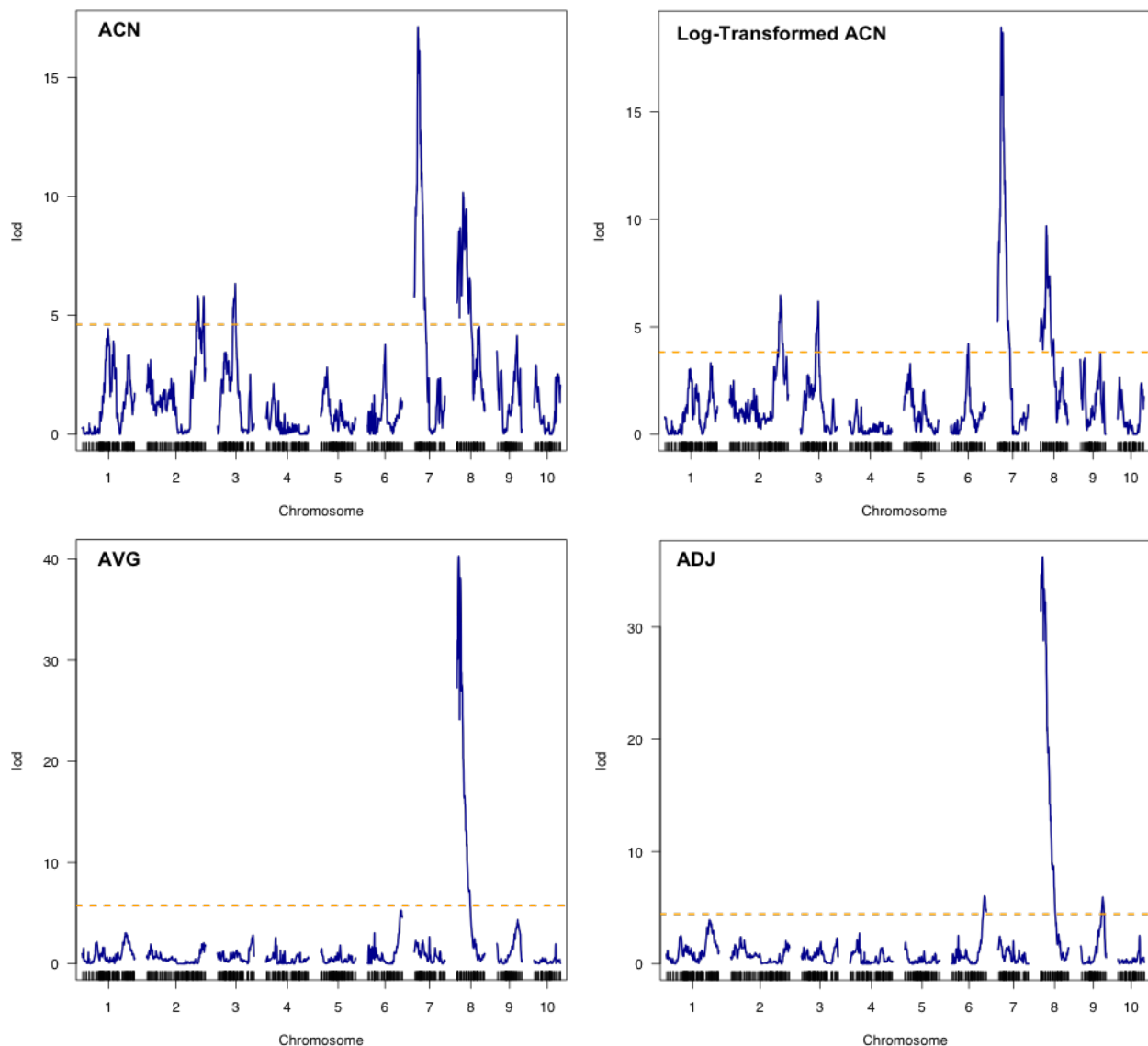
Figure 4.5 (Cont.)
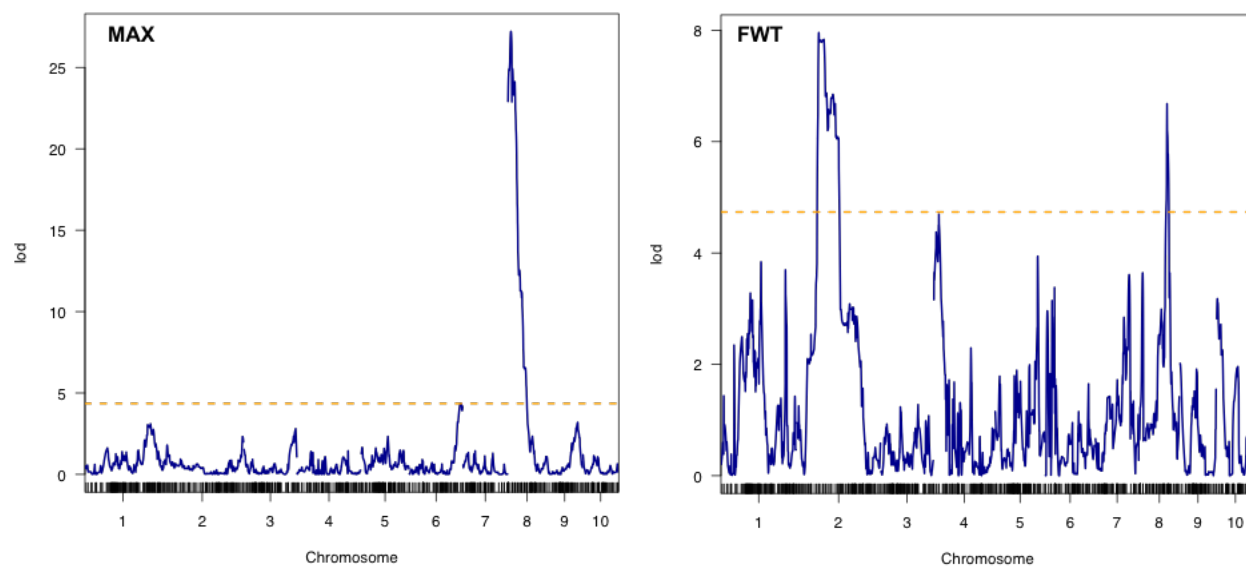
Figure 4.6 SSR Markers validated using the MAL1 population

Examples of SSR marker segregation patterns for MAL1 genotypes: A) *umc1974*, B) *MPMAL2*, C) *MPMAL4*, D) *umc1530*, E) *umc1778*, F) *umc2146*.
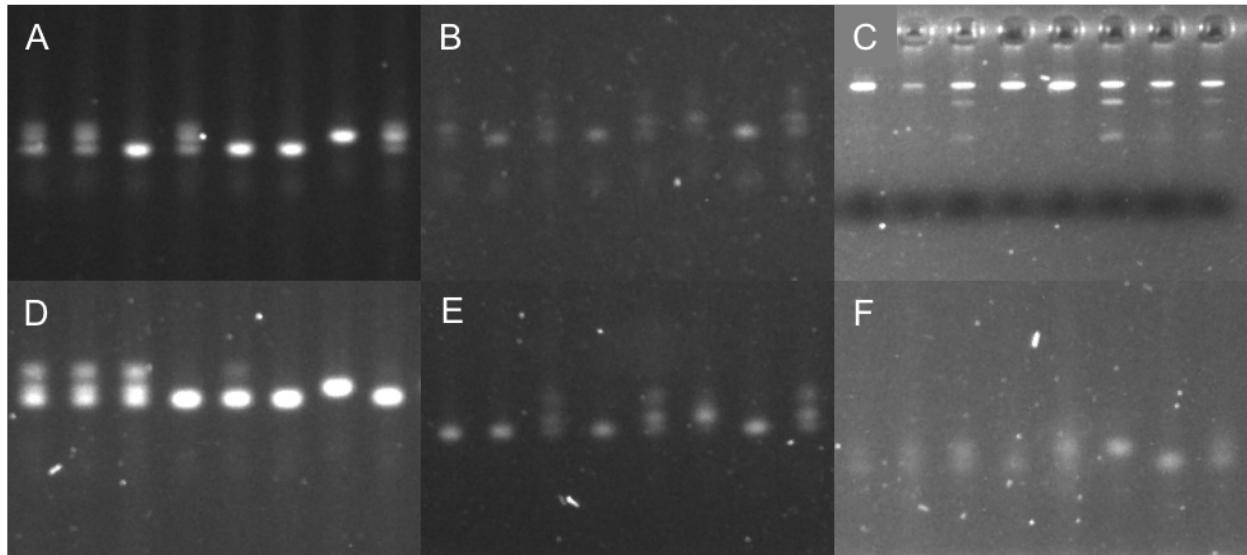
Table 4.1 Structure of the two MAL populations

	MAL1		
Summer 2015	San Martin 105 × Mo17	Crossed	
Summer 2016	F_1 × Mo17	Backcrossed	
Chile 2017	BC_1 Plants, BC_1F_1 Aleurone	Selfed	Phenotyped/Genotyped
Summer 2018	BC_1F_1 Plants, BC_1F_2 Aleurone	Selfed	
	BC_2 × Mo17	Backcrossed	
Summer 2019	BC_1F_2 Plants, $BC_1F_{2:3}$ Aleurone	Selfed	Phenotyped/Fine-Mapped
	BC_3 × Mo17	Backcrossed	
Summer 2020	BC_4 × Mo17	Backcrossed	Phenotyped
	Mo17 × BC_4	Backcrossed	Phenotyped
	$BC_1F_{2:3}$ × (Mo17, 707G)	Crossed	Phenotyped
	(Mo17, 707G) × $BC_1F_{2:3}$	Crossed	Phenotyped
	MAL2		
Summer 2015	San Martin 119 × 707G	Crossed	
Chile 2016	F_1 × 707G	Backcrossed	
Summer 2016	BC_1 Plants, BC_1F_1 Aleurone	Selfed	
Summer 2017	BC_1F_1 Plants, BC_1F_2 Aleurone	Selfed	
Summer 2018	BC_1F_2 Plants, $BC_1F_{2:3}$ Aleurone	Sib-Pollinated	Phenotyped/Genotyped
Summer 2019	$BC_1F_{2:3}$ Families	Sib-Pollinated	Phenotyped
Summer 2020	$BC_1F_{2:3}$ × (Mo17, 707G)	Crossed	Phenotyped
	(Mo17, 707G) × $BC_1F_{2:3}$	Crossed	Phenotyped

Table 4.2 Primer sets used in this study

Marker	Forward	Reverse	B73 RefGen v5 Position	Mo17 position	Annealing Temp (˚C)
umc1974	ACAAGGAGACCCTCCTCAGCTAGT	GTAAGCTGTGGCCATACTACCACC	17,458,078	19,178,123	51
MPMAL1	AATACTAACTGCCTACTGGTACTGT	GTGCAGCTCTCGTTAGAAAGAT	19,517,295	21,311,272	ND
MPMAL2	AATTTCAGTGTGCCGTGCTA	CTTTACCTGCCTGCTTTGGAG	20,286,420	21,935,151	56
MPMAL3	ATAGCAGAACGGTGCCTGTC	CCTGCCTGCTTTGGAGTCTT	20,286,490	21,935,222	ND
MPMAL4	ACATTTGCATGGCACTCACG	ACGTCGAGAGGTGCATAGAG	21,695,168	23,412,424	53
umc1530	GTGGCTCAACCTCTTCCTCCC	GGTTCATGGGGTAGACCAGCA	22,842,164	24,429,960	53
umc1778	GTGAACCATTGTAGCTGTCCCTG	GAGCTCGTACCTGTTCATGAGGAT	24,406,686	25,951,420	57
umc2146	GTCTCCGTCCACCTCCTGTG	GTCATGGGAATGTGCTGGATG	25,667,471	27,095,521	55

ND = Not Determined

Table 4.3 Summary of phenotypes in the two MAL populations

| Trait*		MAL1	MAL2
Prop Kernel MAL	%	38.20	28.80
Prop Samples MAL	%	74.40	60.20
AVG	Avg	1.36	1.20
	H^2	—	0.68
ADJ	Avg	1.49	1.36
	H^2	—	0.86
ACN	Avg (mg kg^{-1})	**Recessive *in1***	
		Single	178.54
		Multiple	238.54
	Avg (mg kg^{-1})	**Dominant *In1***	
		Single	122.61
		Multiple	90.32
	H^2	—	0.95
Protein	Range %	8.45–13.82	—
	Avg %	11.51	—
Oil	Range %	2.51–5.29	—
	Avg %	3.77	—
Density	Range (g cm$^{-3)}$)	1.21–1.36	—
	Avg (g cm$^{-3)}$)	1.29	—
Volume	Range (cm$^{3)}$)	a	—
	Avg (cm$^{3)}$)	2.69	—
KWT	Range (mg)	140.49–469.49	—
	Avg (mg)	348.62	—
PCT	Range %	2.99–7.22	—
	Avg %	4.16	—
PWT	Range (mg)	4.63–18.78	—
	Avg (mg)	12.93	—

* ACN = Anthocyanin content, ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers, KWT = kernel weight, MAX = maximum aleurone layers per taxa, PCT = percent pericarp per kernel, PWT = pericarp weight per kernel.

132

Table 4.4 Correlation matrix of phenotypes in MAL1

	AVG	ADJ	KWT	PWT	PCT	Protein	Oil	Density	Volume
MAX	0.89 ***	0.83 ***	-0.01	0.01	0.03	-0.04	0.13	-0.04	0
AVG		0.93 ***	-0.02	-0.04	-0.01	-0.09	0.24 **	-0.12	0
ADJ			-0.04	-0.06	0	-0.08	0.21 **	-0.10	-0.03
KWT				0.79 ***	-0.27 ***	0.18 *	-0.32 ***	0.27 ***	0.99 ***
PWT					0.35 ***	0.33 ***	-0.38 ***	0.22 **	0.78 ***
PCT						0.26 ***	-0.08	-0.09	-0.27 ***
Protein							-0.32 ***	0.27 ***	0.14
Oil								-0.67 ***	-0.22 **
Density									0.12

Asterisks denote significance where (***) < 0.001, (**) < 0.01, and (*) < 0.05.
ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers, KWT = kernel weight, MAX = maximum aleurone layers per taxa, PCT = percent pericarp per kernel, PWT = pericarp weight per kernel.

Table 4.5 Elemental analysis results

		Counts	P	K	Mg	S	B	Fe	Mn	Cu	Zn	Al
		No.	pct	pct	pct	pct	ppm	ppm	ppm	ppm	ppm	ppm
Single	Avg	1.000	0.2945	0.3625	0.1101	0.1133	2.5000	25.2375	7.1375	0.9313	22.5125	6.7625
	Std dev		0.0175	0.0141	0.0077	0.0049	0.4899	5.5689	0.7984	0.5339	1.2550	3.8181
Multiple	Avg	2.255	0.2791	0.3336	0.1046	0.1078	2.2818	29.6136	6.6955	1.1545	23.0727	9.7045
	Std dev		0.0174	0.0000	0.0035	0.2121	1.5749	0.4435	0.0900	1.0221	1.5428	6.5440
Mo17	Avg	1	0.2810	0.3175	0.09875	0.10875	2.075	21.925	6.1250	1.075	20.0000	6.6000
	Std dev		0.0134	0.0177	0.0060	0.0004	0.1768	1.0960	0.3182	0.1061	0.9192	0.4243

Table 4.6 Summary of QTL analysis results for MAL1

Trait*	Chr	Pos	Interval (cM)	LOD	% R^2
ADJ	8	21306878	28.2–37.9	7.86	19.39
AVG	1	219711643	122.7–151.1	2.97	5.98
	8	21306878	28.1–37.9	9.97	22.14
KD	2	40455703	50.2–53.8	17.24	26.13
	3	230689895	126.4–145.6	5.23	6.70
	4	165684575	45.5–88.7	4.27	5.40
	5	24463369	36.5–78.7	5.04	6.43
KV	6	5669131	0–9.7	4.67	11.75
KWT	5	68327847	31.8–71.4	3.74	8.41
	6	5669131	0–10.1	5.04	11.55
MAX	8	19128271	26–36	9.07	22.01
Oil	2	43972433	46.6–62.5	3.18	6.07
	4	161861261	59.9–77.6	3.59	6.87
	8	125739791	29.8–65.7	4.19	8.09
	9	23085929	21.6–68.7	2.48	4.68
PCT	5	88876985	59.3–70.1	5.74	14.25
PWT	6	6443052	0–3.6	9.91	23.30

* ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers, KWT = kernel weight, MAX = maximum aleurone layers per taxa, PCT = percent pericarp per kernel, PWT = pericarp weight per kernel.

Table 4.7 Summary of QTL analysis results for MAL2

Trait*	Chr	Pos	Interval (cM)	Add	Dom	LOD	% R²
ACN	2	219323824	101.8–118.4	0.188	-0.070	2.16	3.06
	3	187782813	31.6–37.8	-0.071	0.016	0.38	0.53
	7	11678763	6.8–11.2	-0.370	-0.155	8.63	13.55
	8	72353093	6.7–19.9	0.214	0.163	6.70	10.20
ADJ	6	174199500	65–71.3	-0.043	0.002	0.19	0.21
	8	27568159	2.5-–.9	0.468	-0.015	23.87	39.47
	9	156275580	42.7–48.4	-0.030	0.052	1.03	1.17
AFT	2	24208985	30.7–56.1	0.007	-0.126	6.40	15.25
	8	152203849	39.2–42.3	-0.070	0.149	5.11	11.90
AVG	8	27568159	2.8–5	0.302	-0.085	28.59	59.41
LogACN	2	219323824	101.9–108	0.188	-0.071	2.03	2.78
	3	187782813	33–37.7	-0.071	0.016	0.30	0.41
	7	11678763	6.9–11.4	-0.370	-0.155	9.87	15.33
	8	72353093	12–14.9	0.215	0.162	6.11	8.93
Max	8	28934979	2.4–22.9	1.085	-0.159	23.97	53.05

* ACN = Anthocyanin content, ADJ = adjusted number of aleurone layers based on maximum layers possible in individual kernels. AVG = average number of aleurone layers, FWT = flowering time measured as average deviation from the average, MAX = maximum aleurone layers per taxa.

4.9 References

AACC Approved Methods of Analysis. (2009). *Method 44-40.01. Moisture—Modified Vacuum-Oven Method* (11th ed.). Cereals & Grains Association. https://dx.doi.org/10.1094/AACCIntMethod-44-40.01

Andorf, C. M., Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Campbell, D. A., & Sen, T. Z. (2010). The Locus Lookup tool at MaizeGDB: Identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics*, *26*(3), 434–436. https://doi.org/10.1093/bioinformatics/btp556

Barazesh, S., & McSteen, P. (2008). Barren inflorescence1 functions in organogenesis during vegetative and inflorescence development in maize. *Genetics*, *179*(1), 389–401. https://doi.org/10.1534/genetics.107.084079

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beavis, W. D., Smith, O. S., Grant, D., & Fincher, R. (1994). Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop Science*, *34*(4), 882–896. https://doi.org/10.2135/cropsci1994.0011183X003400040010x

Becraft, P. W. (2007). Aleurone cell development. In O.-A. Olsen (Ed.), *Endosperm* (Vol. 8, pp. 45–56). Springer Berlin Heidelberg. https://doi.org/10.1007/7089_2007_108

Becraft, P. W., & Asuncion-Crabb, Y. (2000). Positional cues specify and maintain aleurone cell fate in maize endosperm development. *Development*, *127*(18), 4039–4048.

Becraft, P. W., & Yi, G. (2011). Regulation of aleurone development in cereal grains. *Journal of Experimental Botany*, *62*(5), 1669–1675. https://doi.org/10.1093/jxb/erq372

Bendix, C., Mendoza, J. M., Stanley, D. N., Meeley, R., & Harmon, F. G. (2013). The circadian clock-associated gene gigantea1 affects maize developmental transitions. *Plant, Cell & Environment*, *36*(7), 1379–1390. https://doi.org/10.1111/pce.12067

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab007

Bouis, H. E., Hotz, C., McClafferty, B., Meenakshi, J. V., & Pfeiffer, W. H. (2011). Biofortification: A new tool to reduce micronutrient malnutrition. *Food and Nutrition Bulletin*, *32*(1 Suppl), S31-40. https://doi.org/10.1177/15648265110321S105

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Brown, R. C., & Lemmon, B. E. (2007). The developmental biology of cereal endosperm. In O.-A. Olsen (Ed.), *Endosperm* (Vol. 8, pp. 1–20). Springer Berlin Heidelberg. https://doi.org/10.1007/7089_2007_106

Cannon, E. K. S., Birkett, S. M., Braun, B. L., Kodavali, S., Jennewein, D. M., Yilmaz, A., Antonescu, V., Antonescu, C., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., Campbell, D. A., Andorf, C. M., Andorf, D., Lisch, D., Koch, K. E., McCarty, D. R., Quackenbush, J., Grotewold, E., … Lawrence, C. J. (2011). POPcorn: An online resource providing access to distributed and diverse maize project data. *International Journal of Plant Genomics*, *2011*, e923035. https://doi.org/10.1155/2011/923035

Carter, T. C., & Falconer, D. S. (1951). Stocks for detecting linkage in the mouse, and the theory of their design. *Journal of Genetics*, *50*(2), 307–323. https://doi.org/10.1007/BF02996226

Castelletti, S., Tuberosa, R., Pindo, M., & Salvi, S. (2014). A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL vgt1. *G3 Genes|Genomes|Genetics*, *4*(5), 805–812. https://doi.org/10.1534/g3.114.010686

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Center for Food Safety and Applied Nutrition. (2020). *Report on the Certification of Color Additives: 4th Quarter, Fiscal Year 2020, July 1-September 30*. Food and Drug Administration. https://www.fda.gov/industry/color-certification-reports/report-certification-color-additives-4th-quarter-fiscal-year-2020-july-1-september-30

Chen, Z., Wang, B., Dong, X., Liu, H., Ren, L., Chen, J., Hauck, A., Song, W., & Lai, J. (2014). An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F2 maize population. *BMC Genomics*, *15*(1), 433. https://doi.org/10.1186/1471-2164-15-433

Choe, E., & Rocheford, T. R. (2012). Genetic and QTL analysis of pericarp thickness and ear architecture traits of Korean waxy corn germplasm. *Euphytica*, *183*(2), 243–260. https://doi.org/10.1007/s10681-011-0452-8

Costa, L. M., Gutierrez-Marcos, J. F., Brutnell, T. P., Greenland, A. J., & Dickinson, H. G. (2003). The globby1-1 (glo1-1) mutation disrupts nuclear and cell division in the developing maize seed causing alterations in endosperm cell fate and tissue differentiation. *Development*, *130*(20), 5009–5017. https://doi.org/10.1242/dev.00692

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Danecek, P., Schiffels, S., & Durbin, R. (2014). *Multiallelic calling model in bcftools (-m)*. http://samtools.github.io/bcftools/call-m.pdf

de Miranda, L. (1980). Inheritance and linkages of multiple aleurone layering. *Maize Genetics Cooperation Newsletter*, *54*, 15.

Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13–15.

Duangploy, S., Zuber, M. S., & Cumbie, B. S. (1976). Inheritance of multiple aleurone layering. *Maize Genetics Cooperation Newsletter*, *50*, 90–91.

