CHARACTERIZATION OF ANTHOCYANIN PRODUCTION IN MAIZE GERMPLASM

BY

MICHAEL PAULSMEYER

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Master's Committee:

Professor John Juvik, Chair
Assistant Professor Patrick Brown
Adjunct Professor Leslie West

Abstract

Anthocyanins are the visually appealing red-orange to blue natural pigments present in most plant species. Because of their ease of aqueous extraction and rich color, anthocyanins make suitable natural replacements for synthetic dyes like FD&C Red 40. Increasing consumer demand for natural ingredients in foods and beverages has justified a search for more economic sources of natural colorants. Maize has been used a natural source of color for centuries. South American cultures in particular use purple corn varieties termed "*Maize Morado*" in foods and beverages. Maize has economic potential as a value-added source of natural colors. To characterize the diversity of anthocyanin production in maize germplasm, 398 diverse accessions of pigmented maize were analyzed with High Performance Liquid Chromatography (HPLC). To the best of our knowledge, this is the largest collection of pigmented maize investigated for anthocyanin diversity and composition. A subset of this collection was also used to test the repeatability of anthocyanin content and composition in several environments. Within the collection, 167 accessions could produce detectable amounts of anthocyanins. Clusters of accessions were created based on the abundance of pigments in the aleurone or pericarp layers and on compositional variations. These clusters were confirmed by principal component analysis and hierarchical clustering. Pericarp-pigmented accessions that could produce flavanol-anthocyanin dimers called "condensed forms" were the most important accessions in the survey in terms of total anthocyanin content. In the survey acylated anthocyanins were typically the most predominant pigments types, except in a few unique lines that only produce cyanidin 3-glucoside as their major pigment. The hypothesis tested was that this trait is due to partial loss of function anthocyanin acyltransferase. To characterize this phenotype, coined the "reduced acylation" trait, a mapping population was created from a mutant phenotype line crossed to B73,

the reference genome. Genotyping-by-sequencing was used to generate single nucleotide polymorphisms that could be used to map the location of the trait in the maize genome. QTL analysis found one significant loci at the end of chromosome 1 which corresponds to a candidate anthocyanin acyltransferase GRMZM2G387394. A UniformMu *Mu* transposon knockout of this gene maintained the phenotype when crossed to reduced acylation mutants meaning the gene is responsible for a majority of the anthocyanin acyltransferase activity in maize. Overall, information provided here will assist plant breeders looking to develop anthocyanin-rich purple corn hybrids as a source of natural colorants for food and beverages.

Acknowledgements

I am very grateful for all of the opportunities that I have had so far at the University of Illinois and I have quite a few people to thank for my great experiences. First of all, I would like to thank Kraft Heinz for sponsoring this project in the first place. I hope this project serves as a model for how industry collaboration with universities can work in the future. I am thankful that I got to work alongside two excellent scientists within Kraft Heinz: Dr. Megan West and Dr. Leslie West. The following work would not have even been possible without such driven and motivated scientists. I would also like to thank the Illinois Corn Growers Association for sponsoring my fellowship.

I have numerous people to thank here at the University of Illinois for their help in developing my research projects. First of all, I want to thank my committee: Dr. Jack Juvik, Dr. Pat Brown, and Dr. Les West, for their guidance. This research would not have moved forward without their encouragement. I would also like to thank all fellow graduate students in my lab, past and present: Laura Chatham, Alicia Gardner, Talon Becker, Kang-Mo Ku, and Justin Gifford. Thank you for being a good resource when I needed help with my project and thank you for being good listeners when I just needed someone to vent to. I am very grateful I got to work with you all.

Finally, I want to thank my family for being supportive of me all through grad school. Thank you especially to my parents for instilling a good work ethic in me from a young age. Thank you all again.

Table of Contents

Chapter 1: Literature Review

1.1 Anthocyanin Abundance and Function in Plants

Anthocyanins are the visually appealing pigments responsible for most of the red-orange to blue colors exhibited in nearly every plant species. Anthocyanins belong to the diverse class of secondary metabolites called “flavonoids”. Anthocyanins and the other flavonoids serve many useful functions in plants. The first anthocyanins were thought to have been synthesized 450 million years ago when the first plants colonized the planet (Campanella et al., 2014). The hypothesis is that they developed these compounds for protection. A major obstacle for plant survival, even today, is photoprotection. For this reason, the conservation of anthocyanin pigments among plant species is thought to be associated with UV radiation protection and free radical scavenging (Hatier and Gould, 2008). An obvious function for anthocyanins is their role as signaling compounds. Anthocyanins expressed in the floral organs signal to pollinators or seed dispersers that a reward is contained within (Koes et al., 1994). Moreover, bright anthocyanins can be used as a signal to warn herbivores of defenses like poison or thorns. Conversely, darker anthocyanins are thought to act as camouflage against backgrounds like shade or soil, and can mimic dead leaves to detract herbivores (Lev-Yadun and Gould, 2008). Anthocyanins are diverse compounds that enhance a plant’s fitness for the production of viable offspring.

1.2 Human Health Benefits

Besides their functions in plants, anthocyanins are known health-promoting compounds. Average daily intake of anthocyanins worldwide is thought to be as high as 180-215 mg per person (He and Giusti, 2010). In another estimation for the US consumer, average daily intake of anthocyanins based on the consumption of 100 common foods was estimated to be around 12.5 mg per person (Wu et al., 2006). Anthocyanins have a wide range of health benefits that have

been studied extensively and widely reviewed. To summarize, anthocyanins have been suggested to possess anti-inflammatory, anticarcinogenic, antiangiogenesis, antimicrobial, cardioprotective, and neuroprotective activity. They also have suggested roles in preventing obesity and diabetes as well as roles in improving eye health (Zafra-Stone et al., 2007; He and Giusti, 2010).

1.3 FD&C Red 40

With the ever-increasing demand for more natural ingredients in foods and beverages, sources of natural pigments are becoming increasingly important. Moreover, the published association of synthetic dyes and increased hyperactivity has generated consumer distrust of synthetic dyes (McCann et al., 2007). Because of their attractive color and ease of extraction, anthocyanins are suitable natural replacements for many synthetic dyes such as FD&C Red 40, also known as Allura Red. Red 40 is the most common color additive in the US market, comprising over 25% of all food and beverage additives. In total, over 2.8 million kg of Red 40 were certified for use in the US in 2015 (Center for Food Safety and Applied Nutrition, 2015). The European Food Safety Authority (EFSA) conducted a risk assessment to re-evaluate the current consumption of Red 40. The acceptable daily intake (ADI) of Red 40 according to the Joint FAO/WHO Expert committee on Food Additives (JECFA) and the EU Scientific Community for Food (SCF) is 7 mg per kg body weight per day. The EFSA survey reported that the average daily intake of Red 40 for adults (>18 years old) was 0.9 mg per kg body weight and for children 1 to 10 was 0.5 to 3.0 mg per kg body weight. When applying a 95% percentile to the children 1 to 10 population, they estimated maximum daily intake of Red 40 was 8.5 mg per kg, well over the ADI for Red 40 (EFSA Panel on Food Additives and Nutrient Sources Added to Food, 2009).

1.4 Anthocyanins as Natural Colorants

Many crops have been investigated for their potential use as commercial anthocyanin sources. Francis and Markakis (1989) provided a full review of these potential sources of anthocyanins. A few notable crops used today are blueberries, black carrots, and sweet potatoes (Somavat et al., 2016). Purple corn has also been getting attention lately as a source of anthocyanins because of its potential as value-added coproduct in the commercial corn supply chain (Tsuda et al., 2003; Li et al., 2008; Somavat et al., 2016). The use of purple corn is not a new idea; purple corn varieties have been utilized as a source of natural colors for centuries. Specifically, these varieties are important in Andean cultures. Open-pollinated varieties of purple corn in Peru are collectively referred to as "*Maize Morado*" and are used in foods, beverages, and even textiles (Petroni et al., 2014).

1.5 Anthocyanin Structure

Anthocyanins are a structurally diverse class of molecules that are subject to numerous modifications. Among all this diversity, all anthocyanins share the same 2-phenylbenzylpyrilium ion (Escribano-Bailón et al., 2004). This core ion is typically glycosylated and hydroxylated or methoxylated in various locations of the A-, B-, and C- rings (Figure 1.1). Main anthocyanin classes are divided by B-ring substitutions. In maize, only three types of anthocyanidins are possible: pelargonidin, cyanidin, and peonidin. Pelargonidin is the simplest anthocyanidin and only contains a hydroxyl group at the 4' position of the B-ring. Cyanidin is a derivative of pelargonidin with an additional hydroxylation at the 3' position. Peonidin is a further modification with methoxylation at the 3' position. Much of the anthocyanin diversity is within the glycone constituent. The glycone in maize is typically glucose, but attachment to other sugars such as arabinose, galactose, rhamnose, and rutinose have also been suggested in maize (Bhatla

and Pant, 1977; Grotewold et al., 1998; Abdel-Aal et al., 2006). The main site of glycosylation in maize is at the 3-position, but there is also evidence that attachment can occur at the 3- and 5-position together (González-Manzano et al., 2008; Li et al., 2008).