FAO. (2018). *OECD-FAO Agricultural Outlook 2018-2027: Cereals*. http://www.agri-outlook.org/commodities/Agricultural-Outlook-2018-Cereals.pdf

Forestan, C., Meda, S., & Varotto, S. (2010). ZmPIN1-mediated auxin transport is related to cellular differentiation during maize embryogenesis and endosperm development. *Plant Physiology*, *152*(3), 1373–1390. https://doi.org/10.1104/pp.109.150193

Geisler-Lee, J., & Gallie, D. R. (2005). Aleurone Cell identity is suppressed following connation in maize kernels. *Plant Physiology*, *139*(1), 204–212. https://doi.org/10.1104/pp.105.064295

Gu, R., Fu, J., Guo, S., Duan, F., Wang, Z., Mi, G., & Yuan, L. (2010). Comparative expression and phylogenetic analysis of maize cytokinin dehydrogenase/oxidase (CKX) gene family. *Journal of Plant Growth Regulation*, *29*(4), 428–440. https://doi.org/10.1007/s00344-010-9155-y

He, J., & Giusti, M. M. (2010). Anthocyanins: Natural colorants with health-promoting properties. *Annual Review of Food Science and Technology*, *1*(1), 163–187. https://doi.org/10.1146/annurev.food.080708.100754

Helm, J. L., & Zuber, M. S. (1972). Inheritance of Pericarp Thickness in Corn Belt Maize. *Crop Science*, *12*(4), 428. https://doi.org/10.2135/cropsci1972.0011183X001200040009x

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W. B., Rashid, Z., Usha, R., Saykhedkar, G. R., & Nair, S. K. (2018). Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theoretical and Applied Genetics*, *131*(7), 1443–1457. https://doi.org/10.1007/s00122-018-3089-3

Hoenisch, R. W., & Davis, R. M. (1994). Relationship between kernel pericarp thickness and susceptibility to fusarium ear rot in field corn. *Plant Disease*, *78*(5), 517. https://doi.org/10.1094/PD-78-0517

Holding, D. R., Otegui, M. S., Li, B., Meeley, R. B., Dam, T., Hunter, B. G., Jung, R., & Larkins, B. A. (2007). The maize Floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *The Plant Cell*, *19*(8), 2569–2582. https://doi.org/10.1105/tpc.107.053538

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., Ricci, W. A., Guo, T., Olson, A., Qiu, Y., Della Coletta, R., Tittes, S., Hudson, A. I., Marand, A. P., Wei, S., Lu, Z., Wang, B., Tello-Ruiz, M. K., Piri, R. D., … Dawe, R. K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, *373*(6555), 655–662. https://doi.org/10.1126/science.abg5289

Ito, G. M., & Brewbaker, J. L. (1991). Genetic analysis of pericarp thickness in progenies of eight corn hybrids. *Journal of the American Society for Horticultural Science*, *116*(6), 1072–1077. https://doi.org/10.21273/JASHS.116.6.1072

Je, B. I., Gruel, J., Lee, Y. K., Bommert, P., Arevalo, E. D., Eveland, A. L., Wu, Q., Goldshmidt, A., Meeley, R., Bartlett, M., Komatsu, M., Sakai, H., Jönsson, H., & Jackson, D. (2016). Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nature Genetics*, *48*(7), 785–791. https://doi.org/10.1038/ng.3567

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider,

K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Kessler, S., Seiki, S., & Sinha, N. (2002). Xcl1 causes delayed oblique periclinal cell divisions in developing maize leaves, leading to cellular differentiation by lineage instead of position. *Development*, *129*(8), 1859–1869.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Lid, S. E., Al, R. H., Krekling, T., Meeley, R. B., Ranch, J., Opsahl-Ferstad, H.-G., & Olsen, O.-A. (2004). The maize disorganized aleurone layer 1 and 2 (dil1, dil2) mutants lack control of the mitotic division plane in the aleurone layer of developing endosperm. *Planta*, *218*(3), 370–378. https://doi.org/10.1007/s00425-003-1116-2

Ludwig, S. R., & Wessler, S. R. (1990). Maize R gene family: Tissue-specific helix-loop-helix proteins. *Cell*, *62*(5), 849–851. https://doi.org/10.1016/0092-8674(90)90259-H

M. J. Wolf, Helm, J. L., & Zuber, M. S. (1969). Multiple aleurone layering in maize. *Maize Genetics Cooperation Newsletter*, *43*, 122–123.

Mohamed, A. A., Ashman, R. B., & Kirleis, A. W. (1993). Pericarp thickness and other kernel physical characteristics relate to microwave popping quality of popcorn. *Journal of Food Science*, *58*(2), 342–346. https://doi.org/10.1111/j.1365-2621.1993.tb04271.x

Nelson, O. E., & Chang, M. T. (1974). Effect of multiple aleurone layers on the protein and amino acid content of maize endosperm. *Crop Science*, *14*(3), 374–376. https://doi.org/10.2135/cropsci1974.0011183X001400030010x

Paulsmeyer, M., Brown, P., & Juvik, J. (2018). Discovery of Anthocyanin acyltransferase1 (AAT1) in maize using genotyping-by-sequencing (GBS). *G3: Genes, Genomes, Genetics*, *8*(11), 3669–3678. https://doi.org/10.1534/g3.118.200630

Paulsmeyer, M., Chatham, L., Becker, T., West, M., West, L., & Juvik, J. (2017). Survey of anthocyanin composition and concentration in diverse maize germplasms. *Journal of Agricultural and Food Chemistry*, *65*(21), 4341–4350. https://doi.org/10.1021/acs.jafc.7b00771

Petroni, K., Cominelli, E., Consonni, G., Gusmaroli, G., Gavazzi, G., & Tonelli, C. (2000). The developmental expression of the maize regulatory gene hopi determines germination-dependent anthocyanin accumulation. *Genetics*, *155*(1), 323–336.

Phillips, K. A., Skirpan, A. L., Liu, X., Christensen, A., Slewinski, T. L., Hudson, C., Barazesh, S., Cohen, J. D., Malcomber, S., & McSteen, P. (2011). Vanishing tassel2 encodes a grass-specific tryptophan aminotransferase required for vegetative and reproductive development in maize. *The Plant Cell*, *23*(2), 550–566. https://doi.org/10.1105/tpc.110.075267

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Shen, B., Li, C., Min, Z., Meeley, R. B., Tarczynski, M. C., & Olsen, O.-A. (2003). Sal1 determines the number of aleurone cell layers in maize endosperm and encodes a class E vacuolar sorting protein. *Proceedings of the National Academy of Sciences*, *100*(11), 6552–6557. https://doi.org/10.1073/pnas.0732023100

Singh, M., Liu, S. X., & Vaughn, S. F. (2012). Effect of corn bran as dietary fiber addition on baking and sensory quality. *Biocatalysis and Agricultural Biotechnology*, *1*(4), 348–352. https://doi.org/10.1016/j.bcab.2012.02.005

Somavat, P., Li, Q., de Mejia, E. G., Liu, W., & Singh, V. (2016). Coproduct yield comparisons of purple, blue and yellow dent corn for various milling processes. *Industrial Crops and Products*, *87*, 266–272. https://doi.org/10.1016/j.indcrop.2016.04.062

Somavat, P., Li, Q., Kumar, D., de Mejia, E. G., Liu, W., Rausch, K. D., Juvik, J. A., Tumbleson, M. E., & Singh, V. (2017). A new lab scale corn dry milling protocol generating commercial sized flaking grits for quick estimation of coproduct yield and composition. *Industrial Crops and Products*, *109*, 92–100. https://doi.org/10.1016/j.indcrop.2017.08.013

Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Zhao, H., Song, W., Zhang, M., Cui, Y., Dong, X., Liu, H., Ma, X., Jiao, Y., Wang, B., Wei, X., Stein, J. C., Glaubitz, J. C., Lu, F., Yu, G.,

142

… Lai, J. (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature Genetics*, *50*(9), 1289–1295. https://doi.org/10.1038/s41588-018-0182-0

Tracy, W. F., Chandravadana, P., & Galinat, W. C. (1978). More on pericarp and aleurone thickness in maize and its relatives. *Maize Genetics Cooperation Newsletter*, *52*, 60–62.

Van Damme, D., De Rybel, B., Gudesblat, G., Demidov, D., Grunewald, W., De Smet, I., Houben, A., Beeckman, T., & Russinova, E. (2011). Arabidopsis α aurora kinases function in formative cell division plane orientation. *The Plant Cell*, *23*(11), 4013–4024. https://doi.org/10.1105/tpc.111.089565

Veldboom, L. R., & Lee, M. (1996). Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: I. Grain yield and yield components. *Crop Science*, *36*(5), 1310–1319. https://doi.org/10.2135/cropsci1996.0011183X003600050040x

Welch, R. M., Smith, M. E., Van Campen, D. R., & Schaefer, S. C. (1993). Improving the mineral reserves and protein quality of maize (Zea mays L.) kernels using unique genes. In N. J. Barrow (Ed.), *Plant Nutrition—From Genetic Engineering to Field Practice: Proceedings of the Twelfth International Plant Nutrition Colloquium, 21–26 September 1993, Perth, Western Australia* (pp. 235–238). Springer Netherlands. https://doi.org/10.1007/978-94-011-1880-4_45

Wolf, M. J., Cutler, H. C., Zuber, M. S., & Khoo, U. (1972). Maize with multilayer aleurone of high protein content. *Crop Science*, *12*(4), 440. https://doi.org/10.2135/cropsci1972.0011183X001200040012x

Wu, X., Wang, B., Xie, F., Zhang, L., Gong, J., Zhu, W., Li, X., Feng, F., & Huang, J. (2020). QTL mapping and transcriptome analysis identify candidate genes regulating pericarp thickness in sweet corn. *BMC Plant Biology*, *20*. https://doi.org/10.1186/s12870-020-2295-8

143

CHAPTER 5: BREEDING FOR ENHANCED LEVELS OF ANTHOCYANINS IN TEMPERATE MAIZE

5.1 Abstract

Maize has a promising future as an economical source of natural colorants. Extensive genetic diversity exists for purple corn. Peruvian landraces contain the highest anthocyanin content but are not adapted to growing in a temperate climate. In this study is a proposed method for reducing photoperiodicity of tropical maize landraces to assist in breeding of these materials. Numerous populations of purple corn were made from crosses of purple corn to previously elite off-patent yellow dent lines. One population was segregating for numerous compositional differences and should be investigated in the future. Deleterious genetic load in purple corn landraces inhibits breeding progress and therefore a backcrossing scheme should be employed. Three near-isogenic purple corn lines were created as high seed set purple donor lines. To assist in breeding for purple corn, some molecular markers were also selected and tested in a population for linkage to pigmentation. Pelargonidin-rich varieties are of interest to the natural colorant industry. Discussed are several methods used to breed for purple corn with high amounts of pelargonidin anthocyanin derivatives. Finally, numerous purple corn hybrids were created and tested over three years and in multiple locations for the selection of promising inbred lines. Anthocyanin yield was highly heritable in the varieties tested. The top purple corn varieties developed matched Midwest USA average yield without a reduction in anthocyanin content. This study proves that maize can be an economical source of natural colorants and that integrating tropical germplasm into temperate yellow dent varieties may have a positive effect on grain yield.

5.2 Introduction

Consumers are becoming increasingly conscious of nutrition labels and the ingredients used by the food industry. With this comes a distrust for synthetic dyes ubiquitous in processed foods and beverages. In particular, FD&C Red No. 40 accounts for the majority of synthetic dyes produced, with nearly 6 million pounds being certified every year in the USA (Center for Food Safety and Applied Nutrition, 2020). Natural substitutions for FD&C Red No. 40 are in the form of betalains and anthocyanins. Betalains are produced by many plant species within the Caryophyllales order. Betalains are mutually exclusive of anthocyanins in this clade and seem to have overlapping functions for the plant (Khan & Giridhar, 2015). Anthocyanins are more

144

prevalent pigments in the plant kingdom and seem to be monophyletic from even the most primitive vascular plants (Markham, 1988). The switch to all-natural colorants in foods and beverages is a challenge since the most common sources of anthocyanins and betalains used today are perishable fruits and vegetables. Purple corn (*Zea mays* L.) is an abundant source of anthocyanins and can be stored long-term without degradation of quality. Maize has an attractive future as an economical source of natural colorants because of the economy of scale of maize production. No special equipment or labor is needed to produce the crop. In addition, pigment may be integrated into the supply chain as a co-product to reduce costs for anthocyanin production. In purple corn, pigment is produced in the outer layer of the kernel referred to as the pericarp. This portion is dietarily insignificant and is usually discarded as the first step of the dry-milling or dry grind processes (Somavat et al., 2016). Gathering this portion of the kernel and utilizing the rest as food, fuel, or feed has potential to make purple corn an economically feasible source of natural colorants. Purple corn is culturally important for many of the peoples in Latin America. Products from purple corn range from foods and beverages to textiles (Petroni et al., 2014). These Latin American landraces are usually open-pollinated varieties not improved by modern plant breeding techniques. The focus of this study was to simultaneously increase the anthocyanin content and grain yield of purple corn varieties by integrating temperate yellow dent corn germplasm.

5.3 Materials and Methods

5.3.1 Plant materials

The major goal of this project was to develop high grain- and anthocyanin-yielding temperate dent varieties for the Corn Belt region of Midwest USA. Maize has numerous genetic resources available for breeders. For those interested in grain yield, previously elite germplasm that used to be covered by Plant Variety Protection (PVP) laws are now free to use by breeders once they are off patent. A previous study investigated grain yield traits in ex-PVP varieties that represent the entire basis for diversity in current commercial cultivars of maize in the Midwest USA. Results of this study provided a list of several ex-PVP inbreds that have superior combining ability (Hauck et al., 2014). In this current study, the main inbreds of focus were B73 and Mo17, two public varieties that have defined their heterotic pools for decades. In addition, LH123 (*Ht*) and PHG84 were also chosen because it was found they confer high additive, line heterosis, and general combining effects for yield traits (Hauck et al., 2014). In addition to high

145

grain-yielding varieties, genetic resources are available for other traits such as pigmentation. When the project started in 2014, a collection of 398 varieties of pigmented maize from all over the world were assayed for their pigment concentrations. It was found that Peruvian landraces were superior in their pigment content and thus would be a good basis for purple corn varieties (Paulsmeyer et al., 2017). Arequipa 35 (Ames 8488), Arequipa 115 (PI 514866), and Arequipa 204 (PI 571427) from the North Central Regional Plant Introduction Station in Ames, IA, USA were chosen as purple corn parents since they consistently produced high anthocyanin yield. In addition, a unique line named Apache Red Cob (PI 213730) from Siskiyou Seeds (Williams, OR, USA) was chosen based on its unique composition. This Southwest USA landrace is capable of producing high amounts of pelargonidin anthocyanins, which are of interest as a source of water-soluble orange colors for the natural colorant industry (Chatham et al., 2018). Two stocks from the Maize Genetics Cooperation Stock Center (MGCSC; Urbana, IL, USA) designated 506B and X131I were utilized for purple corn hybrids too. To expand the purple corn germplasm in our collection, two purple landraces were later integrated into the breeding program. One was referred to as "Maize Morado" and was purchased from Angelina's Gourmet (Swanson, CT, USA), and the other was "Inca Purple" (Amazonas Imports, Irwindale, CA, USA) purchased at a local grocery store. Finally, two varieties were an attempt to expand the purple corn market to a short growing season. The Iodent heterotic pool founder PH207 and LH84 were chosen as parents because they consistently flowered earlier than B73.

5.3.2 Field evaluations

Numerous field locations were utilized during the course of the study. The main nursery location was at the University of Illinois Vegetable Crops Research farm (UIUC) in Champaign, IL, USA (40°04'36.1"N 88°14'24.2"W). All breeding materials were developed here along with yield trials throughout the course of the yield trial study. To accelerate breeding, various winter nurseries were utilized. In 2015 to 2017, selected materials were taken to Buin, San Maipo, Chile. In 2018 on, selected materials were taken to Puerto Rico, USA. Yield trials were held at the UIUC location and other various locations which will be described. In 2018, an additional location in Gays, IL, USA (39°23'47.2"N 88°32'30.2"W) was kindly provided by Corteva Agrisciences (Johnston, IA, USA). Grain yield data from the combine was dubious at this location due to severe lodging, but three to five purple corn ears were harvested from each plot to calculate anthocyanin yield. In 2019, the Northwestern Illinois Agricultural Research and

146

Demonstration Center in Monmouth, IL, USA (40°56'10.5"N 90°43'13.1"W) was used for an additional yield trial location. Severe lodging damaged this yield trial and therefore grain yield data was not collected, but five ears from each plot were harvested to calculate anthocyanin yield. For 2018 and 2019, anthocyanin yield was estimated by gathering a subset of grain (approximately 1 kg) from the combine. In 2020, a major thunderstorm damaged the yield trial at the UIUC location and therefore plots were harvested by hand. Each entry of the yield trial was planted in double-row plots measuring 0.762 m × 5.334 m. Hand-harvested maize was dried in a forced air oven at approximately 40 ˚C for at least a week before analysis.

5.3.3 Shade houses for photoperiod sensitive landraces

The first year growing Peruvian landraces in the collection it was apparent that many of these tropical landraces are photoperiod sensitive and will only flower on short days. Daylight in Peru fluctuates between 11 to 13 hours throughout the year. In Illinois, this daylight period is the end of August or September—much too late for successful crossing. Many of the landraces grew over 4 m tall, which complicated crossing (Figure 5.1). Furthermore, the purple corn landraces in the collection were segregating for purple pericarp as was evidenced in their original source seed packets. This frustrated breeding efforts since the purple corn did not produce ears in Illinois nurseries and the result of the cross would not be known until the next generation. After limited success in 2014, "shade houses" made of opaque black tarps were constructed for the summer of 2015. Each shade house is 0.91 m tall, 1.22 m wide, and 3.05 m long. The runners were made with treated lumber to withstand weather and the supports were made of 19.05 mm diameter metal pipe bent to the proper specifications. Twenty-four plants could fit in a single house. The tarps were fastened using binder clips and the shade houses were held down by numerous cinder blocks. The black tarps were taken off at 9:00 in the morning and put back on 17:00 in the afternoon to stimulate the short-day nature of the plants. The treatment was applied for 5 to 6 weeks, or when the plants reached the V9 or V10 stage and started overgrowing the shade houses.

5.3.4 Backcrossing as a tool for inbred development

Despite issues with photoperiodicity, a few successful crosses were made in the summer of 2014. Hybrids of Arequipa 35 and 204 with both B73 and Mo17 were selfed and backcrossed in the winter nursery in 2015. It was immediately apparent in the summer of 2015 that there was high deleterious genetic load within the Peruvian landraces and selfing lines with high

147

contributions of tropical germplasm was not advantageous. Therefore, a backcrossing scheme was used to adapt these varieties. Arequipa 35 lines were backcrossed five times with their respective yellow inbred parents (B73 and Mo17), and the Inca Purple × LH123 population was backcrossed five times with LH123. During this process, the highest anthocyanin content variety was selected. Logistically, it was ideal to use the purple corn as the female in the cross so that the purple color could be easily tracked each backcross generation. B73 and Mo17 were chosen as recurrent parents because they represent the basis for much of the commercial USA dent corn. In a study of temperate corn germplasm, it was found that 41.3% and 11.3% of modern cultivars in the USA are derived from B73 and Mo17 (Mikel, 2011). LH123 was chosen based on general combining ability in a previous study (Hauck et al., 2014).

Aside from the advanced backcross lines, several other populations were developed. In general, selections were made based on mass of anthocyanins produced *per ear*—a combination of anthocyanin yield with grain yield—, along with plant performance in the field. Fewer backcrosses were used to create lines high in anthocyanin yield. The populations developed were Arequipa 35 crossed to B73, LH123, Mo17, and PHG84; Apache Red Cob crossed to B73, LH123, Mo17, and PHG84; Maize Morado crossed to B73, LH123, Mo17, PH207, and PHG84; Arequipa 204 crossed to B73, LH84, and Mo17; LH1 crossed to Arequipa 115; Inca Purple crossed to LH123; and X131I crossed to Arequipa 35. The most promising candidates were then crossed to testers to evaluate hybrid yield. Initially, the testers were yellow inbred lines because purple inbreds were not available. In those cases, B73 and Mo17 were used where appropriate. Further along in the program, advanced backcross inbreds of B73 and Mo17 were used as purple testcrosses where appropriate. The full pedigree information with male and female contributions to purple corn hybrids tested in this study are available in Table 5.1.

5.3.5 Anthocyanin phenotypic evaluation methods

Visual determination of anthocyanin concentration is impossible with the human eye. Pericarp thickness, anthocyanin extractability, and pigment density all obfuscate the measurement of anthocyanin concentration by eye. A high-throughput screening method was made for anthocyanin concentration evaluation. Anthocyanins were extracted as previously described (Chatham & Juvik, 2021). If the sample was to be used for chromatography, the crude extract was passed through a 0.45 μm wwPTFE 96-well plate filter and stored in a 0.8 mL deep-well 96-well plate. If the sample was for absorbance readings or the pH differential method, the

148

extract was decanted into a 2 mL microcentrifuge tube fitted with a 5 μm polyester mesh cell strainer to remove large debris. Extract passing through the strainer was centrifuged at greater than 8000 $x\ g$ for 10 min to remove suspended starch for 96-well optical plate analysis. Only purple corn was chosen for analysis even though many purple corn hybrids segregated for the ability to produce anthocyanin pigmentation. Every batch of extractions included two check varieties grown in their respective years at the UIUC location to calculate variation due to differences in days of extractions. Anthocyanin concentration throughout this study is expressed as mg anthocyanins per kg weight of maize powder (mg/kg). In 2017 and 2018, anthocyanin extracts were run on an L-7100 series HPLC (Hitachi High-Tech Corporation) using a Poroshell 120 SB-C_{18} (100 mm x 4.6 mm, 2.7 μm) column (Agilent Technologies) with a method previously described (Paulsmeyer et al., 2017). In 2019 moving forward, anthocyanin extracts were run on an Agilent 1290 Infinity II UHPLC using a Poroshell 120 SB-C_{18} (3 mm x 100 mm, 1.9 μm) column (Agilent Technologies) maintained at 50 ˚C. The mobile phase consisted of 5% formic acid (A) and acetonitrile (B) at a flow rate of 1.7 mL/min in a gradient of 4% B for 1 min, 4% to 14% B in 10 min, 14% to 25% in 3 min, 25% to 100% B in 30 s, and back to 4% B in 30 s. The column was equilibrated for 3 min between each sample. The detector was monitored at 520 nm. The UHPLC was calibrated using a linear gradient of 15.63 to 500 μg/mL cyanidin 3-glucoside standard (Polyphenol Laboratories AS). Proportion of flavonol-anthocyanin condensed forms was calculated by summing peak areas before cyanidin 3-glucoside and dividing by total integrated peak area. Proportion of acylation was calculated by subtracting pelargonidin 3-glucoside, peonidin 3-glucoside, cyanidin 3-glucoside, and condensed form peak areas from the total peak area and dividing by total integrated peak area. Anthocyanin extracts in 2020 were not run on the UHPLC, but instead were quantified using the AOAC standard pH differential protocol (Lee et al., 2005). Briefly, anthocyanin extracts were diluted in 10 mM potassium chloride (pH 1.0) and 0.4 M sodium acetate (pH 4.5) in duplicate in 96-well optical plates and run on a Biotek Synergy II plate reader. The A520 nm and A700 nm were measured to calculate anthocyanin concentration based on the molar extinction coefficient of cyanidin 3-glucoside (Lee et al., 2005). The absorbance method was also used to screen for pigment content in the breeding program for more rapid selection.

149

5.3.6 Phenotypic and genotypic markers associated with purple corn pigmentation

Culling non-purple maize is a critical step in purple corn breeding. Once purple corn is pollinated, pigment begins to develop around 10 days after pollination (Chatham et al., 2019). However, pollinating maize takes time and equipment; therefore, a phenotypic or genotypic determination method prior to pollination would increase breeding efficiency. Distinguishing between purple corn and non-purple corn phenotypically cannot be done at the seedling stage. Mature plants have subtle, yet unique pigmentation phenotypes. The entire plant is not pigmented in most purple corn. Instead, the auricle, internodes, base of tassel glumes, and patches of the husk are lightly to darkly pigmented in purple corn lines. The degree and pattern of pigmentation in these tissues is different in every background. Some examples of purple corn plant phenotypes are shown in Figure 5.3.

Marker-assisted selection may be more efficient in certain scenarios. The genes activating anthocyanin synthesis in purple corn are known. Combinations of certain alleles of *Booster1* (*B1*) or *Colored1* (*R1*) along with *Plant color1* (*Pl1*) are necessary to activate pigment synthesis in the pericarp. In this study, several PCR-based genetic markers were chosen to assay for purple corn regulatory genes. The SSR markers for *B1* and *Pl1* were publicly available from the Maize Genetics and Genomics Database (https://maizegdb.org/data_center/ssr) (Andorf et al., 2010). The SSR marker for *Pl1*, *umc1014*, is within the gene itself and is highly variable among B73, Mo17, LH123, Inca Purple, and Ames 8488. The *B1* SSR marker, *umc2195*, has not been validated in the purple corn varieties, but is highly polymorphic among temperate inbred lines (results not shown). For R1, a custom-made indel marker was designed based on polymorphic sequence content between genomic sequences of B73 and Mo17 (Chen et al., 2014; Jiao et al., 2017; Sen et al., 2010). The *R1* indel marker flanks a deletion in the *R1* locus between B73 and Mo17. An alternative *R1* marker, *umc2528*, is 2 Mb downstream from the *R1* gene in B73 RefGen v4 (Cannon et al., 2011; Jiao et al., 2017).

Primer sequences for molecular markers used in this study are as follows: *umc2195* (*B1*)-Forward, 5'- CTCTCGGCTTCTTCTACACGCTAC-3'; *umc2195* (*B1*)-Reverse, 5'-ATCCTTGACGATGAAGTCGTGG-3'; *umc1014* (*Pl1*)-Forward, 5'-GAAAGTCGATCGAGAGACCCTG-3'; *umc1014* (*Pl1*)-Reverse, 5'-CCCTCTCTTCACCCCTTCCTT-3'; *umc2528* (*R1*)-Forward, 5'-CTCATCAACATGCAAAGGACGTAG-3'; *umc2528* (*R1*)-Reverse, 5'-

150

ATTCAAATGCCTCTAAGCTAGCCG-3'; *R1*-indel-Forward, 5'-
ACCCTTTGAGGGACCAGAGA-3'; *R1*-indel-Reverse, 5'- CGTGATCATCCCAGCGTAGT-3'.
The reaction consisted of 1.2 M betaine, 600 nm of each primer, and 1 to 100 ng DNA in *Taq* 5X
Master Mix (New England Biolabs). The thermal cycler program was the same for *B1* and *Pl1*
except for the annealing temperature. The annealing temperature for each marker are as follows:
umc2528, 50 ˚C; *umc2195*, 53 ˚C; and *umc1014*, 56 ˚C. The thermal cycler program used started
with a 1 min initial denaturation at 94 ˚C followed by 35 cycles of 30 s at 94 ˚C, 30 s at the
annealing temperature, and 30 s at 72 ˚C, and ending with 3 min at 72 ˚C. The *R1* marker utilized
a different protocol because the amplified fragment was AT-rich and therefore had a low melting
temperature. The thermal cycler program used for *R1* started with a 1 min initial denaturation at
94 ˚C followed by 35 cycles of 30 s at 94 ˚C, 30 s at 51 ˚C, and 30 s at 65 ˚C, and ending with 3
min at 65 ˚C. The PCR products were separated on a 4% SFR agarose gel (VWR) in ice-cold
TBE at 5.4 V/cm. Running the gel for 1.5 to 2 hours was sufficient for all SSR markers.

5.3.7 Development of pelargonidin-rich varieties

Maize has vast anthocyanin compositional diversity that is important for the natural
colorant industry. The majority of anthocyanins in maize are cyanidin derivatives, which are red-
orange to red in acidified extracts. The critical enzymatic step to creating these anthocyanins is a
gene called *Purple aleurone1* (*Pr1*), which is a flavonoid 3'-hydroxylase. Maize has many
cultivars that are lacking a functional *Pr1* gene and accumulate pelargonidin forms of
anthocyanins instead. These pigments are unique in that they confer a more orange color in
acidified extracts. Natural orange colors utilized by the natural colorant industry are usually
lipid-soluble carotenoids. Many food and beverage applications require water-soluble colors, so
pr1 recessive varieties may be the key to providing these hues (Chatham et al., 2018). Apache
Red Cob is a known *pr1* recessive variety. It is also known that this particular variety has a 17-bp
deletion in *pr1* that inactivates the gene (Chatham & Juvik, 2021). Apache Red Cob was
backcrossed once to B73, LH123, and Mo17 to create a new *pr1* recessive variety that has higher
grain yield than Apache Red Cob and better combining ability. Because there was no testcross
parent with a recessive *pr1*, genetic stock 506B from the MGCSC was used to test hybrid
performance (Table 5.1).

Phenotyping for the pelargonidin-dominant trait is relatively difficult since *pr1* alleles
cannot be discriminated from *Pr1* alleles visually. Liquid chromatography can separate

151

compounds within a sample and determine if pelargonidin or cyanidin forms dominate in a sample. However, chromatography takes more time and reagents to analyze. An alternative is to look at the spectra of an acidified corn extract. The λ_{max} for cyanidin is approximately 520 nm, while the λ_{max} for pelargonidin is approximately 510 nm at a pH less than 2. The absorbance ratio at A510 to A520 (A510/A520) should indicate whether more pelargonidin or more cyanidin forms of anthocyanins are within the sample. In the BC_1S_1 stage, purple corn was extracted using maize kernel powder as indicated above (5.3.5 Anthocyanin phenotypic evaluation methods) and the A510/A520 ratios were determined in 96-well optical plates. There were two major distributions with a cutoff around 1.08 (Figure 5.6). Samples with a A510/A520 of 1.09 and above were considered *pr1* recessive.

An alternative method for determining recessive alleles of *pr1* would be to use MAS. The deletion in the *pr1* allele of Apache Red Cob allows for simple differentiation by PCR. A PCR-based assay for *Pr1* was provided by Dr. Laura Chatham and Dr. Tony Studer (Personal Communication). Primer sequences flanking the indel are: *Pr1*-Forward, 5'-GCGAGGGTGTAGTGGTTCATCA-3'; and *Pr1*-Reverse, 5'-CTGCAGCAGGTTTTGTGAAA-3'. The thermal cycler program and reaction conditions were the same as *B1* and *Pl1* above, but the annealing temperature of the *Pr1* primers is 52 ˚C.

5.3.8 Statistical analyses

Analysis of variance was conducted in R with the *lme4* package using the REML (restricted maximum likelihood) method (Bates et al., 2015, p. 4; R Core Team, 2019). Significance tests for random terms were performed via a drop-one likelihood ratio test using the *lmerTest* package (Kuznetsova et al., 2017). Anthocyanin concentrations and grain yields of less than 50 were discarded from the analysis since those results may be unreliable. The linear mixed model for grain yield is shown below in Equation 1. In this equation, term y is the phenotypic values, μ is the grand mean, β is the random effect of year, γ is the random effect of genotype, and ε is the residual. The α term is the random effect for environment. This term represents location × year for the full datasets or year for the UIUC-only yield datasets. The term αγ is the random genotype × environment interaction term. The linear mixed model for anthocyanin yield in Equation 2 is similar to Equation 1, but includes a random term for batch effect, or δ. This term represents the variation in anthocyanin yield for the standard check samples associated with days of extractions or days quantifying anthocyanins.

152

Equation 1: $y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \varepsilon_{(ijk)l}$

Equation 2: $y_{ijklm} = \mu + \alpha_i + \beta_{j(i)} + \delta_{k(ij)} + \gamma_l + \alpha\gamma_{il} + \varepsilon_{(ijkl)m}$

Estimated variance components were used in Equation 3 to calculate broad-sense heritability. In Equation 3, e is the number of environments and r is the average number of replications per environment.

Equation 3: $H^2 = \dfrac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{g \times e}^2}{e} + \frac{\sigma_e^2}{r*e}}$

5.4 Results

5.4.1 Effectiveness of shade houses on reducing photoperiodicity

The shade houses were resilient against severe thunderstorms and powerful gusts of wind (greater than 60 km/hr tested). A few weeks after shade house treatments, the plants initiated tassels. Many successful crosses could be made to delay-planted temperate corn. To maintain purple corn landrace identity, the tropical corn varieties were used as males and crossed to several yellow temperate inbreds. The anthocyanin content of the females was determined when possible, and the highest content individual crossed was used to make the new breeding materials. New populations were created from a cross of Arequipa 35 and Maize Morado to B73, LH123, Mo17, and PHG84 resulting from the shade house experiments. The hybrids generated in 2015 were evaluated in the 2016 winter nursery to determine which to move forward in the breeding program. The purple corn line from the shade house experiment that conferred the highest amount of anthocyanin was Maize Morado with 416 mg/kg and 449 mg/kg when crossed with Mo17 and PHG84, respectively. These lines were backcrossed and selfed in the 2016 Chile winter nursery and developed into new inbred development lines.

5.4.2 Backcrossing as a tool for pre-breeding and inbred development

The selfed lines maintained high anthocyanin content after selfing five generations and were kept as high anthocyanin content donors. Selection for these lines was purely based on whether they could produce adequate grain at UIUC. Repeated selfing was deleterious to many inbreds in this program and thus most inbred lines were discarded. One inbred line that had not undergone backcrossing had very high amounts of flavonol-anthocyanin condensed form dimers (approximately 46%) and very little acylation (approximately 12%). This was a unique finding since it appeared that acylation was lost after numerous generations of selfing. This indicates that

153

acylation may be spontaneously paramutagenic either due to the acyltransferase being silenced or the acylated anthocyanin transporter.

The initial inbreds that came from the Arequipa 35 advanced backcross populations were not ideal but produced adequate grain. Anthocyanin content averaged 516 mg/kg and 210 mg/kg in the B73 and Mo17 advanced backcross lines, respectively. The Inca Purple advanced backcross lines are still in development, but the BC_5S_2 generation ranged from 347 to 616 mg/kg in 2020. The Mo17 BC_5 line was very similar morphologically and performed very well in terms of grain yield but flowered earlier than Mo17. This made it difficult to cross to B73, especially in the winter nursery where hot summer days and warm nights accelerate pollen shed. It was possible to cross B73 BC_5 pollen to Mo17 BC_5 ears, which is typically the opposite direction for stiff stalk and non-stiff stalk variety development. The B73 BC_5 line developed performed much more poorly than B73 and was not necessarily morphologically similar. Kernels appeared disorganized in this BC_5 line (Figure 5.3). The kernels were susceptible to mold, especially *Fusarium* species. This indicates there may be a factor linked to pericarp pigmentation that confers susceptibility to *Fusarium* spp. Because of these drawbacks, new Mo17 and B73 BC_5 lines are currently in development. These upcoming advanced backcross lines are valuable to breeders as purple donor stocks for future breeding programs looking to increase grain yield. Additionally, these purple donor stocks can be used as females for high anthocyanin yielding inbred lines that may not have good seed set since the BC_5 purple donors tend to produce more grain.

5.4.3 Rapid selection of purple corn inbreds

Efficiency of a breeding program is in part determined by the efficiency of selections and the ability to accurately phenotype the trait of interest. The phenotypic evaluation method needs to be rapid and accurate enough to test the number of samples that is usually generated in a breeding program. Using the phenotyping method described above (5.3.5 Anthocyanin phenotypic evaluation methods), the coffee grinding step and weighing step take about three minutes per sample, which tends to be the bottleneck in the extraction procedure. Extracting and centrifugation takes 1.5 hours to complete but can be done in large batches. Filling two 96-well plates (pH 1.0 and pH 4.5) takes about half an hour to complete for a skilled laborer. A photograph of an example 96-well plate ready for analysis is in Figure 5.2. Reading the plate is rapid, only taking about one minute per plate, so the time reading samples is negligible. From

154

grinding to analysis, it takes about 5 min per sample. This equates to 10 workdays for 960 samples. However, most of these processes can be run in parallel, so the actual time decreases with more laborers. Once selections are made, pigment composition can be determined using the same extract filtered through a 0.45 µm filter. An attractive alternative to the high-throughput absorbance method would be to develop a near-infrared spectroscopy (NIR) method. These methods are non-destructive and should be able to accurately estimate relative performance among trials. NIR is used to measure grain-quality traits like protein, oil, and density. Future work will be aimed at creating a rapid and non-destructive method of anthocyanin yield evaluation.

5.4.4 Unique characteristics of the Inca Purple × LH123 population

In summer 2015, amidst the shade house experiment, a purple corn variety purchased at the local grocery store was grown to see what characteristics the plant would have. The variety, Inca Purple, had photoperiodicity issues as other tropical germplasm, but was able to be crossed to LH123 in that summer. The hybrid resulted in an impressive 628 mg/kg anthocyanin content. This population was backcrossed once and made into a large population for further advancement. A few members of this population are pictured in Figure 5.3. Additionally, this line was backcrossed five times and selected for high anthocyanin content to serve as another elite purple donor inbred line. In addition to having very high content, the composition of this population was very diverse (Figure 5.4). In 2017, a large population of BC_1S_1 and BC_1S_2 lines were grown for analysis. After selection against red or yellow ears that were not pigmented by anthocyanins, the anthocyanin content ranged from 51 to 1281 mg/kg with an average of 504 mg/kg (n=228). This indicates these lines had excellent potential to enhance pigment levels. A total of 29 individuals were selected to advance in the breeding program based on pigment and grain yield and were run on the HPLC. Condensed forms content ranged from 0 to 19.6% and peonidin content from 3.6 to 40.1% in the selected individuals. The Inca Purple × LH123 population is seemingly segregating for the ability to make flavonol-anthocyanin condensed forms and peonidin forms of anthocyanins. The genetic basis for both of these traits is currently unknown. This population could be the key to understanding how these anthocyanins are activated in maize. In addition to segregating for anthocyanin diversity, the population was segregating for two deleterious traits that could be interesting to study further. The Inca Purple parent conferred a recessive lethal white seedling phenotype. Also, an unusual "twisted whorl" and subsequent

155

"crinkly leaf" phenotype appeared in many lines (Figure 5.4). Leaves with this phenotype tend to stay wrapped up in the whorl and are thinner than a normal leaf. When they emerge, they are light green and crinkly from being bound in the whorl. Twisted Whorl Syndrome is common in maize during a period of rapid growth (Elmore, 2008); however, it seems that in this population, it is genotype dependent. There may be a factor in Inca Purple that confers susceptibility to this usually environmentally-controlled syndrome.

5.4.5 Marker assisted selection (MAS) for purple corn

MAS is an effective tool for accelerating breeding progress. However, it must be taken into account that DNA extraction and marker analysis require skilled labor, equipment, and reagents to perform (Arbelaez et al., 2019). MAS can be more efficient in certain scenarios, but a breeder would have to perform a cost analysis to compare the two methods. Given that purple corn requires *R1/B1* and *Pl1* together to activate pigment synthesis, backcross progeny are almost always 25% purple in temperate inbred crosses. Selfing a fully heterozygous purple corn plant would result in 56.25% purple plants. Overall, it would take five rounds of purifying selection for greater than 95% of a randomly mated population to become purely purple corn. Five rounds of purifying selection increases the time it takes to develop a variety, especially if backcrossing is utilized before selfing.

To validate the genetic markers in this study, a small, segregating population of 48 Inca Purple × LH123 advanced backcross individuals were assayed for *umc2528* (*R1*) and *umc1014* (*Pl1*) genetic markers. It was found that all purple corn had the Inca Purple alleles of both *R1* and *Pl1* with the exception of one yellow corn sample that had the Inca Purple allele of *R1* (Figure 5.5). The SSR marker *umc2528* is not within the *R1* gene, but is highly linked (Lago et al., 2014). This recombination frequency (approximately 2%) is also consistent with the genetic distance of *Leaf color1*, which is a tandemly duplicated allele of *R1* that can confer purple pericarp (Ludwig & Wessler, 1990). It is known that Apache Red Cob is pigmented by the *R1* locus, so it is reasonable to hypothesize that Inca Purple may be as well (Chatham & Juvik, 2021).

5.4.6 Efficient selection of pelargonidin-rich hybrids

Described above were several methods for producing pelargonidin-rich hybrids. Each method had their own technical limitations. The absorbance ratio method described can simultaneously differentiate pigment concentration among samples. However, the limitation of

156

this method was grinding and weighing corn kernels to extract the pigment. To improve throughput, whole kernels could be extracted with a highly acidic buffer (less than pH 2.0) in 1.5 mL tubes and the extracts measured in 96-well plates. This method cannot differentiate pigment concentration accurately, however, as kernels are heterogeneous in weight and shape. The limitation of marker-assisted selection of *Pr1* is the specialized reagents and equipment needed for DNA extraction and PCR. Breeders without access to MAS technology might find this method more costly than the absorbance ratio method. The major advantage of MAS is if field or greenhouse space is a limitation, then the *pr1* seedlings could be transplanted after MAS. Based on previous results, it is suggested that breeders use backcrossing to remove deleterious genetic load from the purple corn landraces. However, as the number of backcrosses increases the frequency of *pr1* decreases. Specifically, the frequency of *pr1* in the remaining purple corn lines will be 0.5^{n+1} where *n* is the number of backcrosses. A third-generation backcross, for example, would only have 6.25% *pr1* plants. Of those plants, only 25% will be purple, given that both *B1/R1* and *Pl1* need to be present, for a total of 1.56% purple plants with the *pr1* allele. Furthermore, at the BC_3 stage, the *pr1* allele is not visible because the dominant form is still present. It will take two additional generations of selfing to observe the *pr1* phenotype. It may be advantageous for breeders to use MAS to save the space and resources associated with pollinating.

5.4.7 Purple corn hybrid grain yield results

The final product that will be utilized by the natural colorant industry is the purple corn hybrid. While the natural colorant industry is not strictly interested in grain yield, grain yield needs to be high enough that a farmer would want to grow the variety without perceiving a major risk in investment. The incentive for the farmer is a premium price for the acreage produced compared to normal commercial yellow dent corn. This specialty corn system is in place for organic corn, white corn, blue corn, and popcorn, to name a few. A study of popular blue corn landraces from the Southwest USA found the maximum yield was 33 bu/acre among varieties tested (Nankar et al., 2016). Popcorn yields approximately 75 bu/acre according to USA surveys (NASS, 2021). For comparison, the five-year average (2016–2020) yield for commercial dent corn in the Midwest USA is 164 bu/acre. In Illinois, where the current breeding program is, the five-year average for dent corn yield is 188 bu/acre. Purple corn is at a disadvantage like other specialty corn varieties because it has not been genetically engineered with resistance to major

157

pests and it has not been extensively improved by breeding. Purple corn is susceptible to most major agricultural pests like European corn borer (*Ostrinia nubilalis*) and corn rootworms (*Diabrotica spp.*) to name a few. Timely fungicide and insecticide applications are key to a successful harvest and to prevent lodging by the end of the growing season.

A total of 139 purple corn hybrids were considered over multiple years and multiple environment testing to make decisions in the breeding program. Photographs of a few of the high-performing varieties are in Figure 5.3. The full list of hybrids with their pedigrees is in Table 5.1. Certain varieties would be phased out due to inferior grain yield or limited seed and many new varieties were brought in each year as more inbreds were developed. In this way, the field design was unbalanced. Many varieties were grown in every environment as "checks". In this way, the field design was most similar to an augmented p-rep design (Williams et al., 2011). The check varieties were useful for calculating variance estimates for environments and replications. A total of 26 check varieties had grain yield data in at least one environment in 2018 to 2020 and 25 check varieties had pigment data in all the environments.

A major goal was to achieve an acceptable grain yield so as to be an attractive investment for farmers. Two datasets were considered to calculate yield. One considered full data from all environments measured. The second dataset only considered the UIUC location, which generally performed better. The non-UIUC locations had issues with stalk and root lodging due to inclement weather and pest pressure. Variance components were estimated for grain yield for both of these datasets so broad-sense heritability (H^2) could be calculated. For the full dataset, grain yield was slightly heritable, with genotype × environment interactions and residual variance explaining most of the variability (Table 5.2). The high non-genetic variance was due to lodging and hand-harvesting of some locations making yield estimates inaccurate. When only considering the UIUC dataset, grain yield was highly heritable with genotype and residual variance explaining most of the variability (Table 5.2). Predicted yields were calculated using UIUC data. Grain yield exceeded expectations in this project. The top-five cultivars had predicted yields of 192 to 219 bu/acre, which exceeds the Illinois average yield for corn production. The full predicted yields of purple corn hybrids developed in this study are in Appendix F. The limitation of this study was that the purple corn hybrids were only tested in small plots. Larger-scale testing will allow for a more accurate estimation of true yield. Overall, yield results indicate that many purple corn inbreds developed in this study will be suitable for

158

creating purple corn hybrids. Additionally, integrating tropical germplasm into modern cultivars has the potential to provide genetic variation to increase temperate maize yields for conventional corn breeders. Large-scale tests are underway to test the purple corn hybrids in a more realistic setting.

5.4.8 Purple corn hybrid anthocyanin yield results

The potential for anthocyanin production in the grain is dependent mostly on genetic factors according to most studies (Chapter 4; Nankar *et al.*, 2016; Paulsmeyer *et al.*, 2017). Anthocyanin concentration per weight of corn was calculated in multiple ways depending on machinery availability and time (See 5.3.5 Anthocyanin phenotypic evaluation methods). Therefore, some variability associated with year may be confounded with the difference in anthocyanin measurements. Despite this, anthocyanin concentration data was consistent among all environments tested. The same two dataset used for grain yield analysis were used for calculating variance components and H^2 of various anthocyanin traits in the purple corn hybrids. Results of the variance analysis are in Table 5.3 for the full dataset and in Table 5.4 for the UIUC dataset. Predicted anthocyanin yields for each purple corn hybrid tested in this study are in Appendix F. Anthocyanin concentration was not a normally distributed trait in this study, and therefore variance components of the natural log-transformation of anthocyanin content were also calculated. This study agrees with previous studies that anthocyanin yield is highly heritable trait with most variation coming from genetics regardless of the dataset or anthocyanin yield calculation. Environment and genotype × environment interactions were also a large source of variation in anthocyanin content. Variability among days of extractions (batch effect) was significant, but not a major source of variation overall. Fortunately, anthocyanin concentration was not correlated with grain yield, so there will not be a trade-off for breeders simultaneously increasing these traits. The highest H^2 was with log-transformed anthocyanin concentration from the UIUC location (Table 5.4). In this dataset, the predicted anthocyanin yield was 335 to 366 mg/kg for the top five performing cultivars. The highest-performing cultivars were typically varieties that both parents were purple inbred lines. The exception was with one inbred line developed from a cross between Mo17 and Maize Morado. This inbred produced consistently high predicted anthocyanin yield even when crossed to B73 in hybrid B4251CU. However, the inbred itself was poor performing and rarely produced enough grain for increase. Future work

159

will be aimed at increasing the grain yield of this inbred with seemingly high combining ability for anthocyanin content.

Other anthocyanin traits of interest, proportion of acylation and proportion of flavonol-anthocyanin condensed forms, were also measured because of their association with extract stability in food and beverage matrices. Anthocyanin stability is an important trait for the food industry since consumer perception of natural colorants will in part dictate the feasibility of their use. Composition results are also included in Table 5.3 for the full dataset and Table 5.4 for the UIUC-only dataset. The distribution of proportion of condensed forms was skewed and 92 purple corn hybrids could produce measurable amounts of condensed forms. One variety was capable of producing 23.3% condensed forms of total anthocyanins. The H^2 of natural log-transformed condensed form proportion was moderate for both datasets. Residual variances were greater than genotypic variances. Environment and genotype × environment interactions were just as important as genotypic variation. Proportion of acylation was similar to that of condensed forms, except that it was normally distributed and did not need transformation. The H^2 was moderate and controlled mostly by environment and residual variances. The proportion of cyanidin derivatives is another trait investigated using the chromatography data calculated. This trait was highly heritable in the full dataset with an H^2 of 92.6 (Table 5.3). For both datasets, residual variances and genotypic variances were the most important. This trait is largely controlled by *pr1*, so it is expected to be highly heritable.

Grain yield and anthocyanin yield up until this point have been treated separately. However, purple corn breeders need to consider the total amount of product produced per acre. That is, the yield of pigment per acre, not just anthocyanin concentration on a weight basis. To calculate this, anthocyanin concentrations were converted to pounds anthocyanins per pound of corn and multiplied by the total pounds per acre expected. Anthocyanin yield for each hybrid tested is included in Appendix F. Anthocyanin yield per acre was moderately correlated with grain yield and highly correlated with anthocyanin concentration on a weight basis (Table 5.4). The H^2 for these traits was much higher than anthocyanin concentration alone even when considering the full yield dataset (Table 5.3). These findings indicate that selecting for either grain yield or anthocyanin concentration would both have a positive effect on anthocyanin yield and that a low grain-yielding variety can be compensated with high anthocyanin yield, and *vice versa*. Breeders should focus on increasing anthocyanin concentration in the grain for the most

160

efficient selection as there is much more genetic variability in this trait and it was more highly correlated with overall anthocyanin yield per acre.

5.5 Conclusions

Integrating tropical purple corn germplasm into temperate yellow dent varieties has been challenging but the gains demonstrated in this study show that these varieties have a lot of potential. The natural colorant industry will benefit greatly from genetically improved purple corn varieties. Given the grain yield results presented here, integrating tropical germplasm may also be beneficial to conventional corn breeders trying to increase genetic variation for grain yield. Methods for breeding purple corn were discussed in this paper and it is suggested that an advanced backcross scheme be used to increase grain yield while also maintaining high anthocyanin pigmentation. Shade houses may be necessary for breeders utilizing tropical purple corn landraces. However, three near-isogenic purple corn lines in a B73, Mo17, and LH123 background were developed in this study that can serve as purple donors for future breeding lines. The yield trial indicated several inbred lines that will be useful for increasing anthocyanin yield per acre. Future work will be aimed at ever-increasing the pigment and grain yield of purple corn hybrids to create a more economical source of natural colorants for the natural colorant industry.