1.6 Location of Synthesis in the Maize Kernel

Maize has the capability of producing anthocyanins in nearly every tissue depending on which regulatory genes are present (Ludwig and Wessler, 1990). The main location of anthocyanin synthesis in the grain is within the pericarp and aleurone layers (Figure 1.2). The pericarp layer is the outmost layer of the mature corn kernel composed of nonliving maternal tissue. Beneath the pericarp is the aleurone layer, which is (typically) a single layer of triploid tissue around the peripheral of the endosperm. Maize is capable of producing multiple aleurone layers, but so far this trait has been found naturally in only a few accessions and induced mutants (Wolf et al., 1972; Becraft and Yi, 2011). Aleurone nuclei are comprised of one paternal and two maternal sets of chromosomes.

1.7 Structural Genes of the Maize Anthocyanin Biosynthesis Pathway

1.7.1 *C2*

A complete anthocyanin biosynthetic pathway with gene products is included in Figure 1.3. The first step in anthocyanin synthesis involves the combination of three malonyl-CoA molecules with *p*-coumaryl-CoA to form chalcone. This reaction is catalyzed by chalcone synthase (CHS). In maize, there are fourteen genes with CHS motifs within the genome that are expressed in multiple tissues (Han et al., 2016). However, the CHS gene responsible for anthocyanin synthesis in the kernel is *Colorless2* (*C2*). *C2* was cloned using *SPM/En* transposable element systems and showed strong homology to the previously identified parsley

CHS (Wienand et al., 1986). Another important member of the CHS family in maize is *White pollen1* (*Whp1*), a gene responsible for CHS activity in the pollen (Franken et al., 1991).

1.7.2 CHI

The next step in anthocyanin synthesis is the closure of the C-ring of chalcone to form naringenin. This process can happen spontaneously, albeit slowly, but enzymatically it is accomplished through chalcone isomerase (CHI). A CHI in maize has been sequenced and characterized using petunia CHI as a probe. In maize, no mutants of CHI have been found. It is theorized that this is due to multiple copies of CHI rescuing mutations (Grotewold and Peterson, 1994).

1.7.3 *Pr1*

It is important to note that the flavonoid pathway is not necessarily a linear gene cascade developed for the end goal of producing anthocyanins. There are many branch points due to common substrates among genes that lead to the formation of various flavonoids outside of anthocyanins. The product of CHI, naringenin, for example is a substrate for at least six distinct enzymes. This branch point can shunt portions of the pathway away from producing anthocyanin pigments into producing water insoluble red pigments called phlobaphenes (Springob et al., 2003). If the pathway at this point is specified for cyanidin or peonidin, then the P450 enzyme flavonoid 3'-hydroxylase (F3'H) converts naringenin to eriodictyol by hydroxylation of the 3'-position of the B-ring (Sharma et al., 2011). Eriodictyol can also be converted to phlobaphenes. In maize the F3'H is encoded at the *Purple aleurone1 (Pr1)* locus. Maize with recessive alleles of this gene accumulate more pelargonidin than cyanidin anthocyanins. Typically, lines homozygous dominant have a cyanidin to pelargonidin ratio of 9:1. In homozygous recessive lines, the ratio is around 1:10 (Larson et al., 1986). Phenotypically, *Pr1* recessive alleles color

kernels red or pink, while dominant alleles appear blue. The *Pr1* gene was characterized relatively recently using a putative F3'H sequences from sorghum and rice (Sharma et al., 2011).

1.7.4 *Fht1*

After the formation of flavanones with or without the hydroxylation at the 3' position, the gene flavanone 3-hydroxylase (F3H) hydroxylates the 3-position to form dihydroflavonols. There are two dihydroflavanols produced in maize: the pelargonidin precursor is dihydrokaempferol and the cyanidin/peonidin precursor is dihydroquercetin. Dihydrokaempferol is also a substrate for *Pr1* and can be converted to dihydroquercetin. Because of this, there are two genic levels that determine if cyanidin or peonidin anthocyanins are formed. In maize F3H is referred to as *flavanone 3-hydroxylase1* or *Fht1* (Cone, 2007). The sequence of this gene in maize was probed using snapdragon (*Antirrhinum majus* L.) F3H cDNA and confirmed with homology to other species. There are no known mutants to this gene (Deboo et al., 1995).

1.7.5 *A1*

The next step reduces dihydroflavonols to leucoanthocyanidins by converting the ketone at the 4-position of the C-ring to a hydroxyl group. The gene catalyzing this step is referred to as dihydroflavonol reductase (DFR), which is catalyzed by *anthocyaninless1* (*A1*) in maize (Reddy et al., 1987). This gene was first cloned using the transposable elements *Mu1* and *En* as gene tags and the elements *Mu1*, *Spm-I8,* and *En1* as molecular probes (O'Reilly et al., 1985). *A1* is important for the synthesis of phlobaphenes as well and is the gene responsible for flavanol conversion to flavan-4-ols (Mol et al., 1998).

1.7.6 *A2*

The Anthocyanidin Synthase (ANS) gene forms anthocyanidins from leucoanthocyanidins. In this step, the hydroxyl group at the 4-position is removed while the 3-

OH on the C-ring and the 5- and 7-OH on the A-ring remain (Wilmouth et al., 2002). The ANS gene in maize was found at the *Anthocyaninless2* (*A2*) locus. It was also cloned using the *Spm/EN* transposable element system (Menssen et al., 1990).

1.7.7 *Bz1*

After the formation of anthocyanidins, the next step is to conjugate a sugar molecule to form complete anthocyanins. Glycosylation occurs at the 3-position of the C-ring by the *Bronze1* (*Bz1*) gene in maize. *Bz1* is a uridine diphosphate (UDP) glucose-flavanol glucosyltransferase (Fedoroff et al., 1984). Without glycosylation, anthocyanins cannot be moved into the vacuoles. Within the cytoplasm, anthocyanins are instable. Mutants of *Bz1* produce their titular color due to the oxidation and condensation of anthocyanidins (Marrs and Walbot, 1997). *Bz1* was cloned using tagging with the *Ac* transposable element (Fedoroff et al., 1984).

1.7.8 *Bz2*

The final step of anthocyanin biosynthesis in maize is completed with the maize *Bronze2* (*Bz2*) gene. Non-functioning mutants of *Bz2* contain glucoside anthocyanins, but they are not localized in the vacuoles where they would be most stable. Not much was known about the function of this gene until 1995, when it was discovered to have glutathione S-transferase (GST) activity (Marrs et al., 1995). GSTs are detoxification enzymes, most notorious for their association with herbicide tolerance, but are more generally genes associated with stress response. GSTs tag toxic or endogenous compounds in plant cells with a glutathione tag. Vacuolar transport proteins then recognize these tags and sequester the compounds into the vacuole (Marrs, 1996). The *Bz2* gene was cloned using *Mu*-specific transposon tagging and differential hybridization (McLaughlin and Walbot, 1987).

1.8 Maize Anthocyanin Transcriptional Regulators

1.8.1 *R1*

One loci with the most allelic diversity in maize, with over 100 known variants, is the *R1* family of anthocyanin transcriptional regulators (Chandler et al., 1989). *R1* genes are transcriptional regulators responsible for the expression of pigments in various tissues at various growth stages. The standard *R1* allele, *R1-r:standard,* has two genic components that are separable by recombination. Component (S) for Seed color is responsible for aleurone pigmentation and is designated with *R1* for pigmented aleurones or *r1* for colorless. Component (P) for Plant color is responsible for anther, coleoptile, and first and second leaf sheath pigmentation and is designated *-r* or *-g* for *red* or *green*, respectively (Ludwig and Wessler, 1990). Two map units distal to the *R1-r* locus are two additional transcriptional regulators that are part of the *R1* gene family. These are the *Leaf color1* (*Lc1*) and *Scutellar node* (*Sn*) genes. *Lc1* confers pigmentation to the midrib, ligule, auricle, glume, lemma, palea, and pericarp tissues (Ludwig et al., 1989). *Sn* is a light-inducible gene that confers pigmentation to tissues of the plant not normally pigmented by *R1*: scutellar nodes, mesocotyls, leaf bases and midribs, glumes, and pericarps (Tonelli et al., 1991). The *R1* locus was first cloned using the *Ac*-transposon tagging system on the *R1-navajo* (*R1-nj*) allele. *R1-nj* confers a unique phenotype distinguished by aleurone pigmentation only on the crown of the kernel (Dellaporta et al., 1988). Using this clone as a probe, the first *R1* gene family member sequenced was *Lc1*. *Lc1* was discovered to have a basic Helix-Loop-Helix (bHLH) DNA binding domain and homology to *myc*-like transcriptional regulators. Based on Southern blot analysis from the sequencing study and previous research, *Lc1*, *R1*-(S), *R1*-(P), and *R-nj* were all found to be homologous genes that

encode similar regulatory proteins (Ludwig et al., 1989). It is hypothesized that the homology of all *R1* gene family members is due to duplication of the *R1-r* loci (Ludwig and Wessler, 1990).