5.6 Acknowledgements

161

5.7 Figures and Tables

Figure 5.1 Tropical corn photoperiodicity overcome by shade houses

A) Tropical purple corn landraces next to a 2.43 m ladder for scale. B) Shade houses used to induce flowering for tropical purple corn landraces. C) A tropical landrace stimulated to flower after six weeks in a shade house.
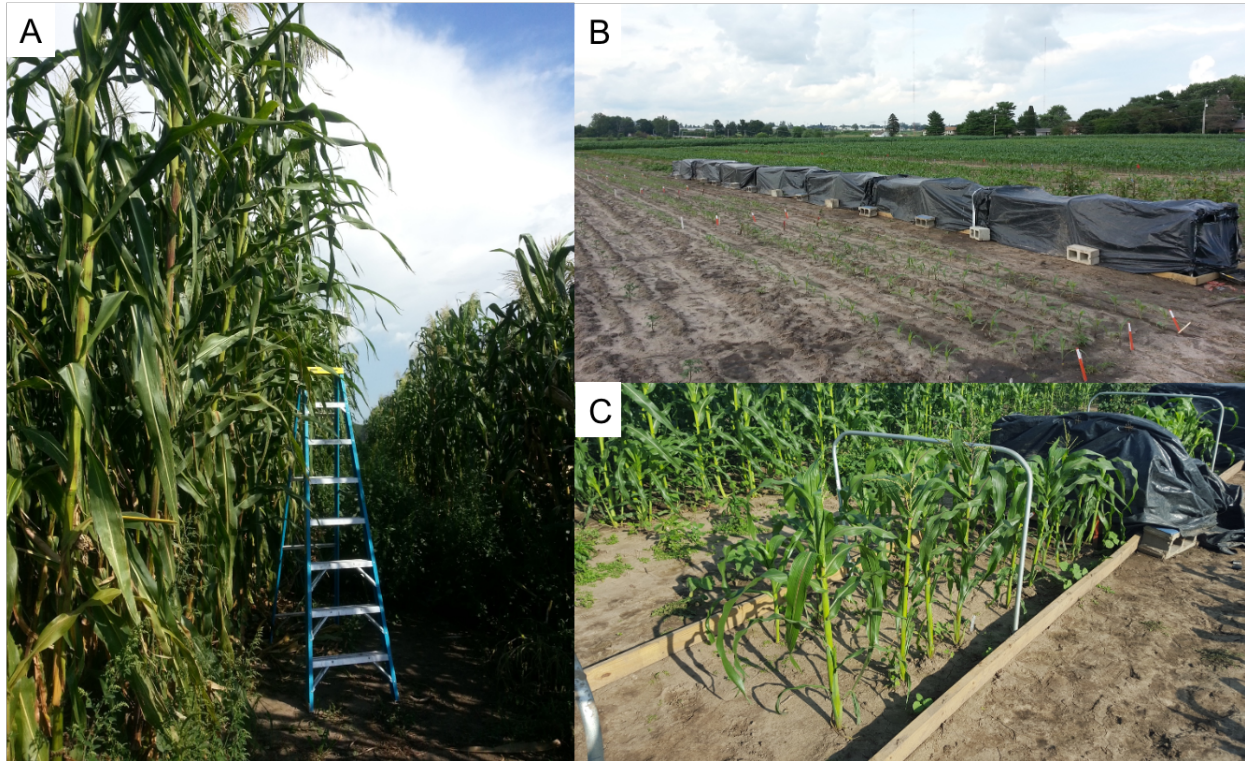
Figure 5.2 High throughput pigment evaluation screen

Top) Anthocyanin extracts in 10 mM potassium chloride (pH 1.0) buffer. Bottom) The same extracts in 0.4 M sodium acetate (pH 4.5).
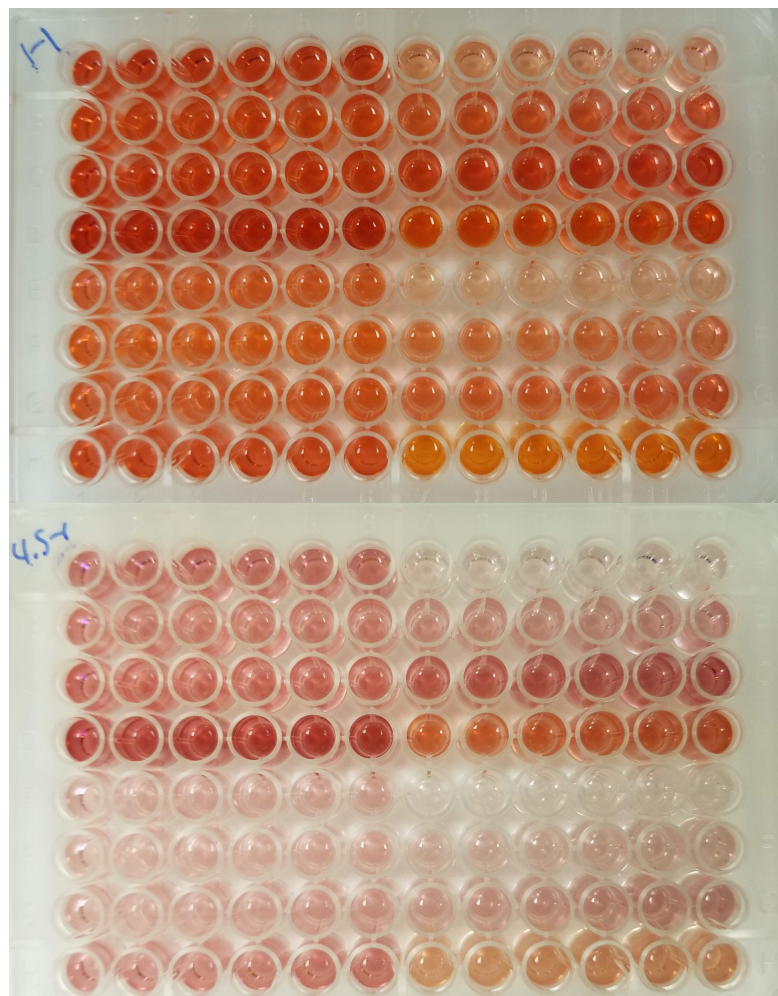
Figure 5.3 Purple corn phenotypes

A) B73 advanced backcross inbred plant demonstrating purple auricles. B) Purple corn inbred with dark purple internodes. C) High-grain and high-anthocyanin yielding purple corn variety. D) High grain-yielding purple corn hybrid with reduced pigmentation. E) B73 advanced backcross inbred. F) Mo17 advanced backcross inbred. Photo Credit E & F UI Public Affairs: L. Brian Stauffer
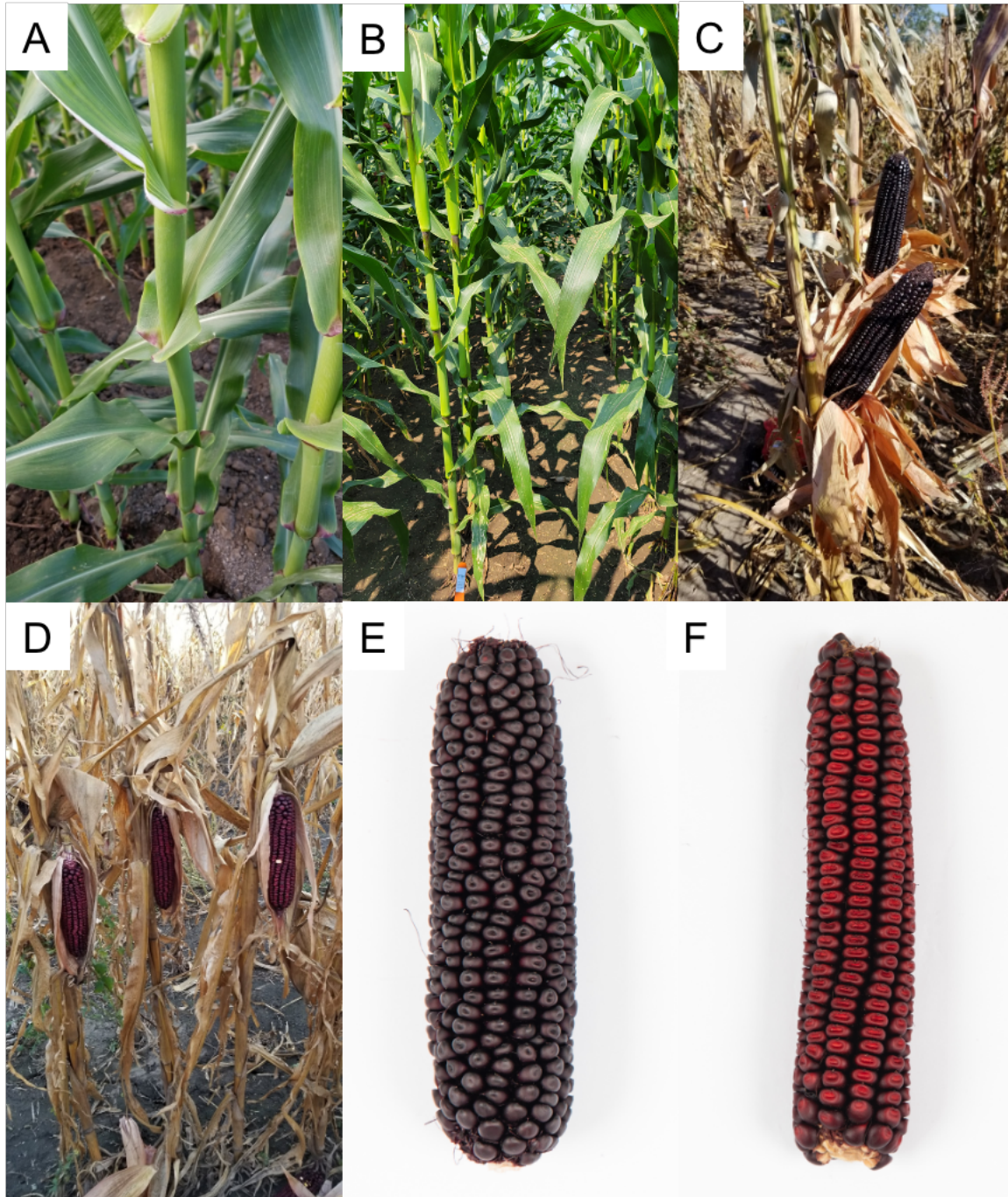
Figure 5.4 Unique characteristics of the Inca Purple × LH123 population

A) HPLC chromatogram of a normal purple corn variety. B) HPLC chromatogram of a high condensed form, low peonidin variety. C) HPLC chromatogram of a high peonidin variety. D) HPLC chromatogram of a normal purple corn variety with condensed forms. E) Member of the Inca Purple × LH123 population with purple auricles and a weak crinkly leaf phenotype. F) Member of the Inca Purple × LH123 population with an unlinked vegetative plant pigmentation phenotype (left) next to a yellow kernel plant (right). G) Crinkly leaf phenotype in the Inca Purple × LH123 population.
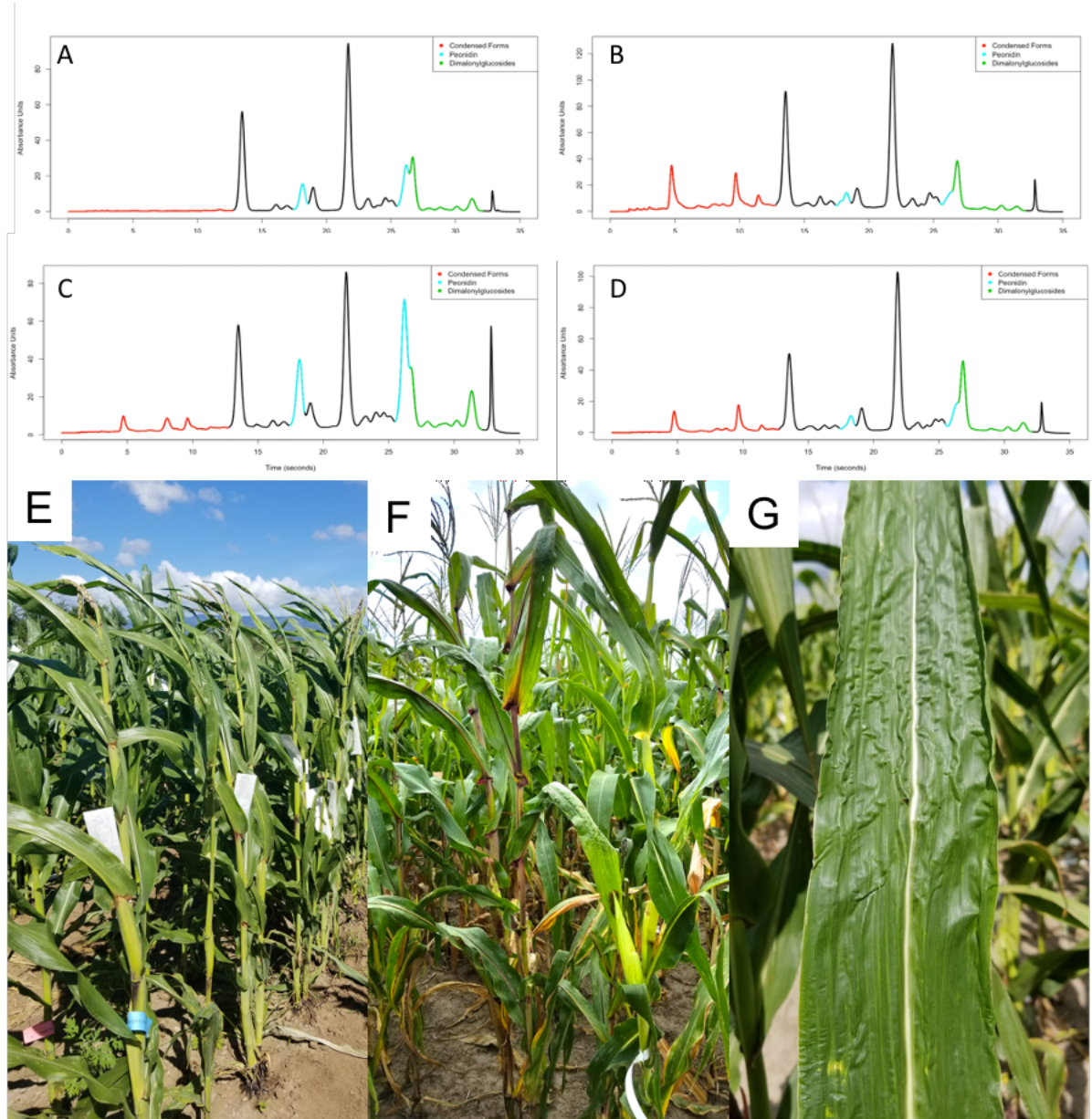
Figure 5.5 Purple corn linkage in the Inca Purple × LH123 population

Segregation of *umc2528 R1* (Top row) and *umc1014* (*Pl1*) SSR marker alleles in the Inca Purple × LH123 population. The DNA ladder is a 100 bp ladder (New England Biolabs.). R= red, Y = yellow, and P = Purple kernels. Lane 21 is a sample that has broken linkage with the markers.

166
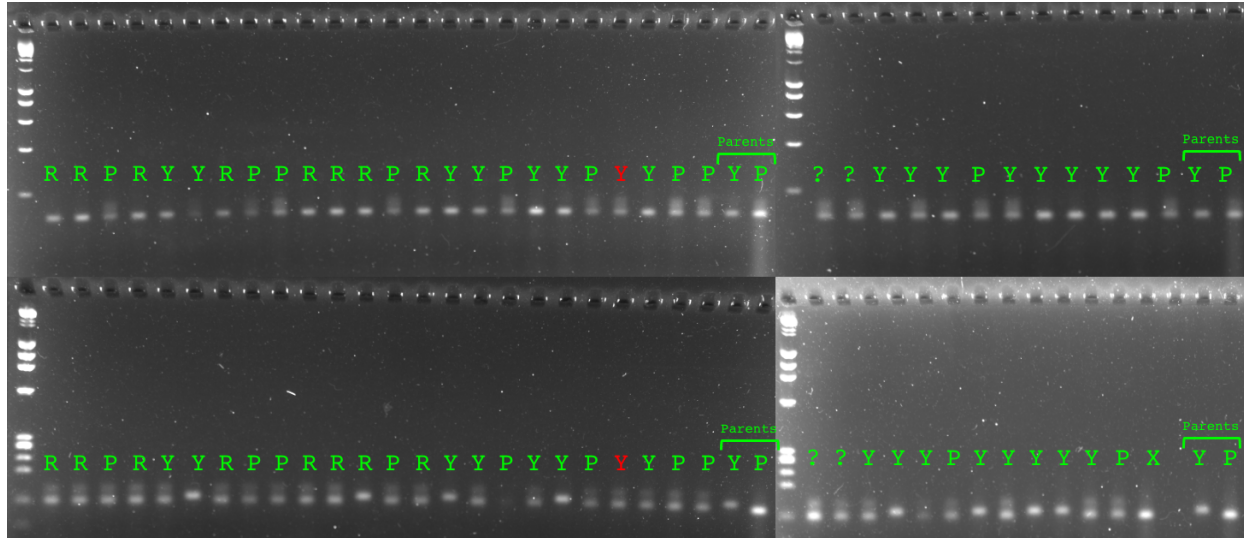
Figure 5.6 Histogram of the A510/A520 Ratio

Distribution of the A510/A520 ratio in a population segregating for *Pr1* alleles. A cutoff of 1.09 indicates recessive alleles where pelargonidin forms are more common.

Histogram of A510/A520
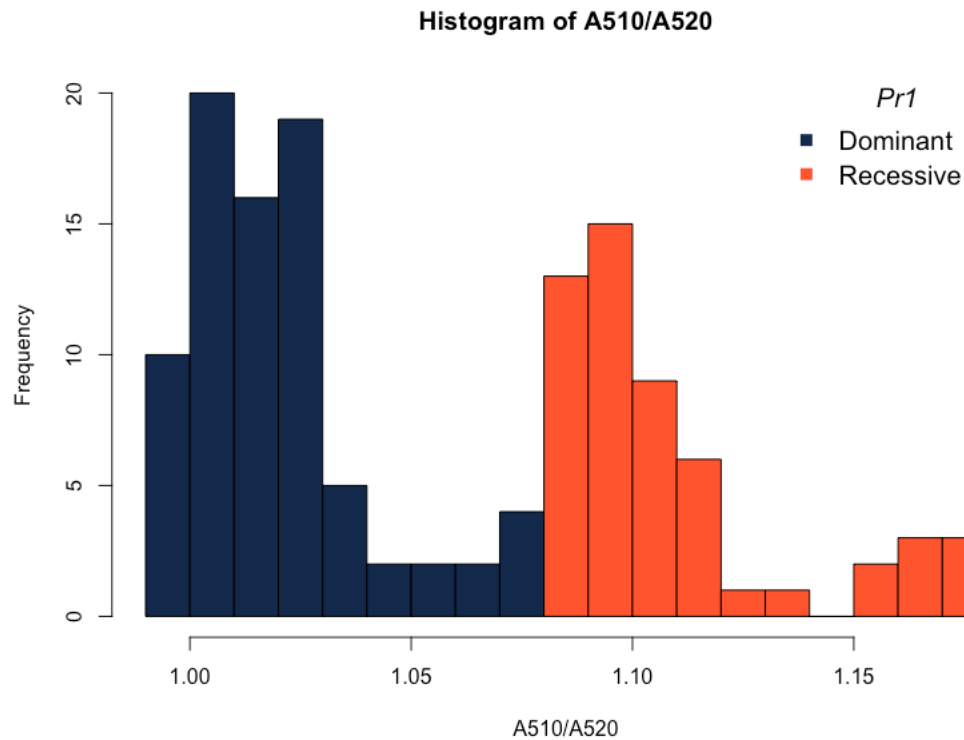
Table 5.1 Pedigrees of purple corn hybrids tested in this study

The hybrid name is the unique code for the genotype. "Females" and "Males" indicate the direction the hybrid was crossed with the female line producing the seed. The first and second parents indicate the varieties used to make the female or male inbreds. An absent second parent indicates the yellow inbred was used as a testcross for the purple corn hybrid. "BC" in the second parent indicates that the inbred was developed from a backcross and the number of times it was backcrossed. Arequipa accessions are from the North Central Regional Plant Introduction Station (Ames, IA, USA), Maize Morado is from Angelina's Gourmet (Swanson, CT, USA), Apache Red Cob is from Siskiyou Seeds (Eugene, OR, USA), Inca Purple is from Amazonas Imports (Irwindale, CA, USA), and 506B and X131I are from the Maize Genetics Cooperation Stock Center (Urbana, IL, USA).

Hybrid Name	Female 1	Female 2	Male 1	Male 2
B0011CU	B73		Arequipa 35	X131I
B0161CU	B73		Arequipa 35	X131I
B2192CU	B73		Inca Purple	LH123 BC1
B2193CU	B73		Inca Purple	LH123 BC1
B2223CU	B73		Inca Purple	LH123 BC1
B2232CU	B73		Inca Purple	LH123 BC1
B2243CU	B73		Inca Purple	LH123 BC1
B2293CU	B73		Inca Purple	LH123 BC1
B2312CU	B73		Inca Purple	LH123 BC1
B2384CU	B73		Inca Purple	LH123 BC1
B2405CU	B73		Inca Purple	LH123 BC1
B2433CU	B73		Inca Purple	LH123 BC1
B2465CU	B73		Inca Purple	LH123 BC1
B2492CU	B73		Inca Purple	LH123 BC1
B2553CU	B73		Inca Purple	LH123 BC1
B2581CU	B73		Inca Purple	LH123 BC1
B2601CU	B73		Inca Purple	LH123 BC1
B2612CU	B73		Inca Purple	LH123 BC1
B2664CU	B73		Inca Purple	LH123 BC1
B2691CU	B73		Inca Purple	LH123 BC1
B2702CU	B73		Inca Purple	LH123 BC1
B2725CU	B73		Inca Purple	LH123 BC1
B2751CU	B73		Inca Purple	LH123 BC1
B2754CU	B73		Inca Purple	LH123 BC1
B2872CU	B73		Inca Purple	LH123 BC1
B3251CU	B73		Arequipa 115	Mo17 BC1
B3253CU	B73		Arequipa 115	Mo17 BC1

Table 5.1 (Cont.)

Hybrid Name	Female 1	Female 2	Male 1	Male 2
B3652CU	B73		Inca Purple	Mo17 BC1
B3701CU	B73		Inca Purple	Mo17 BC1
B3731CU	B73		Inca Purple	Mo17 BC1
B3761CU	B73		Inca Purple	Mo17 BC1
B3792CU	B73		Inca Purple	Mo17 BC1
B3804CU	B73		Inca Purple	Mo17 BC1
B3851CU	B73		Maize Morado	Mo17 BC1
B3884CU	B73		Maize Morado	Mo17 BC1
B3901CU	B73		Maize Morado	Mo17 BC1
B3931CU	B73		Maize Morado	Mo17 BC1
B3952CU	B73		Maize Morado	Mo17 BC1
B4024CU	B73		Maize Morado	Mo17 BC1
B4114CU	B73		Maize Morado	Mo17 BC1
B4221CU	B73		Maize Morado	Mo17 BC1
B4251CU	B73		Maize Morado	Mo17 BC1
B4361CU	B73		Maize Morado	PH207 BC1
B4423CU	B73		Maize Morado	PH207 BC1
B4451CU	B73		Maize Morado	PH207 BC1
B4481CU	B73		Maize Morado	PHG84 BC1
B4482CU	B73		Maize Morado	PHG84 BC1
B4542CU	B73		Maize Morado	PHG84 BC1
B4545CU	B73		Maize Morado	PHG84 BC1
B4632CU	B73		Maize Morado	PHG84 BC1
B4634CU	B73		Maize Morado	PHG84 BC1
B4666CU	B73		Maize Morado	PHG84 BC1
B4691CU	B73		Maize Morado	PHG84 BC1
M0371CU	Mo17		Arequipa 35	B73 BC1
M0422CU	Mo17		Arequipa 35	B73 BC1
M0625CU	Mo17		Arequipa 35	B73 BC1
M0781CU	Mo17		Arequipa 35	B73 BC1
M1655CU	Mo17		Arequipa 35	B73 BC3
M1672CU	Mo17		Arequipa 35	B73 BC3
M1791CU	Mo17		Arequipa 35	B73 BC3
M1884CU	Mo17		Arequipa 35	B73 BC5
M1896CU	Mo17		Arequipa 35	B73 BC5
M1922CU	Mo17		Inca Purple	B73 BC1

169

Table 5.1 (Cont.)

Hybrid Name	Female 1	Female 2	Male 1	Male 2
M3052CU	Mo17		Arequipa 204	LH84 BC2
M3104CU	Mo17		Arequipa 204	LH84 BC2
P0321CU	Maize Morado	PHG84 BC1	Arequipa 35	B73 BC1
P0601CU	Arequipa 115	LH1 BC3	Arequipa 35	Mo17 BC1
P0881CU	Arequipa 35	B73 BC1	Inca Purple	LH123 BC1
P1141CU	Inca Purple	LH123 BC1	Arequipa 35	B73 BC1
P1231CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
P1401CU	Arequipa 35	B73 BC1	Inca Purple	LH123 BC1
P1884CU	Arequipa 35	B73 BC1	Arequipa 35	Mo17 BC1
P2711CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
P3411CU	Arequipa 35	B73 BC1	Maize Morado	PHG84 BC1
P3437CU	Arequipa 204	LH84 BC2	Maize Morado	PH207 BC1
P3591CU	Arequipa 35	B73 BC1	Maize Morado	PHG84 BC1
P3649CU	Apache Red Cob	LH123 BC1	Apache Red Cob	
P4361CU	Arequipa 204	LH84 BC2	Maize Morado	PH207 BC1
P4451CU	Arequipa 204	LH84 BC2	Maize Morado	PH207 BC1
P5349CU	Apache Red Cob	LH123 BC1	Apache Red Cob	
P5649CU	Apache Red Cob	LH123 BC1	Apache Red Cob	
P5746CU	Apache Red Cob	LH123 BC1	Apache Red Cob	
PB021CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB025CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB029CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB033CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB037CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB041CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB045CU	Arequipa 35	B73 BC5	Apache Red Cob	LH123 BC1
PB073CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB077CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB081CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB085CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB089CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB093CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB097CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB101CU	Arequipa 35	B73 BC5	Arequipa 115	Mo17 BC1
PB105CU	Arequipa 35	B73 BC5	Arequipa 115	Mo17 BC1
PB109CU	Arequipa 35	B73 BC5	Arequipa 115	Mo17 BC1

170

Table 5.1 (Cont.)

Hybrid Name	Female 1	Female 2	Male 1	Male 2
PB113CU	Arequipa 35	B73 BC5	Arequipa 115	Mo17 BC1
PB117CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB129CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB135CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB141CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB151CU	Arequipa 35	B73 BC5	Inca Purple	LH123 BC1
PB260CU	Apache Red Cob	Mo17 BC1	Arequipa 35	B73 BC5
PB262CU	Apache Red Cob	Mo17 BC1	Arequipa 35	B73 BC5
PB264CU	Apache Red Cob	Mo17 BC1	Arequipa 35	B73 BC5
PB266CU	Apache Red Cob	Mo17 BC1	Arequipa 35	B73 BC5
PB268CU	Apache Red Cob	Mo17 BC1	Arequipa 35	B73 BC5
PB270CU	Arequipa 35	B73 BC5	Apache Red Cob	Mo17 BC1
PE003CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE004CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE005CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE008CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE009CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE010CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE013CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE014CU	Arequipa 35	Mo17	Arequipa 35	B73 BC5
PE123CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
PE129CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
PE141CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
PE151CU	Maize Morado	PHG84 BC1	Inca Purple	LH123 BC1
PF262CU	506B		Apache Red Cob	LH123 BC1
PF304CU	506B		Apache Red Cob	LH123 BC1
PF313CU	506B		Apache Red Cob	LH123 BC1
PH071CU	Arequipa 35	B73 BC5	Arequipa 204	Mo17
PH154CU	Arequipa 35	B73 BC5	Arequipa 204	Mo17
PH155CU	Arequipa 35	B73 BC5	Arequipa 204	Mo17
PH761CU	Arequipa 35	B73 BC5	Arequipa 204	Mo17
PH7671CU	Arequipa 204	Mo17	Arequipa 204	Mo17
PJ321CU	Arequipa 204	506B	Arequipa 204	506B
PM013CU	Arequipa 35	Mo17 BC5	Arequipa 35	Mo17 BC1
PM017CU	Arequipa 35	Mo17 BC5	Arequipa 35	Mo17 BC1

171

Table 5.2 Analysis of variance and heritability calculations for purple corn hybrid grain yield

	All Environments				UIUC Only			
Term	**σ^2**	**p**		**Effect**	**σ^2**	**p**	
Intercept		4.37E-04	***	**Intercept**		0.0011	**
Year × Location	305.1	0.0103	*	**Year**	156.8	0.0230	*
Year × Location × Rep	0	1		**Year × Rep**	1.4	0.8891	
UIUC	266.5	0.0230	*	**UIUC**	787.1	6.45E-08	***
Year × Location × UIUC	629	5.39E-06	***	**Year × UIUC**	111.9	0.1756	
Residual	994.2			**Residual**	937.1		
Heritability	0.4710			**Heritability**	0.8027		

Asterisks denote significance where (***) < 0.001, (**) < 0.01, and (*) < 0.05.

172

Table 5.3 Analysis of variance and heritability calculations for purple corn hybrid traits in the full dataset

All Environments	Anthocyanin Yield (lbs./bu)			Log-transformed Anthocyanin Yield (lbs./bu)			Anthocyanin Concentration (mg/kg)			Log-transformed Anthocyanin Concentration (mg/kg)		
Term	σ^2	p		σ^2	p		σ^2	p		σ^2	p	
Intercept		0.0561	.		0.18			0.0098	**		6.15E-05	***
Year × Location	0.9889	0.001223	**	0.2156	0.0025	**	7268.7	0.0010	**	0.1284	0.0032	**
Year × Location × Rep	0	1		0	1		37.4	0.8026		0.0007	0.8858	
Year × Location × Rep × Batch	0.0358	0.12475		0.0143	0.0023	**	424.9	0.0072	**	0.0156	2.21E-04	***
UIUC	1.2421	8.67E-08	***	0.1965	1.16E-11	***	8015.0	7.62E-13	***	0.1194	2.20E-16	***
Year × Location × UIUC	0.1964	0.040273	*	0.0348	0.04152	*	4298.7	6.70E-14	***	0.0400	2.62E-04	***
Residual	0.8534			0.1139			3924.1			0.1053		
Heritability	0.8567			0.8653			0.8366			0.8375		

All Environments	Log-transformed Condensed Forms (%)			Acylation (%)			Cyanidin (%)		
Term	σ^2	p		σ^2	p		σ^2	p	
Intercept		0.925			0.00203	**		4.71E-12	***
Year × Location	1.163	0.1462		23.5206	0.016679	*	0.8725	0.6807	
Year × Location × Rep	0	1		0.3609	0.758396		8.51E-09	0.9999	
Year × Location × Rep × Batch	1.076	2.22E-12	***	1.5393	0.092685		9.55	2.74E-06	***
UIUC	1.693	4.42E-06	***	10.5372	0.001002	**	89.8	2.20E-16	***
Year × Location × UIUC	2.13	9.81E-14	***	10.6487	0.002081	**	0	1	
Residual	2.006			33.9386			42.9		
Heritability	0.6185			0.5337			0.9263		

Asterisks denote significance where (***) < 0.001, (**) < 0.01, and (*) < 0.05.

173

Table 5.4 Analysis of variance and heritability calculations for purple corn hybrid traits at UIUC only

UIUC Only	Anthocyanin Yield (lbs./bu)		Log-transformed Anthocyanin Yield (lbs./bu)		Anthocyanin Concentration (mg/kg)		Log-transformed Anthocyanin Concentration (mg/kg)		Log-transformed Condensed Forms (%)		Acylation (%)		Cyanidin (%)	
Term	σ^2	p	σ^2	p	σ^2	p	σ^2	p	σ^2	p	σ^2	p	σ^2	p
Intercept		0.0561 .		0.18		0.0432 *		0.0016 **		0.698		0.00979 **		2.00E-16 ***
Year	0.9889	0.0403 *	0.2156	0.0025 **	9744.9	0.0021 **	0.1646	0.0046 **	1.4845	0.3176	2.7004	0.40970	0	0.9999
Year × Rep	0	8.67E-08 ***	0	1	28.3	0.8395	4.93E-10	1	0.0230	0.9818	0.2096	0.89358	0	1.0000
Year × Rep × Batch	0.0358	0.12475	0.0143	0.0023 **	396.5	0.005295 **	0.0155	2.18E-05 ***	1.5773	2.44E-09 ***	1.4701	0.18353	9.516	0.0032 **
UIUC	1.2421	1	0.1965	1.16E-11 ***	9840.6	3.15E-08 ***	0.1603	1.45E-14 ***	2.1011	0.008854 **	11.08	0.05635 .	36.064	0.0018 **
Year × UIUC	0.1964	0.0012 **	0.0348	0.04152 *	4566.1	2.09E-11 ***	0.0254	0.01235 *	3.2941	8.81E-11 ***	14.7003	0.00417 **	0	0.9998
Residual	0.8534		0.1139		3203.1		0.0763		2.6201		36.153		67.133	
Heritability	0.8567		0.8653		0.8272		0.8833		0.4772		0.4033		0.6824	

Asterisks denote significance where (***) < 0.001, (**) < 0.01, and (*) < 0.05.

Table 5.5 Correlation matrix for purple corn hybrid traits

Trait*	Grain	ACN	Cond	Acyl	C3G	ACN x Grain
Grain	1 ***	0.08	-0.01	0.02	0.13 *	0.49 ***
ACN		1 ***	0.06	0.03	0.24 ***	0.86 ***
Cond			1 ***	-0.28 ***	-0.3 ***	0.09
Acyl				1 ***	-0.07	-0.02
C3G					1 ***	0.24 ***
ACN x Grain						1 ***

Asterisks denote significance where (***) < 0.001, (**) < 0.01, and (*) < 0.05.
*Grain = grain yield, ACN = anthocyanin concentration, Cond = proportion of condensed forms, Acyl = proportion of acylated anthocyanins, C3G = proportion of cyanidin derivatives, ACN x Grain = anthocyanin yield per acre

5.8 References

Andorf, C. M., Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Campbell, D. A., & Sen, T. Z. (2010). The Locus Lookup tool at MaizeGDB: Identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics*, *26*(3), 434–436. https://doi.org/10.1093/bioinformatics/btp556

Arbelaez, J. D., Tandayu, E., Reveche, M. Y., Jarana, A., van Rogen, P., Sandager, L., Stolt, P., Ng, E., Varshney, R. K., Kretzschmar, T., & Cobb, J. (2019). Methodology: Ssb-MASS: a single seed-based sampling strategy for marker-assisted selection in rice. *Plant Methods*, *15*(1), 78. https://doi.org/10.1186/s13007-019-0464-2

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Cannon, E. K. S., Birkett, S. M., Braun, B. L., Kodavali, S., Jennewein, D. M., Yilmaz, A., Antonescu, V., Antonescu, C., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., Campbell, D. A., Andorf, C. M., Andorf, D., Lisch, D., Koch, K. E., McCarty, D. R., Quackenbush, J., Grotewold, E., … Lawrence, C. J. (2011). POPcorn: An online resource providing access to distributed and diverse maize project data. *International Journal of Plant Genomics*, *2011*, e923035. https://doi.org/10.1155/2011/923035

Center for Food Safety and Applied Nutrition. (2020). *Report on the Certification of Color Additives: 4th Quarter, Fiscal Year 2020, July 1-September 30*. Food and Drug Administration. https://www.fda.gov/industry/color-certification-reports/report-certification-color-additives-4th-quarter-fiscal-year-2020-july-1-september-30

Chatham, L. A., & Juvik, J. A. (2021). Linking anthocyanin diversity, hue, and genetics in purple corn. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa062

Chatham, L. A., West, L., Berhow, M. A., Vermillion, K. E., & Juvik, J. A. (2018). Unique Flavanol-anthocyanin condensed forms in apache red purple corn. *Journal of Agricultural and Food Chemistry*, *66*(41), 10844–10854. https://doi.org/10.1021/acs.jafc.8b04723

Chatham, L., Paulsmeyer, M., & Juvik, J. (2019). Prospects for economical natural colorants: Insights from maize. *Theoretical and Applied Genetics*, *132*(11), 2927–2946. https://doi.org/10.1007/s00122-019-03414-0

176

Chen, Y., Blanco, M., Ji, Q., Frei, U. K., & Lübberstedt, T. (2014). Extensive genetic diversity and low linkage disequilibrium within the COMT locus in maize exotic populations. *Plant Science*, *221–222*, 69–80. https://doi.org/10.1016/j.plantsci.2014.02.004

Elmore, R. (2008). *Twisted Whorls*. Iowa State University Extension and Outreach. https://crops.extension.iastate.edu/cropnews/2008/07/twisted-whorls

Hauck, A. L., Johnson, G. R., Mikel, M. A., Mahone, G. S., Morales, A. J., Rocheford, T. R., & Bohn, M. O. (2014). Generation means analysis of elite ex-plant variety protection commercial inbreds: A New public maize genetics resource. *Crop Science*, *54*(1), 174–189. https://doi.org/10.2135/cropsci2013.03.0172

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Khan, M. I., & Giridhar, P. (2015). Plant betalains: Chemistry and biochemistry. *Phytochemistry*, *117*, 267–295. https://doi.org/10.1016/j.phytochem.2015.06.008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lago, C., Cassani, E., Zanzi, C., Landoni, M., Trovato, R., & Pilu, R. (2014). Development and study of a maize cultivar rich in anthocyanins: Coloured polenta, a new functional food. *Plant Breeding*, *133*(2), 210–217. https://doi.org/10.1111/pbr.12153

Lee, J., Durst, R. W., & Wrolstad, R. E. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *Journal of AOAC International*, *88*(5), 1269–1278.

Ludwig, S. R., & Wessler, S. R. (1990). Maize R gene family: Tissue-specific helix-loop-helix proteins. *Cell*, *62*(5), 849–851. https://doi.org/10.1016/0092-8674(90)90259-H

Markham, K. R. (1988). Distribution of flavonoids in the lower plants and its evolutionary significance. In J. B. Harborne (Ed.), *The Flavonoids: Advances in Research since 1980* (pp. 427–468). Springer US. https://doi.org/10.1007/978-1-4899-2913-6_12

Mikel, M. A. (2011). Genetic composition of contemporary u.s. commercial dent corn germplasm. *Crop Science*, *51*(2), 592–599. https://doi.org/10.2135/cropsci2010.06.0332

Nankar, A., Grant, L., Scott, P., & Pratt, R. C. (2016). Agronomic and kernel compositional traits of blue maize landraces from the southwestern united states. *Crop Science*, *56*(5), 2663–2674. https://doi.org/10.2135/cropsci2015.12.0773

NASS. (2021). *USDA/NASS QuickStats Ad-hoc Query Tool*. Quick Stats. https://quickstats.nass.usda.gov/

Paulsmeyer, M., Chatham, L., Becker, T., West, M., West, L., & Juvik, J. (2017). Survey of anthocyanin composition and concentration in diverse maize germplasms. *Journal of Agricultural and Food Chemistry*, *65*(21), 4341–4350. https://doi.org/10.1021/acs.jafc.7b00771

Petroni, K., Pilu, R., & Tonelli, C. (2014). Anthocyanins in corn: A wealth of genes for human health. *Planta*, *240*(5), 901–911. https://doi.org/10.1007/s00425-014-2131-1

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sen, T. Z., Harper, L. C., Schaeffer, M. L., Andorf, C. M., Seigfried, T. E., Campbell, D. A., & Lawrence, C. J. (2010). Choosing a genome browser for a model organism database: Surveying the maize community. *Database*, *2010*, baq007. https://doi.org/10.1093/database/baq007

Somavat, P., Li, Q., de Mejia, E. G., Liu, W., & Singh, V. (2016). Coproduct yield comparisons of purple, blue and yellow dent corn for various milling processes. *Industrial Crops and Products*, *87*, 266–272. https://doi.org/10.1016/j.indcrop.2016.04.062

Williams, E., Piepho, H.-P., & Whitaker, D. (2011). Augmented p-rep designs. *Biometrical Journal*, *53*(1), 19–27. https://doi.org/10.1002/bimj.201000102

CHAPTER 6: PURPLE CORN IN THE FUTURE

6.1 Purple corn as an economical source of natural colorants

Maize has great potential as an economic source of natural colorants. As was described in Chapter 1, the abundance of pigment production and attractive hues makes maize anthocyanins suitable for many natural colorant applications. Maize is unique in that it has the potential to produce water-soluble orange extracts that might be a good replacement for the lipid-soluble carotenoids used as natural colorants. Currently, only maize grain can be used to color foods and beverages, limiting the economic potential for maize anthocyanins (Center for Food Safety and Applied Nutrition, 2017). Husks can produce up to 60 times more anthocyanins than kernels, for example (Table 4.1). The non-vegetable portions of purple corn such as stover and forage can be utilized by pharmaceutical, nutraceutical, cosmetic, and textiles industries since the regulations are less restrictive. Given this, maize anthocyanins can be a more sustainable alternative to synthetic dyes. Especially for the textile industry where synthetic dye in wastewater can pollute watersheds (De Nisi et al., 2021).

Another advantage of using maize as a source of natural colorants is that maize can be produced at a large scale without special farming equipment. A specialty corn supply chain is already in place that can handle the scale of production needed for the natural colorant industry. Given the results in Chapter 5, it will not be difficult to convince farmers to make an investment into purple corn production. Furthermore, maize is not perishable like the other fruit and vegetable sources of natural colorants and can be stored indefinitely until use. In the future, varieties produced as a result of this project will be grown at a larger scale and released to companies attempting to make natural colorants from maize. The hope is that eventually these varieties will be processed through either the dry grind or dry milling process to remove the pericarp while the rest of the kernel can be used for food or fuel. Currently, the waste products of purple corn extractions are being utilized as animal feed and therefore are not being used to their full economic potential. Niche markets could utilize purple corn by-products as "super foods" loaded with antioxidants. Previous studies have shown that purple corn has more antioxidant capacity than any other fruit and vegetable source analyzed (Cevallos-Casals & Cisneros-Zevallos, 2003). Marketing purple corn as a natural and healthy supplement to diets would increase the economic feasibility of purple corn production.

179

6.2 Regulatory genes increasing anthocyanin production in purple corn

The anthocyanin pathway is quite possibly the most well-studied secondary metabolite pathway in plants. Because of the colorful nature of anthocyanins, they easily became the first reporter gene system utilized by geneticists. Beginning with Gregor Mendel and his pea experiments, anthocyanins helped uncover monumental milestones in genetics. Barbara McClintock earned her Nobel Prize using anthocyanins in maize as a proxy for transposon insertion events. Epigenetic phenomena imprinting and paramutation were also characterized using alleles of anthocyanin transcription factors in maize (Brink, 1956; Kermicle, 1970). Mutations enhancing or eliminating anthocyanin production helped decipher the genes within anthocyanin biosynthetic pathway and the regulators affecting it. Despite our vast knowledge of the anthocyanin pathway in maize, there are still gaps in the biosynthetic steps and there is much more to learn about the genes fine-tuning the expression of this pathway.

In Chapter 3, many regulatory genes were upregulated in purple husks. The homology of these transcriptional regulators to other known anthocyanin regulators implicate their role in promoting or possibly inhibiting the anthocyanin pathway in maize. *Wrky33* (Zm00001d024323) is a transcriptional regulator with homology to Arabidopsis *TRANSPARENT TESTA GLABRA2*, which affects late steps in the anthocyanin biosynthetic pathway (Gonzalez et al., 2016). *Nactf25* (Zm00001d023294) and *Nactf44* (Zm00001d028999) have homology to a known transcriptional regulator in peach (*Prunus persica*). Utilizing CRISPR-CAS9-generated knockouts may be the most efficient way to determine the function of these genes. Discovering their modes of action would be possible with a yeast one-hybrid system or systematic evolution of ligands by exponential enrichment (SELEX).

Aside from *A3*, many negative regulators of anthocyanin synthesis are active in maize. Transcripts of *Myb11*, *Myb31*, *Myb42*, *Sro1*, and *In1* were expressed in mutant husks in Chapter 3. *Myb11*, *Myb31*, and *Myb42* regulate lignin content while *Sro1* is a competitive inhibitor of *Pl1* (Agarwal et al., 2016; Qin et al., 2021; Vélez-Bermúdez et al., 2015). *In1* affects aleurone pigmentation and is not known to be associated with husks. This indicates that stacking knockouts of these genes may increase anthocyanin content overall. Over-production of anthocyanin content is good in theory for natural colorants, but possibly detrimental to the plant. Photosynthetic gene transcripts were overall reduced in *a3* maize. Future work should measure photosynthesis in mutant plants to see if the reduction in genes also implies reduced

180

photosynthesis. The mechanism for reduced photosynthetic capacity should also be investigated. It may be that A3 can bind to photosynthesis-related genes. The mechanism for *A3* activation is not currently known either. Understanding how this gene is regulated may be important for understanding how maize induces anthocyanin synthesis in the first place—how maize regulates its regulators, so to speak. Finally, a big question left in Chapter 1 was whether A3 could bind to any *R1* alleles. Theoretically, A3 should bind to Lc1. Work is underway to clone this gene and test with a GST-pulldown assay.

6.3 The biosynthetic pathway in maize is still incomplete

In Chapter 2, a novel anthocyanin was discovered that upset our current understanding of maize acylated anthocyanins. It was surprising to see that the novel pigment was conjugated to an auxin and that the pigment had a detectable odor due to the breakdown products. Future work should be focused at whether *Aat1* is able to add an aromatic acyl group to anthocyanins, or if another gene is involved. Aromatic acyl groups were not tested in Chapter 2, but they might be a possible substrate for the enzyme. An interesting point made in Chapter 2 was the larger binding pocket size due to the Phe-substitution in motif 2 (Tyr-[Phe/Lys]-Gly-Asn-Cys). This larger binding site may be able to accompany phenylacetyl-CoA and other aromatic acyl groups. It would also be interesting to mutate this amino acid to see if dimalonyltransferase activity is reduced.

The anthocyanin methyltransferase in maize has been elusive up until this point. In Chapter 3, the only methyltransferase candidate upregulated was *Omt4*. Upon inspection, it appears there is an error in the gene model and gene models Zm00001d052841 and Zm00001d052842 should be combined. This was based on paired-end reads generated in the RNA-seq dataset (Figure 6.1). Zm00001d052840 is a neighboring gene on the opposite strand. This may be the reason no transcriptomic study has discovered this enzyme to this date. Future work will be aimed at cloning the cDNA of this enzyme and seeing if full-length protein is capable of methyltransferase activity. Knockouts of all the gene transcripts are available through the UniformMu project (Liu et al., 2016). Testcrosses should be able to determine if this gene can make peonidin *in planta*.

Delphinidins are a class of anthocyanidins that have an additional hydroxyl group on 5'-position of the molecule (See Figure 2.1). These pigments shift the color spectrum to a more violet color and are therefore important for some natural colorant applications. Maize abundantly

181

expresses a Flavonoid 3'-Hydroxylase called *Red aleurone1* (*Pr1*) and not a Flavonoid 3'5'-hydroxylase. In Chapter 1, no delphinidins, petunidins, or malvidins were found with an exhaustive LC-MS/MS scan amongst all tissue types. However, a previous study did find that gene model Zm00001d047424 was expressed in their study and they could find detectable amounts of delphinidins and other related anthocyanidins (Liu et al., 2018). Future work should determine the expression pattern of this enzyme and see if it actually has F3'5'H activity. Cytochrome P450 enzymes are hard to purify since they are membrane proteins. However, a transgenic approach can be used. Petunia (*Petunia hybrida*) is a good model for anthocyanin study since the anthocyanins are much simpler than in *Arabidopsis thaliana* and it is easy to transform either stably or transiently. Additionally, protoplasts or cell cultures of maize could be used to study this gene transgenically in maize. Future work will be aimed at expressing Zm00001d047424 in red petunia incapable of producing delphinidins to see if a characteristic purple color will appear in the petals from the presence of delphinidins.

6.4 Multiple aleurone layers for human health

Chapter 4 demonstrated the effectiveness of increasing aleurone yield with the multiple aleurone layer (MAL) trait. This association is important because maize accumulates many phytonutrients in the aleurone. In Chapter 4, the elemental analysis results were inconclusive, but a small increase in iron content was possible in MAL lines (Table 4.5). A problem with the elemental analysis and also the fine-mapping study was there was too much genetic variability. Work is underway to create near-isogenic MAL lines in a B73, Mo17, LH123, and PHG84 background. In this way, the genetic variation in the elemental analysis will be nearly all attributed to the MAL loci or environment. The near-isogenic lines will also be utilized for either fine-mapping, RNA-seq, or both, to determine the causal genes for MAL formation. The fine-mapping experiment is very straight-forward, as the molecular markers have already been developed (Figure 4.5 and Table 4.2). The density of markers could be improved, however. The RNA-seq experiment will be challenging since the numerous layers will inflate the transcript levels of important aleurone-specific genes in MAL samples. Samples will have to be normalized to internal controls so that only MAL-associated genes are differentially expressed. If the micronutrient results are promising, it may be worth it to sequence the genome of a MAL landrace. These landraces have many interesting traits aside from the MAL phenotype. The maize community would benefit from the sequences from exotic races of maize.

182

6.5 The future of purple corn

A lot of work has been done to improve the purple corn varieties that were once exotic tropical landraces to temperate climate adapted inbred lines. A major result of Chapter 5 is the pre-breeding of purple corn lines into favorable genetic backgrounds. The plan is to use these lines in an RNA-seq experiment to determine genes upregulated in purple corn. The genes discovered in this study will assist in our understanding of the regulatory aspects of anthocyanin pigmentation. In addition, these lines could be used for future breeding programs. As discussed briefly, the purple donor lines developed are far from perfect. Breeding is a continuous process, and the next generations will continue to be better. The Inca Purple × LH123 population developed in Chapter 5 would be an excellent choice for genotyping since there were various characteristics that made that population unique (Figure 5.4). Inca Purple × LH123 and many other purple corn inbreds developed did not handle selfing very well. This population and others like it would be an ideal male in a hybrid. Currently, an open-pollinated version of Inca Purple × LH123 is in works and depending on the results, could be its own variety. A few pelargonidin varieties were developed as a result of the breeding program as well, but these varieties also have room for improvement since they were only backcrossed once and still retain unfavorable characteristics from Apache Red Cob. In the future, these lines will be selected with molecular markers during backcrossing to save time pollinating and to reduce field space.

Overall, the work presented in this dissertation demonstrates the efficacy of using purple corn as a source of natural colorants. New anthocyanin regulatory genes and biosynthetic genes have been presented here as new targets for study. Other competitive inhibitors could be new breeding targets. MALs should be integrated into breeding programs looking to increase health benefits of maize. Finally, new purple corn hybrids will continue to be developed to meet the needs for the natural colorant industry. The hope is that this dissertation provides some guidance in the development of new varieties of purple corn with enhanced gain and anthocyanin yield.

6.6 Figure

Figure 6.1 Methyltransferase candidate transcripts in the *A3* bulk-segregant RNA-seq population.