1.8.2 *B1*

While standard *R1* alleles are most often associated with floral tissues, another regulatory gene, *Booster1* (*B1*), is most often associated with pigmenting vegetative tissues. Direct evidence that *B1* works in vegetative tissues came when monitoring *Bz1* activity in backgrounds containing dominant or recessive *B1* alleles in leaf tissues. A dominant *B1* allele was able to induce flavonoid production in the leaf, while recessive *b1* and standard *R1* alleles could not (Gerats et al., 1984). Just like *R1*, *B1* is also a diverse locus with many alleles. Two alleles of *B1*, *B1-Peru* and *B1-Bolivia*, were found to substitute *R1* activity and pigment the aleurone. The ability of *B1* to substitute *R1* activity provided evidence that *B1* might be a functional duplicate of *R1*. To test this, *R1* DNA probes were used to clone the *B1* sequence (Chandler et al., 1989). *B1-Peru* was the first *B1* sequence cloned and was found to share 95% identity with the bHLH domains found in *R1,* which are related to *myc*-like transcriptional regulators. Furthermore, *B1-Peru* shared 78% identity to the *R*-gene family member *Lc1* (Radicella et al., 1991).

1.8.3 *C1*

Early studies in maize anthocyanin synthesis discovered another factor that was required for pigmentation in the aleurone. This gene was termed *Colorless1* (*C1*) because aleurone pigmentation is absent in lines containing recessive alleles. Later it was found that *C1* must be present in combination with *R1* in order for aleurone pigments to be expressed. The regulatory role of *C1* was determined by several independent pieces of evidence, first summarized by Chen and Coe Jr. (1977): 1) Null alleles of *C1* have not been shown to produce flavonoids in the aleurone, but are able to produce flavonoids in other parts of the plant; 2) *C1* has several alleles

with different expression states: null, inhibiting, conditional, and dominant; 3) finally, different alleles are inducible under different developmental and environmental conditions, suggesting regulatory functions, rather than structural functions. Direct evidence that *C1* was a regulatory gene was accomplished by Paz-Ares et al. (1987) when a transposon tagged *C1* locus was cloned and sequenced. They found that the *C1* sequence has an acidic C-terminal domain with homology to *MYB* transcriptional regulators in animals. Goff et al. (1991) found this acidic domain is truncated in inhibitor alleles of *C1* (*C1-I*) but maintains DNA-binding function, making it a potent inhibitor of anthocyanin synthesis.

1.8.4 *Pl1*

Another regulatory gene involved with anthocyanins synthesis is *Purple leaf1* (*Pl1*). *Pl1* was found to be responsible for pigmentation in vegetative and floral organs and was most often associated with *B1* in producing color. *Pl1* has several expression states similar to *C1*. One expression state is referred to as "sun-red" because it produces vegetative pigments only under the influence of direct sunlight. As more became known about regulatory loci in anthocyanin synthesis, it became increasingly obvious that *Pl1* might have a similar function to other anthocyanin transcriptional regulators. Since *B1* is functionally a copy of *R1*, then it would stand to reason that *Pl1* might be a copy of *C1*. Cone et al. (1993) tested this hypothesis by using a *C1* DNA probe to find similar sequences. Using this technique *Pl1* was successfully sequenced and found to be 80% similar to *C1*. The origin of *Pl1* is thought to be due to a translocation and duplication event in the ancestors of modern maize.

1.8.5 *P1*

Anthocyanins are not commonly found in the ears of today's commercial hybrids. One flavonoid pigment is though, and that is phlobaphenes. Phlobaphenes are a divergence from the

anthocyanin pathway that produces non-water soluble, red pigments in the pericarp, cob and tassel glumes, and husks (Sharma et al., 2012). The difference is that phlobaphenes are comprised of polymerized flavan-4-ols, while anthocyanins are modified flavan-3,4-diols (Grotewold et al., 1991). The divergence in the pathway occurs before F3H can convert flavanols to dihydroflavonols. The locus responsible for regulating phlobaphene synthesis is *Pericarp color1* or *P1*. *P1* is described with a two letter-suffix that informs whether phlobaphenes are expressed in the pericarp or cob. For example, *P1-ww* refers to white/colorless pericarp and white cobs, while *P1-rr* refers to red pericarp and red cobs. *P1* gene cloning was accomplished using variegated alleles of *P1* that have natural *Ac*-insertions. *P1* was found to be another *MYB*-like transcription factor with slight homology to *C1*. There is no bHLH family gene like *R1/B1* associated with *P1*, so *P1* may combine both transcriptional activities (Grotewold et al., 1991).

1.8.6 *Pac1*

In a mutagenesis screen for new genes involved with anthocyanin pigmentation, a new transcriptional regulator of anthocyanin synthesis was found. This new gene is referred to as *Pale aleurone color1* (*Pac1*) because of its titular phenotype (Selinger and Chandler, 1999). *Mu1* transposon tagging was used to sequence *Pac1*. The gene was found to be homologous to WD40 repeat proteins that are known to be involved with anthocyanin regulation in *Arabidopsis* and petunia (Carey et al., 2004).

1.8.7 *In1*

In early studies of maize anthocyanins, A.C. Fraser (1924) discovered a few ears from Emerson's aleurone color collection that were segregating for red versus dark and purple versus black kernels in a Mendelian ratio of 1:3. Linkage analysis concluded this trait was due to a new factor separate from any known genes in the anthocyanin pathway at the time. The new factor

was termed *Intensifier1* or *In1* because of its apparent ability to intensify aleurone color when homozygous recessive. The effect of the *In1* gene is most apparent in *pr1* backgrounds. Lines with dominant *In1* alleles appear pink or red, while *in1* lines produce dark colored kernels similar to *Pr1* lines. The *In1* gene was cloned by Burr et al. (1996) using the *En/SPM* transposon tagging system. They found that *In1* shares homology to the *R1/B1* gene family of bHLH transcription factors. The method by which *In1* represses anthocyanin synthesis may be due to competitive binding at *R1* sites, or by the formation of HLH heterodimers.

1.8.8 *A3*

Another recessive intensifier was found near the *A1* gene and was named *Anthocyanin3* (*A3*). In combination with *B1* and *Pl1*, recessive *a3* alleles are shown to enhance pigment production in the plant. Dominant *A3* alleles have been shown to lower transcript levels of *B1* and structural genes involved in anthocyanin biosynthesis (Lauter et al., 2004). Currently, there is only genetic map information available for this gene and no sequence information.

1.9 Anthocyanin Modifications

1.9.1 Acylation

The addition of an acyl group to the glycoside of an anthocyanin is a common modification to anthocyanins, especially in maize. This process is referred to as acylation. There are two types of acylation: aromatic and aliphatic. Aromatic acylation is the addition of acyl groups containing aromatic rings. Examples are sinapoyl or *p*-coumaroyl anthocyanins. Aliphatic acylation adds aliphatic groups like malonyl, succinyl, or acetyl groups (Bakowska-Barczak, 2005). The genes responsible for attaching these acyl groups are the diverse family of genes referred to as acyltransferases. Acyltransferases are in the BAHD superfamily, named after the first four enzymes discovered: Benzylalcohol *O*-acetyltransferase, Anthocyanin *O*-

hydroxycinnamoyltransferase, anthranilate *N*-Hydroxycinnamoyl/benzoyltransferase, and Deacetylvindoline 4-*O*-acetyltransferase (D'Auria, 2006). BAHD members are functionally diverse. The minimum level of homology within the superfamily is 25-34% identity (St-Pierre and De Luca, 2000). Despite this wide diversity, there are two motifs that appear to be conserved among all members: HXXXDG (motif 1) near the N-terminal end and DFGWGKP (motif 2) near the C-terminal end (St-Pierre et al., 1998). These motifs are not strictly conserved, however. Yu et al. (2009) found variations in motif 2 in black cottonwood (*Populus trichocarpa* L.). Anthocyanin acyltransferases appear to have a specific motif themselves. Sequence YFGNC (motif 3) appears in almost every anthocyanin acyltransferase regardless of whether it is aliphatic or aromatic (Suzuki et al., 2001). Motif 3 is generally used as a probe to discover anthocyanin acyltransferases since motif 1 and 2 are conserved among the superfamily (Suzuki et al., 2002). In maize, malonyl groups are most common and produce malonylglucosides and dimalonylglucosides. Succinyl and caffeoyl groups have also been reported as acyl groups in maize (Li et al., 2008; Salinas-Moreno et al., 2012). Currently, the anthocyanin acyltransferase in maize has not been characterized (See Chapter 3).