At the top is a histogram of read depth. Boxes below indicate reads. Thin blue lines between reads connects read gaps. Only the first few reads are shown. In the bottom track are the annotated gene transcripts with arrows for strand orientation. Image was generated using Integrative Genomics Viewer Version 2.8.13 (Robinson et al., 2011) and the B73 RefGen_v4 genome (Jiao et al., 2017).
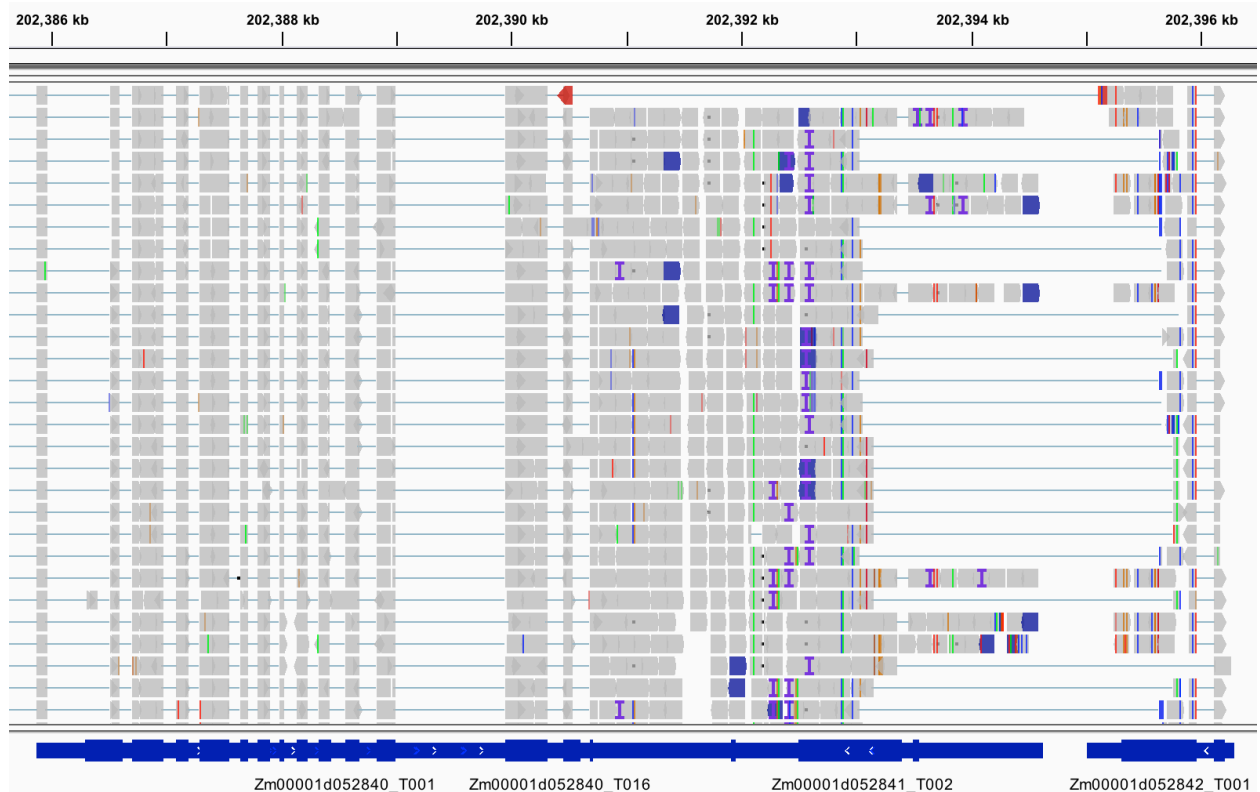
6.7 References

Agarwal, T., Grotewold, E., Doseff, A. I., & Gray, J. (2016). MYB31/MYB42 syntelogs exhibit

　　divergent regulation of phenylpropanoid genes in maize, sorghum and rice. *Scientific*

　　Reports, *6*, 28502. https://doi.org/10.1038/srep28502

Brink, R. A. (1956). A genetic change associated with the r locus in maize which is directed and

　　potentially reversible. *Genetics*, *41*(6), 872–889.

Center for Food Safety and Applied Nutrition. (2017). *21 CFR Part 73: Color additives exempt*

　　from certification. Summary of color additives for use in the united states in foods, drugs,

　　cosmetics, and medical devices. https://www.fda.gov/industry/color-additive-

　　inventories/summary-color-additives-use-united-states-foods-drugs-cosmetics-and-

　　medical-devices

Cevallos-Casals, B. A., & Cisneros-Zevallos, L. (2003). Stoichiometric and kinetic studies of

　　phenolic antioxidants from andean purple corn and red-fleshed sweetpotato. *Journal of*

　　Agricultural and Food Chemistry, *51*(11), 3313–3319. https://doi.org/10.1021/jf034109c

De Nisi, P., Borlini, G., Parizad, P. A., Scarafoni, A., Sandroni, P., Cassani, E., Adani, F., &

　　Pilu, R. (2021). Biorefinery approach applied to the valorization of purple corn cobs. *ACS*

　　Sustainable Chemistry & Engineering, *9*(10), 3781–3791.

　　https://doi.org/10.1021/acssuschemeng.0c08717

Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., Moore, J., Montes,

　　D., Mosley, T., Resendez, J., Nguyen, H., Wilson, L., Campbell, A., Sudarshan, D., &

　　Lloyd, A. (2016). TTG2 controls the developmental regulation of seed coat tannins in

　　Arabidopsis by regulating vacuolar transport steps in the proanthocyanidin pathway.

　　Developmental Biology, *419*(1), 54–63. https://doi.org/10.1016/j.ydbio.2016.03.031

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X., Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T. K., May, M. R., Springer, N. M., … Ware, D. (2017). Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), 524–527. https://doi.org/10.1038/nature22971

Kermicle, J. L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics*, *66*(1), 69–85.

Liu, P., McCarty, D. R., & Koch, K. E. (2016). Transposon mutagenesis and analysis of mutants in UniformMu maize (Zea mays). *Current Protocols in Plant Biology*, *1*(3), 451–465. https://doi.org/10.1002/cppb.20029

Liu, X., Li, S., Yang, W., Mu, B., Jiao, Y., Zhou, X., Zhang, C., Fan, Y., & Chen, R. (2018). Synthesis of seed-specific bidirectional promoters for metabolic engineering of anthocyanin-rich maize. *Plant and Cell Physiology*, *59*(10), 1942–1955. https://doi.org/10.1093/pcp/pcy110

Qin, L., Sun, L., Wei, L., Yuan, J., Kong, F., Zhang, Y., Miao, X., Xia, G., & Liu, S. (2021). Maize SRO1e represses anthocyanin synthesis through regulating the MBW complex in response to abiotic stress. *The Plant Journal*, *105*(4), 1010–1025. https://doi.org/10.1111/tpj.15083

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Vélez-Bermúdez, I.-C., Salazar-Henao, J. E., Fornalé, S., López-Vidriero, I., Franco-Zorrilla, J.-M., Grotewold, E., Gray, J., Solano, R., Schmidt, W., Pagés, M., Riera, M., & Caparros-

Ruiz, D. (2015). A MYB/ZML complex regulates wound-induced lignin genes in maize.

The Plant Cell, *27*(11), 3245–3259. https://doi.org/10.1105/tpc.15.00545

187

APPENDIX A: PROSPECTS FOR ECONOMICAL NATURAL COLORANTS: INSIGHTS FROM MAIZE[1]

Abstract:

Replacing artificial dyes with natural colorants is becoming increasingly popular in foods and beverages. However, natural colorants are often expensive, have lower stability, and reduced variability in hue. Purple corn is rich in anthocyanins and offers a scalable and affordable alternative to synthetic dyes ranging in color from orange to reddish-purple. This diversity is attributable to differences in anthocyanin composition and concentration. Here we review the chemistry, biosynthesis, and genetics of purple corn and outline key factors associated with the feasibility of producing an economical source of natural colorants. Anthocyanin compositional modifications including acylation, methylation, and polymerization with flavan-3-ols can influence color stability and hue, yet there is more to learn regarding the genetic factors responsible for these modifications. Activators and repressors of anthocyanin biosynthesis structural genes as well as factors controlling trafficking and storage largely control anthocyanin yield. Further knowledge of these mechanisms will allow breeders to apply molecular strategies that accelerate the production of purple corn hybrids to meet growing demands for natural colorants.

Diversity in colored corn.

A. Aleurone-pigmented corn kernels derived from a common commercial variety of glass gem corn containing mixtures of carotenoids and anthocyanin in aleurone to create most colors of the visible spectrum. Purple, pink, red, and blue kernels likely contain anthocyanins in aleurone; orange and yellow likely contain carotenoids, and red–orange and green kernels likely contain mixtures of both. B. Various ears of corn illustrating diversity in pigment production. The three major classes of pigments in corn: anthocyanins (1st, 2nd, 3rd, 4th, and 6th ears), phlobaphenes (6th, 8th, and 9th), and carotenoids (4th, 5th, 6th, and 7th). C–G Purple corn; pericarp-pigmented kernels with anthocyanins (F and G may have co-localized phlobaphene pigments). H–L Pericarp-pigmented kernels with phlobaphenes, i.e., red corn. M–Q Anthocyanin-pigmented aleurone lines: blue corn (M, N) and pink corn (O–Q). Figure Credit, Laura Chatham.

APPENDIX B: STANDARD OPERATING PROCEDURE FOR UHPLC ANALYSES

Section 1: Normal UHPLC Operation

1. While the Desktop is off, turn on all units. Wait until all have orange lights before turning on the computer
2. Open "UPLC (online)" (Chemstation Online).
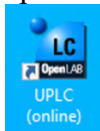
3. Once the machine is booted up, all modules should be orange. Go to the top and select the folder icon next to "Methods" to select the method you would like to use for the day. Make sure the proper column is put in the proper position in the column compartment.
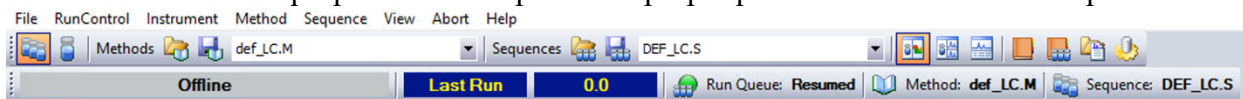
4. Turn on the column thermostat.
5. Check your solvent levels. If you need to add more solvent, make more now. Every month replace the "syringe wash" and the "seal wash". Seal wash is 10% isopropanol in water. Syringe wash is 50% methanol in water. Once finished, tell the UPLC how much is solvent is in the mobile phases by right clicking on the pump module and selecting "Bottle Fillings…".
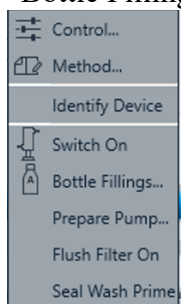
6. Right click on the pump module again and select "Seal Wash Prime"
7. Right click on the pump module again and select "Prepare Pump…"
8. At the top, change the time to 1.5 min and flow rate to 3 mL/min. In the table, change your aqueous solvent to 100%. Click OK and let the pump flush solvent out and remove air bubbles.
9. Now repeat the process with your organic solvent.
10. Right click on the pump module again and select "Method". Change to 100% organic solvent using the table on the left side.
11. Press OK and turn on the pump.
12. Now turn on the DAD.
13. Wait 15 min for organic solvent to clean the machine and any residual compounds in the column.
14. Now go back to "Method" and change back to the starting conditions for the method. Click OK.
15. Right click on the pump module again and select "Prepare Pump…"
16. Select "Pump Conditioning" and change the time to 15 min. This removes air bubbles in the pump.

190

17. Open the desired sequence by selecting the folder icon next to "Sequences". Open the sequence table by clicking on the blue box under Sequences. Fill in sample data.
    a. It is suggested to repeat the first sample and discard the first run, especially if the machine has not been run in a while.
    b. Utilize the "fill down" option to minimize mistakes.
    c. Make sure your last sample is a wash. Choose "Wash.M" for the left column (Pos2) or "WashPos1.M" for the right column. These methods run Solvent D (Methanol) through the system for 15 min.

		Location	Sample Name	Sample Info	Datafile	Data Dir.	Inj. per Vial	Done	Total
						C:..\Data			
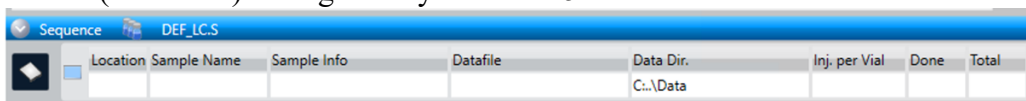
18. Right click on the blue box under sequence and select "Sequence Parameters". Make sure that the "Shutdown" command is "Shut Instrument Down".
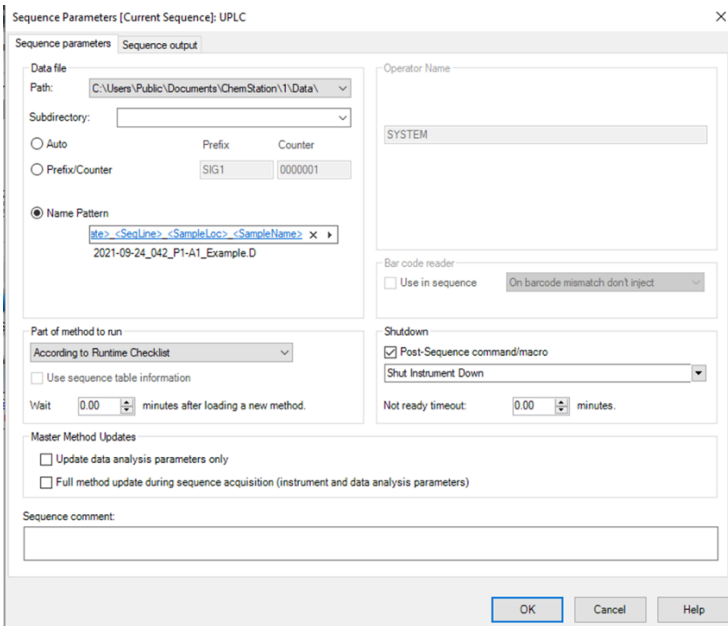
19. Once the 15 min conditioning is over, you are ready to start running samples. Complete the checklist below before proceeding.
20. Click on the Play button next to Sequence above the instrument modules to run.
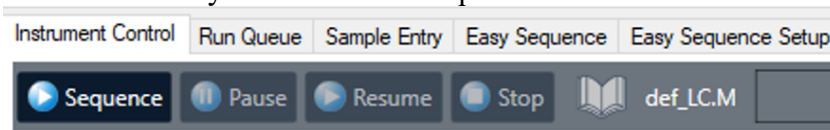
21. Save the method and sequence. Watch the first sample.
22. When complete, make sure all units are orange and none are red. The Wash Method should appear at the top.
23. If desired, turn off the desktop before turning off all UPLC units. We usually keep the machine running.

UPLC Checklist:

Check solvent levels. Are all of them filled? Will you need to refill them before the run is over?

Are bottle fillings correct in Chemstation?

Check the waste container. Does it need to be emptied?

Did you purge all the air from solvent lines?

Check the sample table. Did you select the correct tray? Pay attention to front and back. Double check vials are in the right position and order.

Is the instrument scheduled to shut down after the run? Check "Sequence Parameters" and make sure the instrument shutdown macro is selected.

Section 2: Chromatogram Processing
1. Open Chemstation Offline.
2. On the left-hand side, double-click the batch that you want to analyze.
3. Click on Integration. Does the chromatogram look OK? If not here are things to fix:
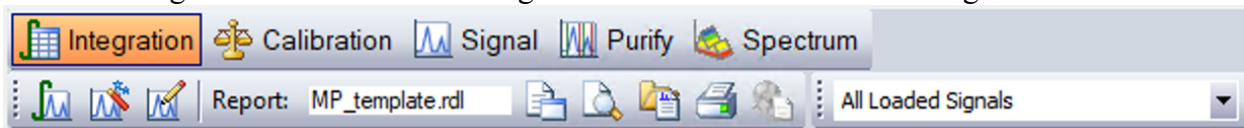
   a. Click on the icon with a pencil (Edit/ Set Integration Events Table).

   b. Change peak width, area reject, and height reject to 0.1. Click away from it and it will update.

   c. Change the Slope Sensitivity. 2 is usually what it is set to, but a higher number will detect minor peaks

   d. Tell Chemstation to ignore the solvent peak/ noise from the beginning of the run by adding a new integration event. This is done by clicking the fourth icon from the left on the top. It will append new rows to the table at the bottom.
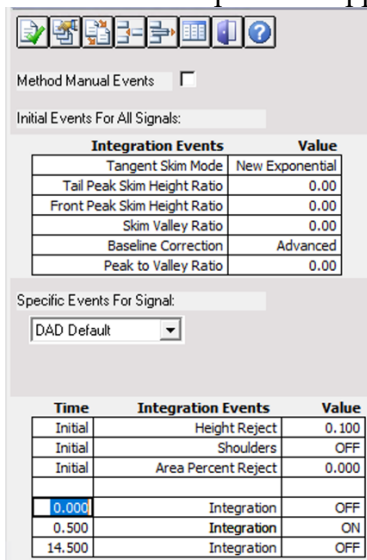
   e. Change the middle column to "Integration" and input the time to turn integration off and on.

   f. There are many other options for manual integration. One scenario is that you may want to change Slope Sensitivity for a small section of the chromatogram if peaks overlap. This can be done by appending new rows to the Integration Events Table and adding a time and function.

   g. If all these parameters fail, you can try clicking on the icon with a red wand for Auto Integrate. It will determine the best parameters for you.

   h. Once satisfied, click on the far left paper icon with a green checkmark.

4. Click on Calibration

a. If a calibration table is already made:
  i. Check the reference compound(s) for several chromatograms. It is indicated as Ref=Yes. Find an approximate retention time for the reference compound during the whole batch excluding the first sample. Input that time in RT. Now press OK. It should update all retention times.
b. If a calibration table is not made:
  i. Go to the Calibration tab at the top and select "New Calibration Table"
  ii. Make sure "Calculate Signals Separately" is chosen.
  iii. If you want to make a fast calibration table, click on the icon with balance over a table. This will add all peaks from the loaded chromatogram into the table.

  iv. If you have peaks to add manually, click on a peak of interest. Choose the icon with a table with a plus sign and a chromatogram (Add new peaks from current chromatogram). Press OK. It will now add that peak to the table.

  v. Continue until all desired peaks are added.
c. You may want to make a standard curve.
  i. Select the "Add new levels to the chromatogram" icon

  ii. Change the level sequentially for as many points as you have in the standard curve. Change the "Amount" to reflect the concentration of your sample. For example, Level 1 to 5 could be 1, 5, 10, 20, 50 µg/mL.

5. Calculate Standard Curves
  a. If you have run the calibration samples already, open "Edit Sequence Table" by clicking on the icon with the vials over a table.

  b. If you have not run samples, start here. Change "Sample Type" to "Calibration" input the level and the amount corresponding to that level.

Line	Sample Location	Sample Name	Method Name	Injection Source	Injection Volume	Inj/Loc	Sample Type	Cal Level	Update RF	Update RT	Cal Interval	Sample Amount
1	D2B-A1	C3G	Maize_Poroshell_1.9_5pct	As Method	10	1	Calibration	5	Average	Average		50

  c. Now save your changes and click the green arrow for "Start Sequence Reprocessing". Wait a while for it to reprocess every sample.
6. Export Data.
  a. Checklist before exporting data:
    i. Make sure Chemstation is behaving and naming and integrating all your compounds of interest.

      ii.   If there are two peaks overlapping, you may need to tinker with the "Signal" tab to manually split peaks.

     iii.   If you have not already, press the green arrow for "Start Sequence Reprocessing" and wait until all the samples are reprocessed.

     iv.   If you would like to add other completed samples from other batches, right click on the batch name in the left-hand table and select "Add Data Files…". This will pull up another table with all the samples in that batch. Left click and choose "Select all rows" to add them.

b. Once data is all present and clean, highlight all the data by clicking and dragging in the sample table.

c. Right click on the table and select "Send Selection to Report Layout"

d. This will take you to the "Report Layout" tab

e. Choose the template to format your data

     i.   "MP_template" compiles all the chromatograms and integration tables for all samples. It is best saved as a PDF

     ii.   "Seq_sum_area" is the most powerful tool. This makes one tab with the RT + areas + amount for each named compound and another tab with the raw integration table + peak areas + total peak area. It only calculates the first wavelength, or DAD1A as it is coded. Export as an Excel file (xlsx)

     iii.   "Seq_sum_area_allDAD" is the same as above, but also gives peak areas for ALL wavelengths quantified.

Section 3: Data Analysis

1. Open the Excel file generated with "Seq_sum_area". Open tab two. Select all data with Cmd/Ctl + A. Now unmerge all cells. Click save. If you have multiple datafiles that need to be processed, combine the data into one sheet of one Excel file.

2. Open R Studio. Run the code below. Replace `"<Path to excel file.xlsx>"` with the real name of your file from Step 1. Also make sure that "sheet" points to the write tab in the Excel document. Change `as.character(c(1:8)))` to 1:7 if R asks you to do so. This will require you to change some code in Step 5 below.

```
install.packages("readxl");install.packages("dplyr")
library("readxl"); library("dplyr")
sheet = read_xlsx("<Path to excel file.xlsx>", sheet = 2,
col_names = as.character(c(1:8)))
write.table(sheet,"~/Desktop/combinedData.txt", sep = "\t")
```

3. Now you have a file with the first column as line number and the data in the remaining columns. However, there are mistakes. Here are the steps to fix these mistakes:

a. Download or open "BBEdit" (https://webstore.illinois.edu/shop/product.aspx) and open the file in this program

b. Remove all quotation marks with Cmd/Ctl + F and replace with nothing.

c. Replace \x{0D}\n with nothing to remove the upside down question marks

d. Replace \t\. (numbers with no beginning "0", e.g. ".14159") with \t0\. (adds the zero).

4. Now open the "processChr.py" script. It is provided after this section and at https://github.com/mpaulsmeyer
   a. Change the "infile" variable to the *absolute* path name of your combinedData.txt file (e.g. "/Users/<username>/Desktop/combinedData.txt") on Line 3 AND Line 58.
   b. Make sure that the vector of compound names at Line 8 is correct for the list of compounds in your samples.
5. Run the "processChr.py" script:
   a. Open "Terminal" on a Mac.
   b. Point Terminal to the directory where "processChr.py" is saved by using the command: `cd /path/to/file`
   c. If using for the first time, use `chmod 755 processChr.py` to make it executable
   d. Now run using: `./processChr.py`
   e. If you receive a "Key error" check that Line 29/30 is correct.
   f. If the data is "out of range", add a column to the first line.
   g. Check that Line 4/5 is correct. Sometimes R changes column number to 7 instead of 8. If this happens, change `float(line[7])` to `float(line[6])`
   h. Open "itemizedData.txt" on the Desktop to see if everything looks correct.
6. Open "itemizedData.txt" in Excel and make a new column with the sample names.
7. Now to calculate peak area of acylation or condensed forms per sample, run the script "peakSumming.py". Code is provided at the end of this section and at https://github.com/mpaulsmeyer.
   a. Open "Terminal" on a Mac.
   b. Point Terminal to the directory where "peakSumming.py" is saved by using the command `cd /path/to/file`
   c. If using for the first time, use `chmod 755 peakSumming.py` to make it executable
   d. The script has two options: `-p` for which peaks to sum, and -o for the output. The options for `-p` are "condensed" or "acylation".
   e. The input file is "/Users/<username>/Desktop/combinedData.txt". Change this to file path in Line 3 AND Line 22 if desired.
   f. Here is some example usage: `./peakSumming.py -p condensed -o ~/Desktop/peakSum.txt`
   g. Of course, you can manipulate this python script by changing Line 33 or Line 37 to reflect any retention time you would like.

```python
#!/usr/bin/env python
#processChr.py Script
inFILE = open("/Users/home/Desktop/combinedData.txt", "r")
#Check: col0=line number, col1=sampleName, col2=date, col3=NA
#col4=compoundName, col5=RT, col6=Amount
outFILE = open("/Users/home/Desktop/itemizedData.txt", "w")

#list of anthocyanins example
compounds = ["AFZ-Pg35DG","anther-CY","anther-PG",\
"C3DMG","CAT-C35DG","CAT-C35DMG","C3DMG","C3G",\
"C3MG","D3DMG","D3G","D3MG","Pg3DMG","Pg3G",\
"Pg3MG","Pn3DMG","Pn3G"]

sample = {}
for line in inFILE:
    line = line.strip('\n')
    line = line.split("\t")
    if line[3][-2:]==".D":
        ID = line[3]
        sample[ID] = {}
inFILE.close()

inFILE = open("/Users/home/Desktop/combinedData.txt", "r")

for line in inFILE:
    line = line.strip('\n')
    line = line.split("\t")
    if line[3][-2:]==".D":
        ID = line[3]
    if (line[1] != "Sample Name") & (line[2] != "NA") \
    & (line[0] != "")& (line[1]!="Injection Data File Name"):
        sample[ID]["Date"] = line[2].split(" ")[0]
        if line[4] != "NA": #[4] is compound name
            sample[ID][line[4]] = float(line[7]) #name=amount
    if line[5] == "Sum": #col6 will say "Sum" check
        sample[ID]["Total"] = float(line[7])
inFILE.close()

outFILE.write("ID\tDate\t")
for i in compounds:
    outFILE.write("{}\t".format(i))
outFILE.write("Total\n")

for ID in sample:
    tmp = [ID]
    tmp.append(sample[ID]["Date"])
    for i in compounds:
        if i in sample[ID]:
```

197

```
                tmp.append(sample[ID][i])
        else:
                sample[ID][i] = 0
                tmp.append(sample[ID][i])
    tmp.append(sample[ID]["Total"])
    for j in tmp:
        outFILE.write("{}\t".format(j))
    outFILE.write("\n")
outFILE.close()
```

```python
#!/usr/bin/env python
import argparse
inFILE = open("/Users/home/Desktop/combinedData.txt", "r")

p = argparse.ArgumentParser()
p.add_argument('-p', '--peaks', required=True, type = str, help
= "Choose condensed or acylation")
p.add_argument('-o', required=True, type = str, help = "Output")
# Parse the command line
args = p.parse_args()
outFILE = open(args.o, "w")

sample = {}
for line in inFILE:
    line = line.strip('\n')
    line = line.split("\t")
    if line[3][-2:]==".D":
        ID = line[3]
        sample[ID] = {}
inFILE.close()

inFILE = open("/Users/home/Desktop/combinedData.txt", "r")
for line in inFILE:
    line = line.strip('\n')
    line = line.split("\t")
    if line[3][-2:]==".D":
        ID = line[3]
        sample[ID]["sum"] = 0.0
    if (line[1] != "Sample Name") & (line[2] != "NA") &\
(line[0] != "") & (line[1]!="Injection Data File Name"):
        sample[ID]["Date"] = line[2].split(" ")[0]
        if(args.peaks == "condensed"):
            if (float(line[5]) < 3.2):
                sample[ID]["sum"] = sample[ID]["sum"] +\
float(line[7])
        if(args.peaks == "acylation"):
            if (float(line[5]) > 5.6):
                sample[ID]["sum"] = sample[ID]["sum"] +\
float(line[7])
inFILE.close()

for ID in sample:
    i = sample[ID]["Date"]
    j = sample[ID]["sum"]
    outFILE.write("{}\t{}\t{}".format(ID,i,j))
    outFILE.write("\n")
outFILE.close()
```

199

APPENDIX C: RNA-SEQ PIPELINE CODE

Analyzing RNA-seq data involves multiple data cleaning and alignment steps. Here is the general workflow that I used:

- Data cleaning with *Fastp* (https://github.com/OpenGene/fastp): Trim the start and end of reads based on quality scores, discard short fragments, and keep fragments with high quality scores.