1.9.2 Methylation

Methylation is a common modification to secondary plant metabolites. Methylation is driven by a diverse family of genes known as methyltransferases. Many economically important plant products are modified by methyltransferases. For example, both morphine and caffeine biosynthetic pathways require three methyltransferases (Ashihara et al., 2008; Galanie et al., 2015). There are several classes of methyltransferases that are distinguished by their substrate specificity (Poulton, 1981). Type I methyltransferases methylate hydroxyl groups of phenylpropanoids including caffeic acid, flavonoids, coumarins, and alkaloids. Type II

methyltransferases are coenzymeA phenylpropanoid derivatives important for lignin synthesis (Hugueney et al., 2009). *Brown midrib* genes (*Bm1-Bm4*) in maize are Type II methyltransferases involved with lignin synthesis that have large effects on forage quality and grain yield (Chen et al., 2014). With the characterization of the *Paeonia suffruticosa* Andrews and *Paeonia tenuifolia* L. anthocyanin methyltransferases, it was found that anthocyanin methyltransferases share a lineage to Type II methyltransferases (Du et al., 2015). The maize anthocyanin methyltransferase has not been mapped or characterized to date. Possible reasons it has not been found because a lack of known mutants, high homology between methyltransferase sequences, or several multiple functional copies of anthocyanin methyltransferases.

1.9.3 Flavanol-Anthocyanins

Flavanol-anthocyanins (F-As) are formed by the condensation of flavanols like catechin and epicatechin with an anthocyanin molecule, usually at the 4- to the 8-position of the flavanol to the anthocyanin. Vivar-Quintana et al. (1999) were the first to experimentally detect these pigments. They detected malvidin 3-glucoside conjugated to a catechin through LC-MS in aged red wines. Initially, the formation of these pigments was thought to be a product of fermentation. Instead, several other studies have reported detecting these pigments in diverse plant species, including strawberries, black currants, and maize. The most abundant F-A in maize is catechin-4,8-cyanidin-3,5-diglucoside. Interestingly, the anthocyanin cyanidin 3,5-diglucoside is only detected in trace amounts while the monoglucoside is fairly abundant (González-Manzano et al., 2008). The genetics underlying F-A condensation and associated glycosylation are unknown.

1.10 Molecular Marker Systems

Since the beginning of the understanding of genetics, discovering the location of traits within a genome has always revolved around the concept of Linkage Disequilibrium (LD). LD is

the principle that genes or molecular markers physically closer to each other on chromosomes will tend to be inherited together. Modern studies use LD to correlate DNA sequence information with phenotypic information in an attempt to find molecular markers associated with the trait of interest. The underlying principles to discover trait associations are still the same today, the only difference is the technology with which molecular markers are developed.

The first molecular marker technology to be widely utilized was Restriction Fragment Length Polymorphisms (RFLPs). These molecular markers take advantage of differences in fragment sizes due to polymorphisms at restriction enzyme cut sites. The fragments of a specific sequence are then hybridized to fluorescent probes to view the differences in fragment size (Botstein et al., 1980). RFLP markers were important for establishing genetic linkage maps for simple traits, but were inefficient due to the limited number of probes available and the time it took to scan the probes (He et al., 2014). With the advent of PCR, new, more efficient molecular marker technologies were developed. The most notable examples of PCR-based molecular markers were random amplifications of polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), sequence characterized amplified regions (SCARs), cleaved amplified polymorphic sequences (CAPS), and transposable element based molecular markers (for a review of PCR-based markers, see Agarwal et al., 2008).

The development of so-called “Next Generation Sequencing” (NGS) once again revolutionized the way molecular markers were developed. Genome sequencing projects like Human Genome Project in the 1990s brought the demand for low cost, high-throughput genomic sequencing. Previously, Sanger Sequencing was the most efficient sequencing technology. Before the introduction of NGS technologies, sanger sequencing could produce 96 reads in parallel with fragment sizes up to 750 base pairs (Schuster, 2007). Today, NGS technologies

utilize massively parallel sequencing and imaging technologies to produce hundreds of millions or even hundreds of billions of reads in a single sequencing run (He et al., 2014). NGS driving down sequencing costs allowed for single nucleotide polymorphisms (SNPs) to become the most prevalent molecular marker in use today.

NGS can produce molecular markers in several ways. For many species, whole genome sequencing (WGS) of several specimens of interest is the most straightforward way to discover SNPs. With WGS, the genome of each sample is randomly sheared into fragments small enough for efficient reads and attached to sequencing primers compatible with NGS machines. For highly complex genomes like maize and barley, genome re-sequencing is not as simple. For example, the maize genome is large and full of presence/absence variations, repetitive regions, and transposon-mediated rearrangements that increase the complexity for sequencing (Elshire et al., 2011). Instead of whole genome re-sequencing, Baird et al. (2008) developed a method to reduce the complexity of the genome with restriction-site associated DNA sequencing (RADseq). In this method, whole genomic DNA is fragmented with restriction enzymes and attached to Solexa sequencing adapters (Illumina, Inc., San Diego, CA, USA) fitted with sample-specific barcodes. The barcodes are usually 4 to 8 base pairs in length and are useful for distinguishing individuals when samples are pooled in a sequencing machine. Restriction enzyme digested fragments are then size selected by gel electrophoresis and ligated to a common reverse Solexa adapter (Baird et al., 2008). The advantage of RAD sequencing is that restriction enzyme sites are widely and randomly distributed across the genome and restriction enzymes can be methylation-sensitive, so repetitive regions are ignored (Elshire et al., 2011).

Elshire et al. (2011) modified the RAD sequencing protocol to make it less complicated and less expensive than before. This new method is called Genotyping-By-Sequencing (GBS).

Where the two mainly differ is after library construction. GBS skips the size selection and gel purification steps to reduce handling time and complexity significantly. The protocol also combines many reactions in a single well to reduce error and handling. Finally, the barcode sequences were designed more efficiently so that now 384 samples can be multiplexed in one lane.

1.11 Figures

Figure 1.1: Structure of an anthocyanidin molecule with molecular nomenclature.

R = H, Pelargonidin; R = OH, Cyanidin; R = OCH_3, Peonidin.

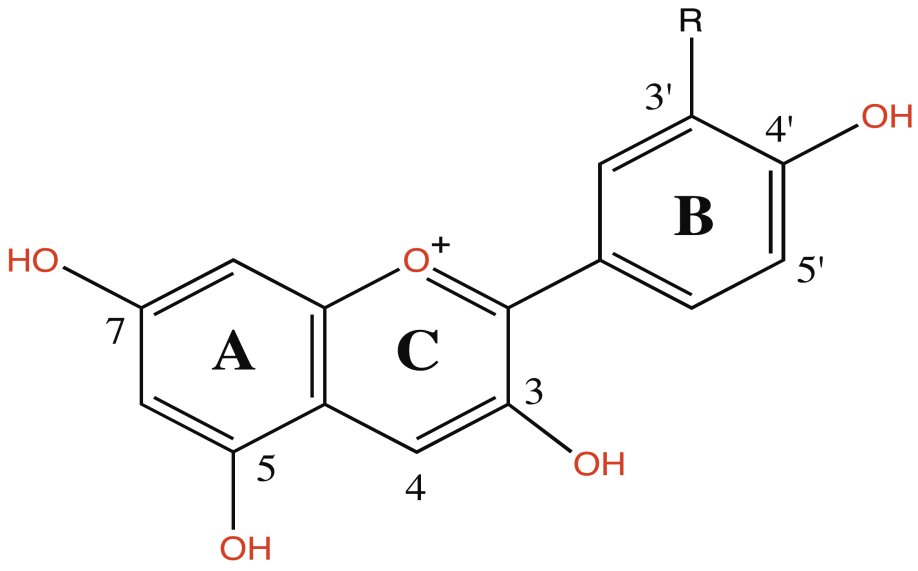


Figure 1.2: Major anthocyanin-producing tissues in the maize kernel. *Left*: Cross-section of a maize kernel. Figure adapted from Coe (2001). *Right*: Microscopic sections of kernels demonstrating anthocyanin pigmentation in various kernel tissues. Black bars represent 100 microns (Photo Credit, L. Chatham, University of Illinois Urbana-Champaign).

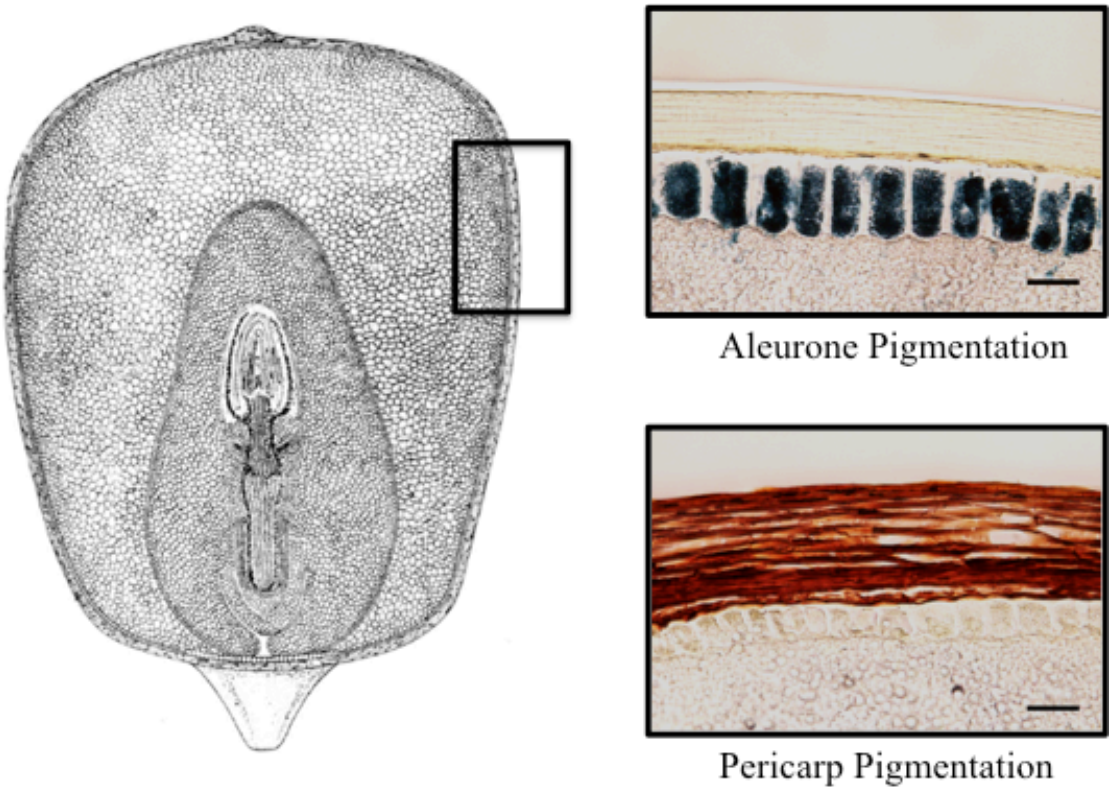
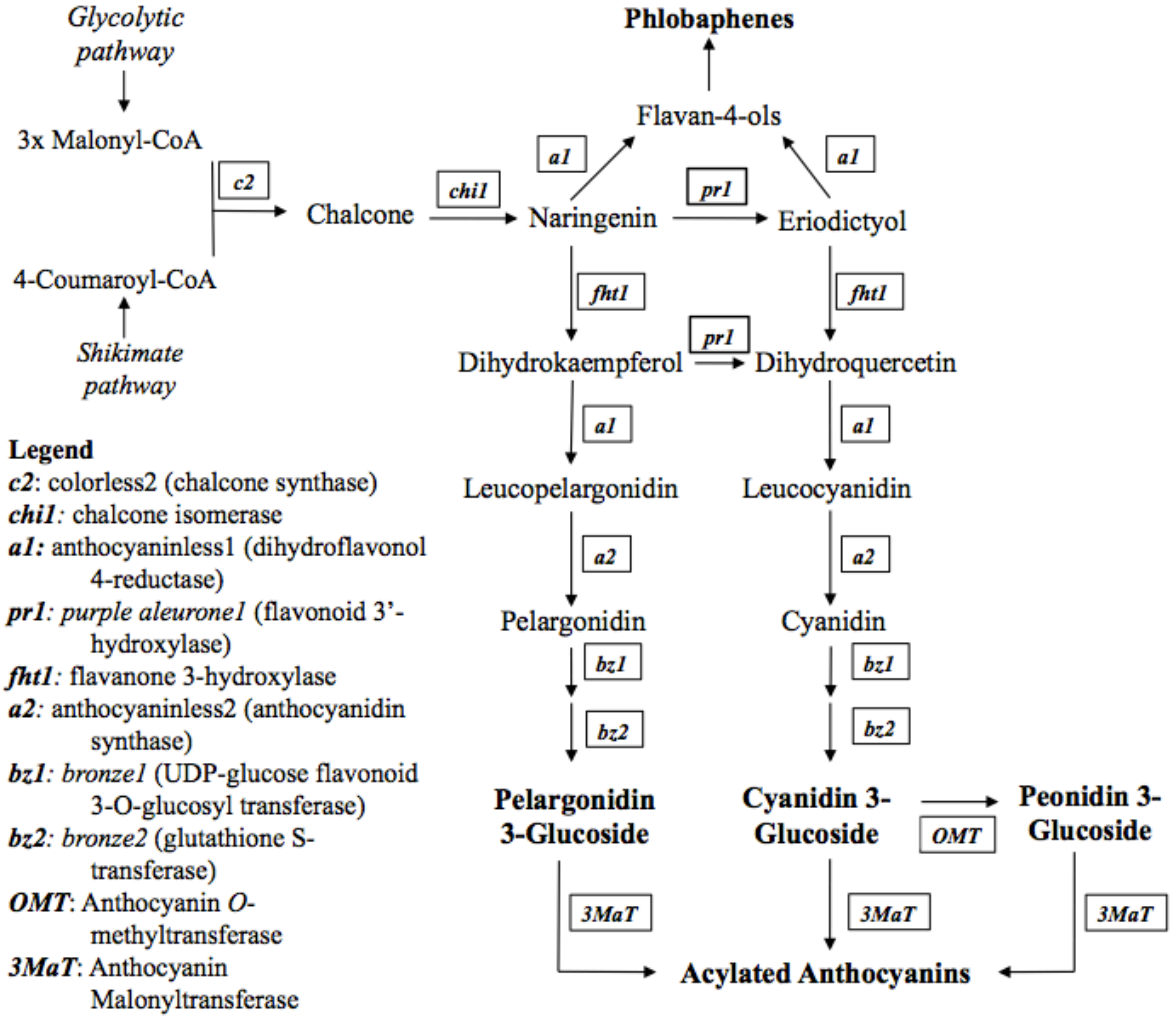


Figure 1.3: Anthocyanin biosynthetic pathway in maize with gene products.



1.12 References

Abdel-Aal, E.-S. M.; Young, J. C.; Rabalski, I. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J. Agric. Food Chem.* **2006**, *54* (13), 4696–4704.

Agarwal, M.; Shrivastava, N.; Padh, H. Advances in Molecular Marker Techniques and Their Applications in Plant Sciences. *Plant Cell Rep.* **2008**, *27* (4), 617–631.

Ashihara, H.; Sano, H.; Crozier, A. Caffeine and Related Purine Alkaloids: Biosynthesis, Catabolism, Function and Genetic Engineering. *Phytochemistry* **2008**, *69* (4), 841–856.

Baird, N. A.; Etter, P. D.; Atwood, T. S.; Currey, M. C.; Shiver, A. L.; Lewis, Z. A.; Selker, E. U.; Cresko, W. A.; Johnson, E. A. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **2008**, *3* (10), e3376.

Bakowska-Barczak, A. Acylated Anthocyanins as Stable, Natural Food Colorants - a Review. *Pol. J. Food Nutr. Sci.* **2005**, *14/15* (2), 107–116.

Becraft, P. W.; Yi, G. Regulation of Aleurone Development in Cereal Grains. *J. Exp. Bot.* **2011**, *62* (5), 1669–1675.

Bhatla, S.; Pant, R. Isolation and Characterisation of Anthocyanin Pigment from Phosphorus-Deficient Maize Plants. *Curr. Sci.* **1977**, *46*, 700–702.

Botstein, D.; White, R. L.; Skolnick, M.; Davis, R. W. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* **1980**, *32* (3), 314.

Burr, F. A.; Burr, B.; Scheffler, B. E.; Blewitt, M.; Wienand, U.; Matz, E. C. The Maize Repressor-like Gene Intensifier1 Shares Homology with the R1/B1 Multigene Family of Transcription Factors and Exhibits Missplicing. *Plant Cell* **1996**, *8* (8), 1249–1259.

Campanella, J. J.; Smalley, J. V.; Dempsey, M. E. A Phylogenetic Examination of the Primary Anthocyanin Production Pathway of the Plantae. *Bot. Stud.* **2014**, *55* (1), 10.

Carey, C. C.; Strahle, J. T.; Selinger, D. A.; Chandler, V. L. Mutations in the Pale Aleurone Color1 Regulatory Gene of the Zea Mays Anthocyanin Pathway Have Distinct Phenotypes Relative to the Functionally Similar Transparent Testa Glabra1 Gene in Arabidopsis Thaliana. *Plant Cell* **2004**, *16* (2), 450–464.

Center for Food Safety and Applied Nutrition. Color Certification Reports - Report on the Certification of Color Additives: 4th Quarter, Fiscal Year 2014, July 1-September 30 http://www.fda.gov/ForIndustry/ColorAdditives/ColorCertification/ColorCertificationReports/ucm418381.htm (accessed Sep 5, 2016).

Chandler, V. L.; Radicella, J. P.; Robbins, T. P.; Chen, J.; Turks, D. Two Regulatory Genes of the Maize Anthocyanin Pathway Are Homologous: Isolation of B Utilizing R Genomic Sequences. *Plant Cell* **1989**, *1* (12), 1175–1183.

Chen, S.-M.; Coe Jr, E. Control of Anthocyanin Synthesis by the C Locus in Maize. *Biochem. Genet.* **1977**, *15* (3–4), 333–346.

Chen, Y.; Blanco, M.; Ji, Q.; Frei, U. K.; Lübberstedt, T. Extensive Genetic Diversity and Low Linkage Disequilibrium within the COMT Locus in Maize Exotic Populations. *Plant Sci.* **2014**, *221–222*, 69–80.

Cone, K. C. Anthocyanin Synthesis in Maize Aleurone Tissue. In *Endosperm*; Olsen, O.-A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; Vol. 8, pp 121–139.