- Alignment with *STAR* (https://github.com/alexdobin/STAR)

- Counting reads with *featureCounts* (http://subread.sourceforge.net/)

- Differential gene expression analysis using *EdgeR* (https://bioconductor.org/packages/release/bioc/html/edgeR.html)

I used the campus Biocluster to run all of my samples. They refer to jobs as "SLURMs". Each script starts with parameters that the cluster will use to allocate resources to my job. The SLURM parameters are close to optimized and were the most efficient parameters tested for parallelization and memory requirements. More information about SLURM parameters can be found here: https://help.igb.illinois.edu/Biocluster2.

I do not go into detail about every option included within the data processing steps. One must change the options depending on their application. Especially note file handles to make sure inputs and outputs are going to the right places.

Code is included below and also at https://github.com/mpaulsmeyer/

Fastp

```bash
#!/bin/bash
# ----------------SLURM Parameters----------------
#SBATCH -p normal
#SBATCH -n 16
#SBATCH --mem=64g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J fastp
# ----------------Load Modules------------------
module load fastp/0.20.0-IGB-gcc-4.9.4
cwd="/home/trim/fastp" #working directory

#each file was named [MUT/WT]_A3_REP[1-3]_R[1-2].fastq.gz. Line
15 pulls sample name
for f in $(ls -1 /home/raw/ | grep "A3" | cut -d "_" -f 1-3 |
sort | uniq); do
    echo "Filtering ${f}"
    fastp -w $SLURM_NTASKS \
        -i /home/raw/"${f}"_R1.fastq.gz \
        -I /home/raw/"${f}"_R2.fastq.gz \
        -e 30 -l 50 -x --
adapter_fasta="/home/raw/myAdapter.fa" \
        -t 3 -T 3 -f 10 -F 10 --overrepresentation_analysis \
        -o "${cwd}"/"${f}"_trimmed_FP.fastq.gz \
        -O "${cwd}"/"${f}"_trimmed_RP.fastq.gz \
        --unpaired1="${cwd}/${f}_trimmed_UP.fastq.gz" \
        --unpaired2="${cwd}/${f}_trimmed_UP.fastq.gz" \
        -j "${cwd}/${f}_fastp_results.json" \
        -h "${cwd}/${f}_fastp_results.html" \
        -R "${cwd}/${f}.results";
done
```

STAR

```bash
#!/bin/bash
# ---------------SLURM Parameters---------------
#SBATCH -p normal
#SBATCH -n 20
#SBATCH --mem=64g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J STAR
# ---------------Load Modules------------------
module load STAR/2.7.6a-IGB-gcc-8.2.0

mkdir /scratch/starTMP/
cwd="/home" #working directory
clean="${cwd}/A3/trim/fastp" #path to cleaned data
scratch="/scratch/starTMP/$SLURM_JOB_ID" #temp directory
date="1Jan"

echo "Running Genome Indexer"
STAR --runThreadN $SLURM_NTASKS \
--runMode genomeGenerate --limitGenomeGenerateRAM 62000000000 \
--sjdbOverhang 136 --genomeDir "${cwd}"/db/STARfull \
--genomeFastaFiles "${cwd}"/ref/Zea_mays.AGPv4.dna.toplevel.fa \
--sjdbGTFfile "${cwd}"/ref/Zea_mays.B73_RefGen_v4.49.gtf

#each file was named [MUT/WT]_A3_REP[1-3]_trimmed_[FP/RP].fastq.gz
for f in $(ls -1 "${clean}" | cut -d "_" -f 1-3 | sort | uniq); do
      echo "Running Aligner for ${f}"
      #pull WT or MUT for SAM file
      s=$(echo "${f}" | sed -E 's/([MUTW].+)_A3_REP[1-3]/\1/')
      STAR --runThreadN $SLURM_NTASKS --genomeDir "${cwd}"/db/STARfull \
      --readFilesIn "${clean}"/"${f}"_trimmed_FP.fastq.gz \
            "${clean}"/"${f}"_trimmed_RP.fastq.gz \
      --readFilesCommand zcat --sjdbGTFfile \
            "${cwd}"/ref/Zea_mays.B73_RefGen_v4.49.gtf \
      --outSAMtype BAM SortedByCoordinate \
      --outFileNamePrefix "${cwd}"/A3/aligned/STAR/"${date}"/pass/"${f}". \
      --outTmpDir "${scratch}" \
      --outSAMattrRGline ID:"${f}" LB:1 PL:Illumina PU:HYGYHDRXX SM:"${s}"
      rm -fr "${scratch}";
done

echo "Begin Two-pass Mode"
echo "Indexing with the new transcriptome"
STAR --runThreadN $SLURM_NTASKS --runMode genomeGenerate \
      --limitGenomeGenerateRAM 62000000000 --sjdbOverhang 136 \
      --genomeDir "${cwd}"/db/two-pass \
      --genomeFastaFiles "${cwd}"/ref/Zea_mays.AGPv4.dna.toplevel.fa \
      --sjdbGTFfile "${cwd}"/ref/Zea_mays.B73_RefGen_v4.49.gtf \
      --sjdbFileChrStartEnd "${cwd}"/A3/aligned/STAR/"${date}"/*.SJ.out.tab
```

```
echo "Begin the second pass"
for f in $(ls -1 "${clean}" | cut -d "_" -f 1-3 | sort | uniq); do
    echo "Running Second Aligner for ${f}"
    s=$(echo "${f}" | sed -E 's/([MUTW].+)_A3_REP[1-3]/\1/')
    STAR --runThreadN $SLURM_NTASKS --genomeDir "${cwd}"/db/two-pass \
        --readFilesIn "${clean}"/"${f}"_trimmed_FP.fastq.gz \
        "${clean}"/"${f}"_trimmed_RP.fastq.gz --readFilesCommand zcat \
        --sjdbGTFfile "${cwd}"/ref/Zea_mays.B73_RefGen_v4.49.gtf \
        --sjdbFileChrStartEnd \
            "${cwd}"/A3/aligned/STAR/"${date}"/*.SJ.out.tab \
    --outSAMtype BAM SortedByCoordinate \
    --outFileNamePrefix "${cwd}"/A3/aligned/STAR/"${date}"/pass2/"${f}". \
    --outTmpDir "${scratch}" \
    --outSAMattrRGline ID:"${f}" LB:1 PL:Illumina PU:HYGYHDRXX SM:"${s}"
    rm -fr "${scratch}";
done
```

featureCounts

```bash
#!/bin/bash
# ----------------SLURM Parameters----------------
#SBATCH -p normal
#SBATCH -n 6
#SBATCH --mem=64g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J Fcounts
# ----------------Load Modules--------------------
module load Subread/2.0.0-IGB-gcc-8.2.0
cwd="/home" #working directory
date="1Jan" #date ran STAR pipeline

for f in $(ls -1 "${cwd}"/A3/trim/fastp | \
      cut -d "_" -f 1-3 | sort | uniq); do
   echo "Running featureCounts for ${f}"
   featureCounts -p -T $SLURM_NTASKS -s 2 \
      -a "${cwd}"/ref/Zea_mays.B73_RefGen_v4.49.gtf \
      -t exon -g gene_id --verbose \
      -o "${cwd}"/deg/featureCounts/"${date}"/"${f}"_rev_featurecounts.txt \
      "${cwd}"/aligned/STAR/"${date}"/"${f}".Aligned.sortedByCoord.out.bam;
done
```

EdgeR

```
library(BiocManager)
#install("edgeR")
library(edgeR)

############################################################
############### Import Featurecounts Data ##################
############################################################

setwd(<Featurecounts results directory>)
options(stringsAsFactors=FALSE)
dirs = list.files(getwd(), pattern = "*.txt$")
for (i in 1:length(dirs)){
  #assign a variable to each file and only keep sample name
  name = paste(substr(dirs[i],1,11))
  assign(name,
         read.delim(dirs[i], sep = "\t",
                    row.names=1, skip = 2,
                    header = F))
}

#now that you have loaded all data, make the counts file
Counts <- cbind(MUT_A3_REP1[,6],MUT_A3_REP2[,6],
                MUT_A3_REP3[,6],WT_A3_REP1_[,6],
                WT_A3_REP2_[,6],WT_A3_REP3_[,6])
row.names(Counts) = row.names(WT_A3_REP1_)
colnames(Counts) = c("MUT_A3_REP1","MUT_A3_REP2",
                     "MUT_A3_REP3","WT_A3_REP1",
                     "WT_A3_REP2","WT_A3_REP3")
str(Counts)

write.table(Counts,"Combined_A3_featureCounts.tsv")
Counts = read.table("Combined_A3_featureCounts.tsv",
                    header = T, row.names = 1)
```

```
###############################################################
###################### Run EdgeR ###########################
###############################################################

#change each column to a treatment
group = as.factor(c("MUT","MUT","MUT","WT","WT","WT"))
dge <- DGEList(counts=Counts, group=group)

#remove samples with less than 5 counts per sample on average
keep.dge <- rowSums(dge$counts) >= 30
dge <- dge[keep.dge,]

#Design Matrix
design.er <- model.matrix(~0 + group, data = dge$samples)
design.er

# Before we conduct the hypothesis tests,
# we need to estimate that dispersion parameter for each gene.
# Estimate common, trended, and tagwise dispersions
dge.norm <- calcNormFactors(dge, method = "TMM")
dge.norm <- estimateGLMCommonDisp(dge.norm, design.er)
dge.norm <- estimateGLMTrendedDisp(dge.norm, design.er)
dge.norm <- estimateGLMTagwiseDisp(dge.norm, design.er)

#Now we fit the full model incorporating dispersion estimates!
fit.QL <- glmQLFit(dge.norm, design.er,) #Quasi F likelihood

#adjust using the log2-fold cutoff
QLPop.adj <- glmTreat(fit.QL, contrast=c(1,-1), lfc=log2(1.2))
topTags(LrtPop, adjust.method = "fdr", sort.by = "PValue")
summary(decideTests(QLPop.adj,
                    adjust.method = "fdr",
                    p=0.01, lfc = 1.3))
plotMD(QLPop.adj, p.value = 0.01)
```

APPENDIX D: VARIANT CALLING PIPELINES

Genotyping-by-sequencing (GBS) is one of the most efficient and cost-effective sequencing methods for maize. However, GBS is fraught with genotype errors due to low coverage. In order to circumvent this problem, several SNP discovery pipelines were utilized for Chapter 4.

- TASSEL5 (https://www.maizegenetics.net/tassel) is a widely used pipeline in maize research. However, it appears to be optimized for inbred lines and not highly heterozygous materials like I used in Chapter 4.
- SAMtools/BCFtools (http://www.htslib.org/): For my population, this pipeline provided the most high-confidence SNPs.
- GATK: I did not complete the entire GATK pipeline. The last step is a neural network step that utilizes a database of known SNPs. However, the developers of this program obviously do not care about the functionality of their program for researchers outside human genetics. There are no databases for SNPs for maize to run this algorithm. In addition, this pipeline uses a lot of resources and takes several days to run on the Biocluster. I have altered the code to be extremely efficient. I hope that future researchers use my code to speed up the pipeline.

Code is included below and also at https://github.com/mpaulsmeyer/

TASSEL 5.0 Pipeline

```bash
#!/bin/bash
#Download the latest reference genome from
#https://download.maizegdb.org/
#Download the latest version of Bowtie from http://bowtie-
#bio.sourceforge.net/index.shtml
#Download TASSEL 5.0 standalone from
#https://www.maizegenetics.net/tassel
#Add your sequence to "mySequence"
#Make empty dirs for db, maizeTags, logfiles, hdf5, and keys

#Define the directory where run_pipeline.py
#and all directories are stored
cwd="/home/shared/MAL"
cd "${cwd}"

"${cwd}"/run_pipeline.pl -Xmx16g -GBSSeqToTagDBPlugin \
    -i "${cwd}"/mySequence/Lane1 \
    -db "${cwd}"/db/maizeTags.db \
    -e PstI -k "${cwd}"/keys/MAL1.txt \
    -kmerLength 64 -minKmerL 20 -batchSize 1 -c 1 \
    -mxKmerNum 100000000 -endPlugin | |
    tee "${cwd}"/logfiles/logfile1.txt
"${cwd}"/run_pipeline.pl -Xmx16g -TagExportToFastqPlugin \
    -db "${cwd}"/db/maizeTags.db \
    -o "${cwd}"/mergedTags/maizeTags.fa.gz \
    -c 1 -endPlugin | tee "${cwd}"/logfiles/logfile2.txt
"${cwd}"/Bowtie/bowtie2 -p 4 --very-sensitive \
    -U "${cwd}"/mergedTags/maizeTags.fa.gz \
    -x "${cwd}"/build/Mo17build \
    -S "${cwd}"/mergedTags/aligned.MAL1.sam  | \
    tee "${cwd}"/logfiles/logfile4.txt
"${cwd}"/run_pipeline.pl -Xmx16g -SAMToGBSdbPlugin \
    -db "${cwd}"/db/maizeTags.db \
    -i "${cwd}"/mergedTags/aligned.MAL1.sam \
    -aProp 0.0 -aLen 0 -endPlugin | \
    tee "${cwd}"/logfiles/logfile5.txt
${cwd}/run_pipeline.pl -Xmx16g -DiscoverySNPCallerPluginV2 \
    -db "${cwd}"/db/maizeTags.db \
    -sC 1 -eC 10 -mnLCov 0.001 -mnMAF 0.001 \
    -endPlugin | tee "${cwd}"/logfiles/logfile6.txt
"${cwd}"/run_pipeline.pl -Xmx16g -SNPQualityProfilerPlugin \
    -db "${cwd}"/db/maizeTags.db \
    -statFile "outputStats.MAL1.txt"
    -deleteOldData true -endPlugin | \
    tee "${cwd}"/logfiles/logfileMo7.txt
${cwd}/run_pipeline.pl -Xmx16g -ProductionSNPCallerPluginV2 \
```

```
-db "${cwd}"/db/maizeTags.db \
-e PstI -i "${cwd}"/mySequence/Lane1 \
-k "${cwd}"/keys/MAL1.txt
-o "${cwd}"/hdf5/MAL1.genotypes.h5
-batchSize 1 -mnQS 30 -endPlugin | \
tee "${cwd}"/logfiles/logfile8.txt
```

Alignment Using BWA

```bash
#!/bin/bash
# ---------------SLURM Parameters---------------
#SBATCH -p normal
#SBATCH -n 16
#SBATCH --mem=64g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J bwa
# ---------------Load Modules------------------
module load Stacks/2.54-IGB-gcc-8.2.0
module load BWA/0.7.17-IGB-gcc-8.2.0
module load SAMtools/1.10-IGB-gcc-8.2.0


cwd="/home/MAL" #working directory
ref="/home/ref/Zm-B73-REFERENCE-NAM-5.0.fa"
bwa index -p "${home}"/db/bwa/B73v5 -a bwtsw "${ref}"


#barcodes are in a file with sample\tbarcode on each line
echo "Demultiplex MAL2"
process_radtags -p "${cwd}"/raw/Lane2 \
     -o "${cwd}"/samples -b "${cwd}"/MAL2_barcodes.txt \
     -e pstI --len_limit 50 --retain_header -r -c -q


echo "Demultiplex MAL1"
process_radtags -p "${cwd}"/raw/Lane1 -o "${cwd}"/samples -b
"${cwd}"/MAL1_barcodes.txt \
     -e pstI --len_limit 50 --retain_header -r -c -q


echo "Begin MAL2 Alignment"
for f in $(ls -1 "${cwd}"/Aligned/MAL2/ | cut -d "." -f 1 | sort | uniq); do
    echo "Aligning sample ${f}"
    RGID=$(zcat "${cwd}"/samples/"${f}".fq.gz | head -n 1 | cut -c 13-21)
    RGLB=$(zcat "${cwd}"/samples/"${f}".fq.gz | head -n 1 | cut -c 23)
    RGPU=$(zcat "${cwd}"/samples/"${f}".fq.gz | head -n 1 | cut -c 9-23)
    bwa mem -t $SLURM_NTASKS \
      -R "@RG\tID:${RGID}\tPU:${RGPU}\tSM:${f}\tPL:ILLUMINA\tLB:${RGLB}" \
        "${ref}"/db/bwa/B73v5 "${cwd}"/samples/"${f}".fq.gz | \
    samtools view -bh -o "${cwd}"/Aligned/MAL2/"${f}".aligned.bam
    samtools sort -@ $SLURM_NTASKS-1 \
      -o "${cwd}"/Aligned/MAL2/"${f}".aligned.sort.bam \
      "${cwd}"/Aligned/MAL2/"${f}".aligned.bam
    samtools index -@ $SLURM_NTASKS-1 \
      "${cwd}"/Aligned/MAL2/"${f}".aligned.sort.bam;
done
```

SAMtools / BCFtools variant calling pipeline

```bash
#!/bin/bash
# ----------------SLURM Parameters----------------
#SBATCH -p normal
#SBATCH -n 8
#SBATCH --mem=64g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J MP_snp
##########################################################################
############# BCFtools Variant Calling Pipeline  #########################
##########################################################################

module load SAMtools/1.10-IGB-gcc-8.2.0

cwd="/home/MAL" #working directory
ref="/home/ref" #ref directory
log="/home/MAL/SNP/samtools/variant.log" #optional log
depth=6 #read depth per SNP for MAL
max=250 #max read depth for MAL

#Need to index the reference genome for the next steps
echo "index Mo17" | tee -a "${log}"
samtools faidx "${ref}"/Zm-Mo17-REFERENCE-CAU-1.0.fa
module purge

module load BCFtools/1.12-IGB-gcc-8.2.0
# max is depth saved, B is do not readjust alignments,
# Ou is a way to pipe commands
# -mv means variants only and rare allele caller
echo "Starting variant pipeline for MAL 1" | tee -a "${log}"
bcftools mpileup -Ou --skip-indels -d "${max}" --threads $SLURM_NTASKS \
      -f "${ref}"/Zm-Mo17-REFERENCE-CAU-1.0.fa \
      "${cwd}"/Aligned/MAL1/*.sort.bam | \
bcftools call --threads $SLURM_NTASKS -mv > \
      "${cwd}"/SNP/samtools/MAL1.raw.SNP.vcf

#Filter depth < $depth and qual < 30. Remove indels
echo "filtering final VCF" | tee -a "${log}"
bcftools filter -i '%QUAL>=30 & DP>=6 & TYPE="snp"' \
      "${cwd}"/SNP/samtools/MAL1.raw.SNP.vcf > \
      "${cwd}"/SNP/samtools/MAL1.filtered.SNP.vcf
module purge

#optional step to impute missing data
# module load beagle/5.1-Java-1.8.0_152
# beagle --java-options "-Xmx56g" \
      gt="${cwd}"/SNP/samtools/CBlock.filtered.SNP.vcf \
      out="${cwd}"/SNP/samtools/CBlock.imputed.b1.SNP.vcf \
      gp=true window=10 overlap=5 ne=2
# beagle --java-options "-Xmx56g" \
      gt="${cwd}"/SNP/samtools/CBlock.filtered.SNP.vcf \
      out="${cwd}"/SNP/samtools/CBlock.imputed.b2.SNP.vcf gp=true
```

211

GATK variant calling pipeline

Step 1: Haplotype Calling

```bash
#!/bin/bash
# ----------------SLURM Parameters----------------
#SBATCH -p normal
#SBATCH -n 64
#SBATCH --mem=192g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J GATK_step1
# ----------------Load Modules------------------
module load GATK/4.0.9.0-IGB-gcc-4.9.4-\
     Java-1.8.0_152-Python-3.6.1
module load BCFtools/1.9-IGB-gcc-4.9.4

cwd="/home/MAL" #working directory
chr=(chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10)

echo "Begin GATK Pipeline for MAL1"
ref="/home/ref/Zm-Mo17-REFERENCE-CAU-1.0.fa" #Mo17 genome
geno="$cwd/SNP/GATK/MAL1"

for f in $(ls -1 "${cwd}"/Aligned/MAL1/ | cut -d "." -f 1 | \
     sort | uniq); do
     timestamp=$(date +%H:%M:%S);
     echo "Calling SNPs For $f at $timestamp"
     for int in $(echo ${chr[@]}); do
        gatk --java-options "-Xmx16g" HaplotypeCaller \
           -R "${ref}" \
           -I "${cwd}"/Aligned/MAL1/"${f}".aligned.sort.bam \
           -O "${geno}"/"${int}".c.vcf.gz \
           -ERC GVCF \
           -L "${int}" \
           #Most important line for parallelizing the shell:
           --native-pair-hmm-threads 6 & done;
     wait
     echo "${f}     ${f}.raw.g.vcf.gz" >> \
         "${geno}"/MAL1_gatk_map.txt
     bcftools concat -o "${geno}"/"${f}".raw.g.vcf \
         "${geno}"/chr1.c.vcf.gz "${geno}"/chr2.c.vcf.gz \
         "${geno}"/chr3.c.vcf.gz "${geno}"/chr4.c.vcf.gz \
         "${geno}"/chr5.c.vcf.gz "${geno}"/chr6.c.vcf.gz \
         "${geno}"/chr7.c.vcf.gz "${geno}"/chr8.c.vcf.gz\
         "${geno}"/chr9.c.vcf.gz "${geno}"/chr10.c.vcf.gz
     rm "${geno}"/*.c.gz "${geno}"/*.tbi
```

212

```
      gzip "${geno}"/"${f}".raw.g.vcf #optional
      timestamp=$(date +%H:%M:%S);
      echo "Ended Calling SNPs For $f at $timestamp";
done

#You can split this next part into a different job
#and use 24 CPUs to be more efficient

echo "Begin Compiling MAL1 Genomic Database"
geno="$cwd/SNP/GATK/MAL1"
ref="/home/ref/Zm-Mo17-REFERENCE-CAU-1.0.fa"
for int in $(echo ${chr[@]}); do
    gatk --java-options "-Xmx186g" \
          GenomicsDBImport \
          --genomicsdb-workspace-path "${geno}"/genomicsdb \
          --sample-name-map "${geno}"/MAL1_gatk_map \
          --batch-size 0 \
          -L "${int}" \
          --reader-threads $SLURM_NTASKS
    gatk --java-options "-Xmx186g" GenotypeGVCFs \
         -R "${ref}" \
         -V gendb://"${geno}"/genomicsdb \
         -O "${geno}"/MAL1."${int}".combined.raw.vcf.gz
    rm -fr "${geno}"/genomicsdb;
done
```

GATK Continued:

Step 2: Database creation and SNP compilation

```bash
#!/bin/bash
# ---------------SLURM Parameters---------------
#SBATCH -p normal
#SBATCH -n 24
#SBATCH --mem=192g
#SBATCH -N 1
#SBATCH --mail-user=<email>
#SBATCH --mail-type=ALL
#SBATCH -J GATK_step2
# ---------------Load Modules-------------------
module load GATK/4.0.9.0-IGB-gcc-4.9.4-\
        Java-1.8.0_152-Python-3.6.1

chr=(chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10)

echo "Begin Compiling MAL1 Genomic Database"
geno="/home/MAL/SNP/GATK/MAL1"
ref="/home/ref/Zm-Mo17-REFERENCE-CAU-1.0.fa"
for int in $(echo ${chr[@]}); do
    gatk --java-options "-Xmx186g" \
        GenomicsDBImport \
        --genomicsdb-workspace-path "${geno}"/genomicsdb \
        --sample-name-map "${geno}"/MAL1_gatk_map \
        --batch-size 0 \
        -L "${int}" \
        --reader-threads $SLURM_NTASKS
    gatk --java-options "-Xmx186g" GenotypeGVCFs \
         -R "${ref}" \
         -V gendb://"${geno}"/genomicsdb \
         -O "${geno}"/MAL1."${int}".combined.raw.vcf.gz
    rm -fr "${geno}"/genomicsdb;
done
#Filter SNPs using SAMtools/BCFtools
```