Cone, K. C.; Cocciolone, S. M.; Burr, F. A.; Burr, B. Maize Anthocyanin Regulatory Gene Pl Is a Duplicate of C1 That Functions in the Plant. *Plant Cell* **1993**, *5* (12), 1795–1805.

D'Auria, J. C. Acyltransferases in Plants: A Good Time to Be BAHD. *Curr. Opin. Plant Biol.* **2006**, *9* (3), 331–340.

Deboo, G. B.; Albertsen, M. C.; Taylor, L. P. Flavanone 3-Hydroxylase Transcripts and Flavonol Accumulation Are Temporally Coordinate in Maize Anthers. *Plant J. Cell Mol. Biol.* **1995**, *7* (5), 703–713.

Dellaporta, S. L.; Greenblatt, I.; Kermicle, J. L.; Hicks, J. B.; Wessler, S. R. Molecular Cloning of the Maize R-Nj Allele by Transposon Tagging with Ac. In *Chromosome structure and function*; Springer, 1988; pp 263–282.

Du, H.; Wu, J.; Ji, K.-X.; Zeng, Q.-Y.; Bhuiya, M.-W.; Su, S.; Shu, Q.-Y.; Ren, H.-X.; Liu, Z.-A.; Wang, L.-S. Methylation Mediated by an Anthocyanin, *O* -Methyltransferase, Is Involved in Purple Flower Coloration in *Paeonia*. *J. Exp. Bot.* **2015**, *66* (21), 6563–6577.

EFSA Panel on Food Additives and Nutrient Sources Added to Food. Scientific Opinion on the Re-Evaluation of Allura Red AC (E 129) as a Food Additive: Re-Evaluation of Allura Red AC (E 129) as a Food Additive. *EFSA J.* **2009**, *7* (11), 1327.

Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S.; Mitchell, S. E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6* (5), e19379.

Escribano-Bailón, M. T.; Santos-Buelga, C.; Rivas-Gonzalo, J. C. Anthocyanins in Cereals. *J. Chromatogr. A* **2004**, *1054* (1–2), 129–141.

Fedoroff, N. V.; Furtek, D. B.; Nelson, O. E. Cloning of the Bronze Locus in Maize by a Simple and Generalizable Procedure Using the Transposable Controlling Element Activator (Ac). *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81* (12), 3825–3829.

Francis, F. J.; Markakis, P. C. Food Colorants: Anthocyanins. *Crit. Rev. Food Sci. Nutr.* **1989**, *28* (4), 273–314.

Franken, P.; Niesbach-Klösgen, U.; Weydemann, U.; Maréchal-Drouard, L.; Saedler, H.; Wienand, U. The Duplicated Chalcone Synthase Genes C2 and Whp (White Pollen) of Zea Mays Are Independently Regulated; Evidence for Translational Control of Whp Expression by the Anthocyanin Intensifying Gene in. *EMBO J.* **1991**, *10* (9), 2605–2612.

Fraser, A. Heritable Characters of Maize: XVII—Intensified Red and Purple Aleurone Color. *J. Hered.* **1924**, *15* (3), 119–124.

Galanie, S.; Thodey, K.; Trenchard, I. J.; Interrante, M. F.; Smolke, C. D. Complete Biosynthesis of Opioids in Yeast. *Science* **2015**, *349* (6252), 1095–1100.

Gerats, A.; Bussard, J.; Coe Jr, E.; Larson, R. Influence of B and Pl on UDPG: Flavonoid-3-O-Glucosyltransferase in Zea Mays L. *Biochem. Genet.* **1984**, *22* (11–12), 1161–1169.

Goff, S. A.; Cone, K. C.; Fromm, M. E. Identification of Functional Domains in the Maize Transcriptional Activator C1: Comparison of Wild-Type and Dominant Inhibitor Proteins. *Genes Dev.* **1991**, *5* (2), 298–309.

González-Manzano, S.; Pérez-Alonso, J. J.; Salinas-Moreno, Y.; Mateus, N.; Silva, A. M. S.; de Freitas, V.; Santos-Buelga, C. Flavanol–anthocyanin Pigments in Corn: NMR Characterisation and Presence in Different Purple Corn Varieties. *J. Food Compos. Anal.* **2008**, *21* (7), 521–526.

Grotewold, E.; Peterson, T. Isolation and Characterization of a Maize Gene Encoding Chalcone Flavonone Isomerase. *Mol. Gen. Genet. MGG* **1994**, *242* (1), 1–8.

Grotewold, E.; Athma, P.; Peterson, T. Alternatively Spliced Products of the Maize P Gene Encode Proteins with Homology to the DNA-Binding Domain of Myb-like Transcription Factors. *Proc. Natl. Acad. Sci.* **1991**, *88* (11), 4587–4591.

Grotewold, E.; Chamberlin, M.; Snook, M.; Siame, B.; Butler, L.; Swenson, J.; Maddock, S.; St Clair, G.; Bowen, B. Engineering Secondary Metabolism in Maize Cells by Ectopic Expression of Transcription Factors. *Plant Cell* **1998**, *10* (5), 721–740.

Han, Y.; Ding, T.; Su, B.; Jiang, H. Genome-Wide Identification, Characterization and Expression Analysis of the Chalcone Synthase Family in Maize. *Int. J. Mol. Sci.* **2016**, *17* (2), 161.

Hatier, J.-H. B.; Gould, K. S. Anthocyanin Function in Vegetative Organs. In *Anthocyanins*; Winefield, C., Davies, K., Gould, K., Eds.; Springer New York: New York, NY, 2008; pp 1–19.

He, J.; Giusti, M. M. Anthocyanins: Natural Colorants with Health-Promoting Properties. *Annu. Rev. Food Sci. Technol.* **2010**, *1*, 163–187.

He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.; Li, Z. Genotyping-by-Sequencing (GBS), an Ultimate Marker-Assisted Selection (MAS) Tool to Accelerate Plant Breeding. *Front. Plant Sci.* **2014**, *5*.

Hugueney, P.; Provenzano, S.; Verries, C.; Ferrandino, A.; Meudec, E.; Batelli, G.; Merdinoglu, D.; Cheynier, V.; Schubert, A.; Ageorges, A. A Novel Cation-Dependent O-Methyltransferase Involved in Anthocyanin Methylation in Grapevine. *Plant Physiol.* **2009**, *150* (4), 2057–2070.

Koes, R. E.; Quattrocchio, F.; Mol, J. N. The Flavonoid Biosynthetic Pathway in Plants: Function and Evolution. *BioEssays* **1994**, *16* (2), 123–132.

Larson, R.; Bussard, J.; Coe Jr, E. Gene-Dependent Flavonoid 3′-Hydroxylation in Maize. *Biochem. Genet.* **1986**, *24* (7–8), 615–624.

Lauter, N.; Gustus, C.; Westerbergh, A.; Doebly, J. The Inheritance and Evolution of Leaf Pigmentation and Pubescence in Teosinte. *Genetics* **2004**, *167* (4), 1949–1959.

Lev-Yadun, S.; Gould, K. S. Role of Anthocyanins in Plant Defence. In *Anthocyanins*; Winefield, C., Davies, K., Gould, K., Eds.; Springer New York: New York, NY, 2008; pp 22–28.

Li, C.-Y.; Kim, H.-W.; Won, S.; Min, H.-K.; Park, K.-J.; Park, J.-Y.; Ahn, M.-S.; Rhee, H.-I. Corn Husk as a Potential Source of Anthocyanins. *J. Agric. Food Chem.* **2008**, *56* (23), 11413–11416.

Ludwig, S. R.; Wessler, S. R. Maize R Gene Family: Tissue-Specific Helix-Loop-Helix Proteins. *Cell* **1990**, *62* (5), 849–851.

Ludwig, S. R.; Habera, L. F.; Dellaporta, S. L.; Wessler, S. R. Lc, a Member of the Maize R Gene Family Responsible for Tissue-Specific Anthocyanin Production, Encodes a Protein Similar to Transcriptional Activators and Contains the Myc-Homology Region. *Proc. Natl. Acad. Sci.* **1989**, *86* (18), 7092–7096.

Marrs, K. A. The Functions and Regulation of Glutathione S-Transferases in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 127–158.

Marrs, K. A.; Walbot, V. Expression and RNA Splicing of the Maize Glutathione S-Transferase Bronze2 Gene Is Regulated by Cadmium and Other Stresses. *Plant Physiol.* **1997**, *113* (1), 93–102.

Marrs, K. A.; Alfenito, M. R.; Lloyd, A. M.; Walbot, V. A Glutathione S-Transferase Involved in Vacuolar Transfer Encoded by the Maize Gene Bronze-2. **1995**.

McCann, D.; Barrett, A.; Cooper, A.; Crumpler, D.; Dalen, L.; Grimshaw, K.; Kitchin, E.; Lok, K.; Porteous, L.; Prince, E.; et al. Food Additives and Hyperactive Behaviour in 3-Year-Old and 8/9-Year-Old Children in the Community: A Randomised, Double-Blinded, Placebo-Controlled Trial. *The Lancet* **2007**, *370* (9598), 1560–1567.