APPENDIX E: QTL MAPPING IN R

```
###############################################################
################### Smoothing Function ###################
###############################################################
#Dependencies:
#SNPtable is a file in the format of col 1 = 1 to n SNPs, col 2 = SNP name,
#col 3 = chromosome, col 4 = position
#Genotype is a matrix with no header. The row names are samples
#Row 1 is SNP names. Row 2 is chromosome. Row 3 to n is genotype info

#method1 is similar to 10.1186/1471-2164-15-433
#method2 uses mode. This was not used in Chapter 4.

smoothGeno = function(genotype = genotype, SNPtable = SNPtable,
                      window = 20, step = 5, snp = "method1"){
  if((window %% step) != 0){
    warning("Window size not divisible by step size")
    stop()
  }
  #method1
  #if the most common marker is more than 75%, keep it.
  #If not, it is heterozygous
  snpDecide = function(x) {
    if(max(table(as.matrix(x)))/sum(table(as.matrix(x)))>0.75){
      return(as.numeric(names(which.max(table(as.matrix(x))))))
    }
    else{return(0.5)}
  }
  #method2
  #Mode function from
  #https://www.tutorialspoint.com/r/r_mean_median_mode.htm
  getmode <- function(v) {
    uniqv <- unique(as.vector(v))
    uniqv[which.max(tabulate(match(v, uniqv)))]
  }
  smoothSNP = as.vector(genotype[,1])
  for(i in 1:10){
    tmpSNP = genotype[,which(genotype[2,]==i)] #subset by chr
    k = 1 #k is ncol of genotypes
    while(k < (ncol(tmpSNP)-window)){
      if(snp == "method1"){
        smoothSNP = cbind(smoothSNP, #append new genotypes
                    c(as.vector(tmpSNP[1,k + floor((window/2))]),
                      i, #report middle position name
                      #add the vector of genotypes
                      unlist(apply(tmpSNP[3:nrow(genotype),
```

215

```
                                  k:(k + window - 1)], 1, snpDecide))))
      }
      else if(snp == "method2"){
        smoothSNP = cbind(smoothSNP,
                     c(as.vector(tmpSNP[1,k + floor((window/2))]),
                       i, #report middle position name
                       unlist(apply(tmpSNP[2:nrow(genotype),
                              k:(k + window - 1)], 1, getmode))))
      }
      else{warning("SNP decision method not recognized"); stop()}
      k = k + step
    }
  }
  return(smoothSNP)
}

SNPtable = read.table("SNPtable.txt", header = T, stringsAsFactors = F)
smoothSNP = smoothGeno(genotype = genotype, SNPtable = SNPtable,
                       window = 20, step = 2, snp="method1")
dim(smoothSNP)

################################################################
############### Data cleaning using R/qtl ###################
################################################################

#initial data import
library(qtl)
library(parallel)
Crossfile = read.cross(dir = getwd(),format="csvs",
                       genfile = "MAL1.csv",error.prob = 0.0001,
                       phefile = "MAL1.phenotypes.csv",
                       genotypes = c(1,0.5,0,NA,NA),estimate.map = T,
                       map.function = "c-f", crosstype ="bcsft",
                       BC.gen = 1, F.gen=0)
summaryMap(Crossfile); plotMap(Crossfile);
jpeg("genoimage.jpg", height = 1200, width = 1600)
        geno.image(Crossfile) #colorcodes your genotypes
dev.off()

################################################################
#geno summarys
gt = geno.table(CrossfileTEST4, scanone.output=TRUE)
sum(gt$AA);sum(gt$AB);sum(gt$BB); all = sum(gt$AA)+sum(gt$AB)+sum(gt$BB)
sum(gt$AA)/all; sum(gt$AB)/all; sum(gt$BB)/all
```

```
################################################################################
#removing markers with segregation distortion
segdist = c() #is it different from 1:1 ratio?
for (i in 1:nrow(gt)){
  segdist[i] = chisq.test(c(nrow(genotype)*gt$AA[i],
                            nrow(genotype)*gt$AB[i]),p=c(0.5,0.5),
                            correct = F)$p.value
}
sum(segdist<0.05/nrow(gt))

Crossfile = drop.markers(Crossfile,
                         rownames(gt[which(segdist<0.05/nrow(gt)),]))

################################################################################
#estimate recombination frequencies and see if there are unlinked markers
Crossfile = est.rf(Crossfile)
rf = pull.rf(Crossfile, what = "rf")

jpeg(filename = "RF plot.jpg", width = 1600, height = 1200)
        plotRF(Crossfile, chr=c(1:10), what="rf",
               alternate.chrid=F, zmax=12,
               mark.diagonal=FALSE,col.scheme="viridis")
dev.off()

################################################################################
#Using recombination frequency data, remove any markers that appear to be
#unlinked to their own chromosome.

SNPtable = read.table("SNPtable.txt", header = T)

outliers = c()
for(i in 1:10){
  for(j in 1:10){
    avgrf = apply(rf[which(SNPtable[,3]==i),which(SNPtable[,3]==j)],1,mean)
    #anything 3 std dev is an outlier
    outliers = c(outliers,names(avgrf[which(abs(scale(avgrf))>3)]))
  }
}
outliers

CrossfileDrop = drop.markers(cross = Crossfile, outliers)

################################################################################
# What is the genotyping error rate?

#method adapted from https://rqtl.org/tutorials/geneticmaps.pdf but
#added parallelization
```

217

```
library(parallel)
loglik = err = c(0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01)
loglik = unlist(mclapply(err, mc.cores = detectCores()-2, function (x) {
        tempmap = est.map(CrossfileDrop, map.function = "c-f", error.prob=x)
        loglik = sum(sapply(tempmap, attr, "loglik"))
        return(loglik)
    }
  )
)
lod = (loglik - max(loglik))/log(10)
plot(err,lod)
error = 0.0001 #define based on results

################################################################################
#See if there are any markers significantly altering the genetic map

#I added parallelization
dropone = mclapply(c(1:10), function (x) {
  droponemarker(CrossfileDrop, chr = x)
}, mc.cores = detectCores()-2)
dropone = do.call(rbind.data.frame, dropone)

summary(dropone); plot(dropone); hist(dropone[,4])

drop = rownames(dropone[which(dropone[,4]>0),])#change 0 depending on results
CrossfileDrop = drop.markers(cross = CrossfileDrop, drop)

################################################################################
#another diagnostic is looking at pairs of markers that are greater
#than 50 cM apart. That means they are not linked and have genotype errors.

CrossfileDrop = est.rf(CrossfileDrop)

markers = matrix(NA,nrow=nrows(SNPtable), ncol = 1)
m.diff = c()
names = c()
for (i in 1:10){
  #makes a matrix of all markers with cumulative cM
  markers = pull.map(NewMap,i,as.table=T)
  #finds dif between each marker and next in terms of cM
  m.diff = diff(markers[,2])
  names = c(names,markernames(NewMap,chr = i)
          [which(m.diff>50)],
          markernames(NewMap,chr = i)[which(m.diff>50)+1])
          #one can change 50 to whatever cM distance desired
  print(names)
}
```

```
CrossfileDrop = drop.markers(CrossfileDrop,names)

################################################################
#inspect the recombination frequency plot once again before mapping
#inspect the error rate once again before mapping

CrossfileDrop = est.rf(CrossfileDrop)
rf = pull.rf(CrossfileDrop, what = "rf")

jpeg(filename = "RF plot clean.jpg", width = 1600, height = 1200)
        plotRF(CrossfileDrop, chr=c(1:10),
                   what="rf", alternate.chrid=F,
                   zmax=12, mark.diagonal=FALSE,col.scheme="viridis")
dev.off()

loglik = err = c(0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01)
loglik = unlist(mclapply(err, mc.cores = detectCores()-2, function (x) {
        tempmap = est.map(CrossfileDrop, map.function = "c-f", error.prob=x)
        loglik = sum(sapply(tempmap, attr, "loglik"))
        return(loglik)
     }
   )
)
lod = (loglik - max(loglik))/log(10)
plot(err,lod)
error = 0.0001 #define based on results

################################################################
##################### QTL Mapping #######################
################################################################

genoprob = calc.genoprob(CrossfileDrop,map.function="c-f",
                          step=0, stepwidth="fixed", error.prob = error)
summary(genoprob)

library(parallel)
pdf("Manhattan Plots.pdf")
for (i in 2:ncol(pull.pheno(genoprob))){
  threshold = scanone(genoprob, chr = c(1:10), pheno.col=i,
                      model="normal", method="em", use="all.obs",
                      n.perm = 1000,n.cluster=detectCores()-2)
  scanoneTry = scanone(genoprob, chr = c(1:10), pheno.col=i,
                       model="normal", method="em", use="all.obs")
  plot(scanoneTry, chr = c(1:10), col="darkblue",
       main = names(pull.pheno(genoprob)[i]),
       lwd=2, gap=25, mtick = "line", add=F)
  abline(h = summary(threshold, alpha = 0.05),
```

```
            lty = 2, col = "orange", lwd = 2)
}
dev.off()

#If interested in a specific phenotype
scanoneTry = scanone(genoprob, chr = c(1:10),
                     pheno.col=2, model="normal",
                     method="em", use="all.obs")
#gives the results of the mapping
lodint(results=scanoneTry, chr=8, lodcolumn=1,
       expandtomarkers =T,qtl.index=1)

#running a regression in R/qtl
qtl = makeqtl(genoprob, chr = c(6,8,9), #all significant chr
              pos = c(100,20,11), what = "prob") #most sig marker pos
fitQTL = fitqtl(cross = genoprob,forceXcovar = F, dropone = T,
              #each "Q" is a QTL in order of use in makeqtl
              #can add "*" for interaction
              formula = y ~ Q1+Q2+Q3,
              qtl = qtl,
              model="normal",method = "hk",
              pheno.col = 2,
              covar = NULL,get.ests=T)
fitQTL$result.drop
summary(fitQTL)
```

APPENDIX F: SUPPLEMENTARY TABLE

Supplementary Table for Chapter 5.

Hybrid names correspond to Table 5.1. ACN = anthocyanin concentration in terms of mg anthocyanin per kg maize. Log ACN Conc = Transformed values of the natural log of anthocyanin concentration. ACN Yield = anthocyanin yield on a per-acre basis in terms of pounds per acre. Log ACN Yield = Transformed values of the natural log of anthocyanin yield per acre. UIUC = University of Illinois location only. Full dataset includes all locations.

Hybrid Name	Grain Yield (bu/acre) Full Dataset	Grain Yield (bu/acre) UIUC	ACN (mg/kg) Full Data	Log ACN (mg/kg) Full Dataset	ACN (mg/kg) UIUC	Log ACN (mg/kg) UIUC	ACN Yield (lbs/acre) Full Dataset	Log ACN Yield (lbs/acre) Full Dataset
B0011CU	131.55	135.40	209.70	182.38	221.94	184.98	1.75	1.37
B0161CU	138.89	134.38	171.39	139.31	185.94	136.52	1.71	1.20
B2192CU	136.37	144.51	228.75	214.60	252.16	240.63	2.06	1.93
B2193CU	145.65	159.84	225.41	187.63	225.83	175.29	2.02	1.55
B2223CU	155.28	162.35	264.49	234.19	257.98	225.76	2.43	2.10
B2232CU	145.38	164.80	251.74	216.98	261.90	221.81	2.43	2.02
B2243CU	144.92	141.64	246.55	208.76	290.28	280.57	2.31	2.07
B2293CU	153.15	170.89	243.11	208.29	267.98	228.52	2.64	2.14
B2312CU	147.74	154.82	184.14	154.52	170.54	141.62	1.40	1.19
B2384CU	139.58	155.83	236.23	213.64	234.72	201.92	2.17	1.76
B2405CU	145.11	157.85	163.91	132.45	170.05	122.44	1.48	1.08
B2433CU	156.17	144.91	212.92	190.36	227.37	195.46	1.88	1.56
B2465CU	148.56	160.27	246.96	211.32	262.16	229.80	2.51	2.17
B2492CU	152.16	168.64	178.34	157.85	187.16	159.28	1.87	1.53
B2553CU	150.47	169.92	194.64	185.25	199.10	186.68	1.81	1.69
B2581CU	157.25	176.72	204.65	180.26	224.58	196.59	2.18	1.96
B2601CU	147.49	161.66	208.16	192.68	223.02	205.25	2.03	1.91
B2612CU	168.01	174.34	211.18	192.66	223.94	203.25	2.20	1.98
B2664CU	164.62	196.63	202.46	187.45	203.53	186.75	2.23	2.05
B2691CU	154.60	176.51	242.68	225.56	245.40	231.04	2.50	2.26
B2702CU	149.58	162.57	240.66	210.90	247.04	212.36	2.19	1.85
B2725CU	147.69	161.67	171.47	134.04	206.44	160.96	1.88	1.46
B2751CU	141.11	147.53	268.92	247.37	278.63	257.09	2.29	2.00
B2754CU	139.59	132.67	228.83	191.80	282.57	260.36	1.98	1.63
B2872CU	153.70	172.47	188.25	157.98	192.73	147.49	1.85	1.40
B3251CU	142.67	150.79	237.21	203.88	251.30	215.30	2.11	1.79
B3253CU	168.88	174.55	215.59	201.66	211.37	194.12	2.07	1.83

Supplementary Table (Cont.)

Hybrid Name	Grain Yield (bu/acre) Full Dataset	Grain Yield (bu/acre) UIUC	ACN (mg/kg) Full Data	Log ACN (mg/kg) Full Dataset	ACN (mg/kg) UIUC	Log ACN (mg/kg) UIUC	ACN Yield (lbs/acre) Full Dataset	Log ACN Yield (lbs/acre) Full Dataset
B3652CU	157.73	173.93	155.96	136.94	178.88	156.51	1.70	1.48
B3701CU	142.15	149.38	236.73	214.08	252.00	224.63	2.12	1.87
B3731CU	137.64	125.22	169.85	139.61	196.38	150.33	1.62	1.12
B3761CU	134.99	122.43	183.52	138.11	210.70	152.35	1.54	0.94
B3792CU	137.40	142.18	196.62	166.02	227.92	201.11	1.94	1.59
B3804CU	140.04	139.42	197.26	171.72	205.63	165.42	1.70	1.27
B3851CU	149.27	134.26	209.00	171.93	193.70	143.05	1.45	0.99
B3884CU	138.58	147.29	218.93	199.12	222.66	193.36	1.95	1.64
B3901CU	135.29	128.85	198.25	167.23	209.37	169.70	1.70	1.26
B3931CU	151.32	158.34	196.85	184.93	193.77	177.10	1.84	1.63
B3952CU	141.41	147.02	205.41	191.35	209.60	190.13	1.82	1.66
B4024CU	152.64	174.30	249.23	214.66	296.28	280.68	2.96	2.70
B4114CU	146.71	128.61	211.49	177.53	237.19	206.73	1.77	1.39
B4221CU	148.97	166.21	254.10	240.47	259.01	250.88	2.43	2.31
B4251CU	165.79	181.76	431.40	401.33	399.17	404.89	4.18	4.02
B4361CU	135.44	137.48	218.12	187.53	253.60	240.78	2.00	1.75
B4423CU	133.99	125.67	203.16	177.12	236.75	210.56	1.72	1.35
B4451CU	148.65	157.41	254.69	213.59	315.25	301.11	2.78	2.59
B4481CU	145.05	164.30	166.18	136.57	169.77	123.55	1.57	1.15
B4482CU	153.25	166.53	146.91	130.23	155.50	130.39	1.43	1.21
B4542CU	143.07	158.26	184.52	168.29	189.50	166.59	1.73	1.46
B4545CU	146.56	156.98	196.28	168.38	210.21	170.19	1.90	1.48
B4632CU	165.57	197.14	144.72	131.26	139.18	121.20	1.50	1.33
B4634CU	163.33	186.78	141.81	122.65	109.06	90.65	1.12	1.02
B4666CU	155.65	179.05	134.07	114.53	146.50	124.30	1.32	1.25
B4691CU	145.92	141.51	170.92	137.85	171.26	125.06	1.43	0.91
M0371CU	132.78	102.03	151.40	119.11	162.80	113.42	1.14	0.58
M0422CU	150.41	132.13	141.45	104.01	158.78	101.92	1.32	0.73
M0625CU	143.65	146.10	NA	NA	NA	NA	NA	NA
M0781CU	139.85	141.28	220.72	186.35	225.07	201.33	1.79	1.50
M1655CU	138.24	137.46	147.07	114.01	166.04	117.56	1.37	0.87
M1672CU	144.25	144.43	174.81	132.27	209.37	150.85	1.80	1.28

Supplementary Table (Cont.)

Hybrid Name	Grain Yield (bu/acre) Full Dataset	Grain Yield (bu/acre) UIUC	ACN (mg/kg) Full Data	Log ACN (mg/kg) Full Dataset	ACN (mg/kg) UIUC	Log ACN (mg/kg) UIUC	ACN Yield (lbs/acre) Full Dataset	Log ACN Yield (lbs/acre) Full Dataset
M1791CU	160.52	161.97	150.96	131.03	156.23	132.46	1.39	1.24
M1884CU	150.09	160.79	113.98	97.54	120.41	98.58	1.08	0.95
M1896CU	156.14	167.95	170.02	156.26	170.85	149.83	1.58	1.45
M1922CU	144.48	151.89	120.18	107.17	129.07	105.15	1.11	0.87
M3052CU	146.02	161.14	185.61	151.97	190.63	143.10	1.69	1.27
M3104CU	132.92	132.26	135.68	93.95	156.28	92.00	1.33	0.69
P0321CU	156.42	175.14	398.15	285.07	465.20	408.23	5.04	3.66
P0601CU	159.32	183.15	327.04	313.07	335.02	331.04	3.69	3.18
P0881CU	151.36	168.98	319.34	225.61	342.48	248.40	3.44	2.31
P1141CU	141.78	140.09	313.40	247.04	335.29	283.70	2.44	1.96
P1231CU	144.71	148.99	343.31	299.49	439.11	503.77	3.69	3.89
P1401CU	153.52	169.96	396.76	334.39	429.50	380.44	4.36	3.53
P1884CU	145.61	154.23	183.79	164.06	207.00	182.18	1.70	1.56
P2711CU	142.00	143.77	425.19	244.05	524.97	312.70	4.89	2.40
P3411CU	156.02	175.65	320.41	274.05	313.01	273.75	3.52	2.77
P3437CU	141.77	140.14	234.15	201.97	298.43	294.03	2.41	2.17
P3591CU	144.93	149.65	233.96	185.69	248.69	195.44	2.04	1.56
P3649CU	NA	NA	264.32	232.58	281.72	258.42	NA	NA
P4361CU	133.04	114.13	257.43	215.01	274.25	236.88	1.56	1.44
P4451CU	137.70	128.17	203.21	185.59	214.03	196.38	1.31	1.33
P5349CU	134.29	116.96	198.68	150.64	208.73	146.28	1.56	0.61
P5649CU	135.34	120.15	193.94	155.18	201.71	149.72	1.64	0.84
P5746CU	129.88	104.64	181.98	136.52	188.51	125.44	NA	NA
PB021CU	145.56	151.79	206.52	187.69	218.27	200.02	1.81	1.68
PB025CU	144.87	149.72	236.61	207.63	251.46	226.59	2.23	1.91
PB029CU	138.81	131.48	288.66	237.53	307.99	269.07	2.23	1.94
PB033CU	145.87	152.72	220.87	197.07	234.01	211.39	2.23	1.88
PB037CU	152.88	173.81	249.65	212.75	265.48	234.59	2.87	2.27
PB041CU	155.14	180.60	279.51	233.95	298.11	264.39	3.53	2.59
PB045CU	141.00	138.08	227.24	202.03	241.19	218.83	1.87	1.69
PB073CU	155.34	181.22	426.02	301.93	459.28	366.47	5.66	3.58
PB077CU	153.77	176.50	218.04	196.08	230.93	212.00	2.49	2.08

223

Supplementary Table (Cont.)

Hybrid Name	Grain Yield (bu/acre) Full Dataset	Grain Yield (bu/acre) UIUC	ACN (mg/kg) Full Data	Log ACN (mg/kg) Full Dataset	ACN (mg/kg) UIUC	Log ACN (mg/kg) UIUC	ACN Yield (lbs/acre) Full Dataset	Log ACN Yield (lbs/acre) Full Dataset
PB081CU	145.82	152.58	161.06	155.21	169.39	157.67	1.35	1.33
PB085CU	151.48	169.61	206.77	187.07	219.12	200.43	2.06	1.80
PB089CU	151.11	168.45	287.38	228.91	310.02	262.84	3.17	2.37
PB093CU	146.22	153.80	170.21	164.20	177.89	168.06	1.45	1.41
PB097CU	151.75	170.42	168.46	163.14	176.59	167.35	1.70	1.61
PB101CU	152.08	171.41	382.65	280.06	411.70	334.89	4.75	2.87
PB105CU	141.91	140.82	333.23	259.86	356.69	303.32	2.67	2.02
PB109CU	144.17	147.34	340.08	259.93	365.09	304.57	3.28	2.39
PB113CU	150.79	167.47	336.83	251.06	366.01	297.92	3.75	2.65
PB117CU	153.46	175.57	258.27	218.22	274.68	241.34	3.04	2.32
PB129CU	147.49	157.59	293.33	238.31	314.27	272.74	2.98	2.23
PB135CU	161.15	198.69	330.29	256.94	353.56	297.34	5.09	3.23
PB141CU	158.79	191.58	287.53	233.67	306.96	264.32	3.86	2.79
PB151CU	168.04	219.42	259.11	220.26	276.09	244.25	4.31	3.00
PB260CU	143.39	145.26	289.64	235.81	309.87	269.08	2.49	2.04
PB262CU	148.85	161.69	298.65	240.97	319.72	276.59	3.08	2.36
PB264CU	130.75	107.24	352.50	263.67	379.00	309.38	1.82	1.62
PB266CU	147.84	158.65	267.39	223.30	285.59	249.68	2.75	2.21
PB268CU	129.77	104.32	237.98	207.62	252.40	227.99	1.02	1.16
PB270CU	146.14	153.54	235.54	207.30	249.73	227.54	2.14	1.89
PE003CU	136.48	123.97	335.17	256.67	362.93	303.07	2.23	1.93
PE004CU	137.22	126.72	279.75	228.11	299.34	257.32	1.92	1.70
PE005CU	132.44	112.32	338.52	261.59	363.75	307.07	1.97	1.70
PE008CU	131.44	109.34	312.10	247.45	333.65	283.45	1.61	1.55
PE009CU	132.39	111.53	319.70	244.53	344.84	282.90	1.65	1.51
PE010CU	130.26	105.79	242.72	211.48	257.69	232.11	1.20	1.27
PE013CU	132.57	112.74	281.23	233.32	300.12	264.48	1.62	1.44
PE014CU	137.07	126.25	261.83	220.22	278.58	244.15	1.82	1.61
PE123CU	142.73	143.28	404.23	295.23	434.80	356.80	3.65	2.76
PE129CU	151.75	170.42	378.00	280.16	406.26	331.70	4.59	3.11
PE141CU	157.14	186.64	343.02	263.73	368.68	310.25	4.54	3.14
PE151CU	130.94	107.82	363.56	273.49	390.45	321.69	1.98	1.78

Supplementary Table (Cont.)

Hybrid Name	Grain Yield (bu/acre) Full Dataset	Grain Yield (bu/acre) UIUC	ACN (mg/kg) Full Data	Log ACN (mg/kg) Full Dataset	ACN (mg/kg) UIUC	Log ACN (mg/kg) UIUC	ACN Yield (lbs/acre) Full Dataset	Log ACN Yield (lbs/acre) Full Dataset
PF262CU	134.68	127.11	107.60	95.82	120.12	98.30	0.76	0.71
PF304CU	148.01	157.44	229.85	209.00	220.65	199.50	1.94	1.69
PF313CU	144.79	150.75	204.14	185.23	216.25	197.93	1.84	1.64
PH071CU	135.65	129.32	306.10	265.84	324.74	298.57	2.29	2.07
PH154CU	138.01	134.69	332.02	260.21	353.21	293.21	2.63	2.05
PH155CU	134.59	126.92	174.17	167.87	185.64	174.96	1.35	1.18
PH761CU	138.82	135.42	382.05	339.98	397.62	382.99	3.01	2.83
PH7671CU	130.86	107.31	408.04	364.97	438.94	454.09	2.38	2.11
PJ321CU	136.99	125.54	412.34	293.28	443.10	349.56	2.79	2.24
PM013CU	141.64	140.00	222.56	196.25	236.73	212.50	1.76	1.60
PM017CU	139.99	135.04	182.16	173.40	191.37	180.31	1.36	1.38