McLaughlin, M.; Walbot, V. Cloning of a Mutable Bz2 Allele of Maize by Transposon Tagging and Differential Hybridization. *Genetics* **1987**, *117* (4), 771–776.

Menssen, A.; Höhmann, S.; Martin, W.; Schnable, P. S.; Peterson, P. A.; Saedler, H.; Gierl, A. The En/Spm Transposable Element of Zea Mays Contains Splice Sites at the Termini Generating a Novel Intron from a dSpm Element in the A2 Gene. *EMBO J.* **1990**, *9* (10), 3051–3057.

Mol, J.; Grotewold, E.; Koes, R. How Genes Paint Flowers and Seeds. *Trends Plant Sci.* **1998**, *3* (6), 212–217.

O'Reilly, C.; Shepherd, N. S.; Pereira, A.; Schwarz-Sommer, Z.; Bertram, I.; Robertson, D. S.; Peterson, P. A.; Saedler, H. Molecular Cloning of the A1 Locus of Zea Mays Using the Transposable Elements En and Mu1. *EMBO J.* **1985**, *4* (4), 877–882.

Paz-Ares, J.; Ghosal, D.; Wienand, U.; Peterson, P. A.; Saedler, H. The Regulatory c1 Locus of Zea Mays Encodes a Protein with Homology to Myb Proto-Oncogene Products and with Structural Similarities to Transcriptional Activators. *EMBO J.* **1987**, *6* (12), 3553–3558.

Petroni, K.; Pilu, R.; Tonelli, C. Anthocyanins in Corn: A Wealth of Genes for Human Health. *Planta* **2014**, *240* (5), 901–911.

Poulton, J. E. Transmethylation and Demethylation Reactions in the Metabolism of Secondary Plant Products. In *Secondary Plant Products*; Elsevier, 1981; pp 667–723.

Radicella, J. P.; Turks, D.; Chandler, V. L. Cloning and Nucleotide Sequence of a cDNA Encoding B-Peru, a Regulatory Protein of the Anthocyanin Pathway in Maize. *Plant Mol. Biol.* **1991**, *17* (1), 127–130.

Reddy, A. R.; Britsch, L.; Salamini, F.; Saedler, H.; Rohde, W. The A1 (Anthocyanin-1) Locus in Zea Mays Encodes Dihydroquercetin Reductase. *Plant Sci.* **1987**, *52* (1), 7–13.

Salinas-Moreno, Y.; Pérez-Alonso, J. J.; Vázquez-Carrillo, G.; Aragón-Cuevas, F.; Velázquez-Cardelas, G. A. Antocianinas Y Actividad Antioxidante En Maíces (Zea Mays L.) de Las Razas Chalqueño, Elotes Cónicos Y Bolita. *Agrociencia* **2012**, *46* (7), 693–706.

Schuster, S. C. Next-Generation Sequencing Transforms Today's Biology. *Nat. Methods* **2007**, *5* (1), 16–18.

Selinger, D. A.; Chandler, V. L. A Mutation in the Pale Aleurone Color1 Gene Identifies a Novel Regulator of the Maize Anthocyanin Pathway. *Plant Cell* **1999**, *11* (1), 5–14.

Sharma, M.; Chai, C.; Morohashi, K.; Grotewold, E.; Snook, M. E.; Chopra, S. Expression of Flavonoid 3'-Hydroxylase Is Controlled by P1, the Regulator of 3-Deoxyflavonoid Biosynthesis in Maize. *BMC Plant Biol.* **2012**, *12* (1), 1.

Sharma, M.; Cortes-Cruz, M.; Ahern, K. R.; McMullen, M.; Brutnell, T. P.; Chopra, S. Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. *Genetics* **2011**, *188* (1), 69–79.

Somavat, P.; Li, Q.; de Mejia, E. G.; Liu, W.; Singh, V. Coproduct Yield Comparisons of Purple, Blue and Yellow Dent Corn for Various Milling Processes. *Ind. Crops Prod.* **2016**, *87*, 266–272.

Springob, K.; Nakajima, J.; Yamazaki, M.; Saito, K. Recent Advances in the Biosynthesis and Accumulation of Anthocyanins. *Nat. Prod. Rep.* **2003**, *20* (3), 288.

St-Pierre, B.; De Luca, V. Evolution of Acyltransferase Genes: Origin and Diversification of the BAHD Superfamily of Acyltransferases Involved in Secondary Metabolism. In *Recent Advances in Phytochemistry*; Elsevier, 2000; Vol. 34, pp 285–315.

St-Pierre, B.; Laflamme, P.; Alarco, A. M.; De Luca, V. The Terminal O-Acetyltransferase Involved in Vindoline Biosynthesis Defines a New Class of Proteins Responsible for Coenzyme A-Dependent Acyl Transfer. *Plant J. Cell Mol. Biol.* **1998**, *14* (6), 703–713.

Suzuki, H.; Nakayama, T.; Yonekura-Sakakibara, K.; Fukui, Y.; Nakamura, N.; Nakao, M.; Tanaka, Y.; Yamaguchi, M. -a.; Kusumi, T.; Nishino, T. Malonyl-CoA:Anthocyanin 5-O-Glucoside-6’’’-O-Malonyltransferase from Scarlet Sage (Salvia Splendens) Flowers: Enzyme Purification, Gene Cloning, Expression, and Characterization. *J. Biol. Chem.* **2001**, *276* (52), 49013–49019.

Suzuki, H.; Nakayama, T.; Yonekura-Sakakibara, K.; Fukui, Y.; Nakamura, N.; Yamaguchi, M.; Tanaka, Y.; Kusumi, T.; Nishino, T. cDNA Cloning, Heterologous Expressions, and Functional Characterization of Malonyl-Coenzyme A:Anthocyanidin 3-O-Glucoside-6"-O-Malonyltransferase from Dahlia Flowers. *Plant Physiol.* **2002**, *130* (4), 2142–2151.

Tonelli, C.; Consonni, G.; Dolfini, S. F.; Dellaporta, S. L.; Viotti, A.; Gavazzi, G. Genetic and Molecular Analysis of Sn, a Light-Inducible, Tissue Specific Regulatory Gene in Maize. *Mol. Gen. Genet. MGG* **1991**, *225* (3), 401–410.

Tsuda, T.; Horio, F.; Uchida, K.; Aoki, H.; Osawa, T. Dietary Cyanidin 3-O-β-D-Glucoside-Rich Purple Corn Color Prevents Obesity and Ameliorates Hyperglycemia in Mice. *J. Nutr.* **2003**, *133* (7), 2125–2130.

Wienand, U.; Weydemann, U.; Niesbach-Klösgen, U.; Peterson, P. A.; Saedler, H. Molecular Cloning of the C2 Locus of Zea Mays, the Gene Coding for Chalcone Synthase. *Mol. Gen. Genet.* **1986**, *203* (2), 202–207.

Wilmouth, R. C.; Turnbull, J. J.; Welford, R. W. D.; Clifton, I. J.; Prescott, A. G.; Schofield, C. J. Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis Thaliana. *Structure* **2002**, *10* (1), 93–103.

Wolf, M. J.; Cutler, H. C.; Zuber, M. S.; Khoo, U. Maize with Multilayer Aleurone of High Protein Content. *Crop Sci.* **1972**, *12* (4), 440.

Wu, X.; Beecher, G. R.; Holden, J. M.; Haytowitz, D. B.; Gebhardt, S. E.; Prior, R. L. Concentrations of Anthocyanins in Common Foods in the United States and Estimation of Normal Consumption. *J. Agric. Food Chem.* **2006**, *54* (11), 4069–4075.

Yu, X.-H.; Gou, J.-Y.; Liu, C.-J. BAHD Superfamily of Acyl-CoA Dependent Acyltransferases in Populus and Arabidopsis: Bioinformatics and Gene Expression. *Plant Mol. Biol.* **2009**, *70* (4), 421–442.

Zafra-Stone, S.; Yasmin, T.; Bagchi, M.; Chatterjee, A.; Vinson, J. A.; Bagchi, D. Berry Anthocyanins as Novel Antioxidants in Human Health and Disease Prevention. *Mol. Nutr. Food Res.* **2007**, *51* (6), 675–683.

Chapter 2: Survey of Anthocyanin Production in Diverse Maize Germplasm

2.1 Abstract

Increasing consumer demand for natural ingredients in foods and beverages has necessitated a more economic source of natural colorants. In this study, 398 genetically diverse pigmented accessions of maize were analyzed using HPLC to characterize the diversity of anthocyanin composition in maize germplasm. This resulted in 167 accessions identified that could produce anthocyanins in the grain. These accessions were then categorized based on the abundance of pigments in the aleurone or pericarp layers and on compositional variations. Aleurone categories were divided on whether the kernels appeared blue or pink, a trait determined chemically by the abundance of either cyanidin or pelargonidin anthocyanins and genetically by the *Purple aleurone1* (*Pr1*) gene. Pericarp categories were divided based on the presence or absence of flavanol-anthocyanin condensed forms. In addition, a category was discovered that had previously never been characterized in maize; some accessions were unable to produce significant amounts of acylated anthocyanins which are typically the most abundant forms in most accessions. It is hypothesized that a partial loss of function anthocyanin acyltransferase is responsible for the reduced acylation trait. The major finding of this survey was that the highest performing category in terms of total anthocyanin content was pericarp-pigmentation with condensed forms. Breeders should focus on adapting Andean purple corn landraces to the Midwest to create purple corn hybrids. Broad-sense heritability of anthocyanin production in a selected subset of accessions showed that anthocyanin production is highly genetically controlled in maize, meaning anthocyanin production in maize can be improved effectively. This study represents the most comprehensive screening of pigmented maize lines to date. Information from this study will provide information to plant breeders looking to develop

anthocyanin-rich purple corn hybrids as an economic source of natural colorants for food and beverages.

2.2 Introduction

Anthocyanins are the visually appealing water-soluble pigments responsible for most of the red-orange to blue pigments exhibited in plants. The most obvious function of anthocyanins is as a colorful signaler for pollinators and seed dispersers (Koes et al., 1994). However, anthocyanins also protect the plant from photodamage and herbivory (Hatier and Gould, 2008; Lev-Yadun and Gould, 2008). In addition to their role for plants and animals, anthocyanins have been shown to possess anti-inflammatory, anticarcinogenic, antiangiogenesis, antimicrobial, cardioprotective and neuroprotective bioactivity in mammals. They also have suggested roles in preventing obesity and diabetes as well as roles in improving eye health (Zafra-Stone et al., 2007; He and Giusti, 2010).

Because of their attractive color and ease of aqueous extraction, anthocyanins make suitable natural replacements for certain synthetic dyes such as FD&C Red 40. Red 40 is the most abundant synthetic color additive produced in the US with over six million pounds certified each year. It accounts for almost 25% of the color additives produced in the US (Center for Food Safety and Applied Nutrition, 2015). Synthetic dyes are preferred by the industry to most natural colors due to their lower cost and relatively greater stability. However, with the ever-increasing demand for more natural ingredients in foods and beverages by consumers, natural colors are becoming increasingly important. More economic sources of anthocyanin pigments need to be investigated to meet this growing demand. Currently natural anthocyanins are recovered from purple fruits and vegetables and the remaining biomass is processed as animal feed or waste. Pigmented maize represents a unique opportunity from a processing standpoint because

anthocyanin-producing tissue from the grain can be isolated using one of several milling processes available as a value-added coproduct while the remaining fractions can be sold as food, fuel, or feed (Somavat et al., 2016).

Pigmented maize has been utilized as a source of natural colors for centuries. In Andean cultures, high anthocyanin yielding purple corn, referred to as "*Maize morado*", is still important for producing local foods and beverages (Petroni et al., 2014). Purple corn produces anthocyanins most abundantly in the pericarp of the kernel, which is the outermost layer of the kernel (Figure 2.1). Blue and pink corn varieties typically produce anthocyanins in the peripheral layer of the endosperm called the aleurone. Genetically, pericarp is diploid maternal tissue, while aleurone is triploid tissue consisting of two maternal and one paternal genomes (Soave and Salamini, 1984).

The chemical structure of an anthocyanin consists of an anthocyanidin bound to a glycoside. Three types of anthocyanidins are possible in maize: pelargonidin, cyanidin, and peonidin. These differ by hydroxylation or methoxylation of the phenyl group on the flavylium cation. The structure of an anthocyanidin molecule is provided in Figure 2.2. The glycoside in maize is most often glucose, but arabinose, galactose, rutinose, and rhamnose glycosides have been detected (Bhatla and Pant, 1977; Grotewold et al., 1998; Abdel-Aal et al., 2006).

Due to its conspicuous nature, the anthocyanin biosynthetic pathway is the most thoroughly studied plant secondary metabolite pathway known today (Irani et al., 2003). All the essential genes in the maize anthocyanin biosynthetic pathway have been cloned and characterized to date (Sharma et al., 2011). A diagram of the anthocyanin pathway in maize with gene products is given in Figure 2.3. The first step of anthocyanin synthesis involves the formation of naringenin chalcone by three malonyl-CoA and one *p*-coumaroyl-CoA molecules

by chalcone synthase (CHS) (Han et al., 2016). Subsequent modification by chalcone isomerase (CHI) leads to the formation of flavanones (Grotewold and Peterson, 1994). This step represents a major branch point that determines the type of flavonoid made. If the *Pericarp color1* (*P1*) gene is active, phlobaphene pigments can be synthesized from flavanones. Phlobaphenes are water insoluble red pigments found in pericarp, cob and tassel glumes, and husks (Sharma et al., 2012). If a functional flavonoid 3'-hydroxylase (F3'H) is present at this step, then the flavanone naringenin can be converted to eriodictyol by the addition of an extra hydroxyl group at the 3'-position on the flavonoid B-ring (Figure 2.2). This reaction is catalyzed by the maize *Purple aleurone1* or *Pr1* gene (Sharma et al., 2011). Lines homozygous for recessive alleles of *pr1* are mainly composed of pelargonidin anthocyanins. Phenotypically, the change is very noticeable in aleurone-pigmented lines. Recessive *pr1* backgrounds appear red or pink in aleurone while dominant *Pr1* alleles are more blue.

After the formation of flavanones, flavanone 3-hydroxylase (F3H) adds an additional hydroxyl group to create dihydroflavonols (Deboo et al., 1995). One dihydroflavanol, dihydrokaempferol, can also be hydroxylated by *Pr1* in this step to form dihydroquercetin, allowing for two enzymatic steps to create cyanidin and peonidin anthocyanins. Next, dihydroflavanol 4-reductase (DFR) reduces a ketone group to a hydroxyl group at the 4-position, then anthocyanidin synthase (ANS) removes the 4-OH to form an anthocyanidin (Wilmouth et al., 2002). Here, a colored molecule exists, it is just unstable. Anthocyanin structures are in constant flux depending on the pH of the system. They are most stable in their flavilium forms at a pH under 2.0 and quickly change to colorless carbinol pseudobase and chalcone forms at a pH above 4.5 (Lee et al., 2005). Within cytosolic conditions, anthocyanins are unstable and prone to oxidative breakdown (Suzuki et al., 2002). For this reason, the next two steps glycosylate

anthocyanidins and move them into the vacuole where they are most stable. The last two steps are accomplished through a UDP glucose-flavonoid glucosyltransferase and a glutathione *S*-transferase in maize. The genes encoding these enzymes are *Bronze1* and *Bronze2*, respectively (Marrs et al., 1995). Homozygous recessive *Bronze* mutants produce a titular brown color in the grain due to unknown flavonoid degradation products in the cytosol.

Subsequent modifications to anthocyanins compound the diversity of pigment production while also aiding stability. In maize, several anthocyanin modifications are known to produce more stable compounds. The most common modification to maize anthocyanins is acylation. Acylation is the process of adding an acyl group to the glycoside of an anthocyanin molecule. In maize, the addition of a malonyl group is most common. Malonylation increases stability *in planta* by protecting the glycosylated sugar from enzymatic breakdown and stabilizing the anthocyanin structure under more alkaline conditions (Suzuki et al., 2002). Malonylation has also been shown to increase total anthocyanin content (TAC) by enhancing anthocyanin solubility and increasing uptake into vacuoles (Nakayama et al., 2003).

A second modification that may be important for stability is the formation of flavanol-anthocyanin dimers, also known as “condensed forms”. These compounds were first discovered as a product of wine fermentation (Vivar-Quintana et al., 1999). Since their discovery in wine, condensed forms were found to be important naturally forming pigments in many crops, including strawberries, black currants, beans, and maize. The major condensed forms in maize were first characterized by LC-MS/MS and H-NMR in a previous study (González-Manzano et al., 2008).

Maize has potential as an economic source of natural colors. To assess the diversity of anthocyanin production in maize, 398 diverse accessions of pigmented maize were analyzed

using High Performance Liquid Chromatography (HPLC). To test repeatability, a subset of these accessions were grown for several seasons and analyzed. To our knowledge, this is the largest collection gathered to date for the purpose of characterizing anthocyanin production in maize. Data from this investigation will provide information to plant breeders looking to develop varieties of maize with enhanced levels of stable anthocyanins to be used as an economic source of natural colorants in foods and beverages.

2.3 Materials and Methods

2.3.1 Plant Material

Pigmented maize (*Zea mays* L.) accessions were collected from the North Central Regional Plant Introduction Station (NCRPIS) in Ames, IA, USA; the International Maize and Wheat Improvement Center (CIMMYT) in Mexico; Native Seeds/SEARCH (NS) in Tucson, AZ, USA; the Maize Genetics Cooperation Stock Center (MGCSC) in Urbana, IL, USA; and various commercial sources. Representative kernels of the original stock were analyzed on the HPLC as described below and designated as the first pseudo-environment. In 2014, remaining stock from the NCRPIS, MGCSC, and commercial sources were grown in two replications at the University of Illinois Vegetable Research Farm in Champaign, IL, USA (40˚ 04′ 38.89″ N, 88˚ 14′ 26.18″ W) with 25’ rows spaced 30” apart. Individuals were selfed to maintain genetic purity and to produce stock for the next season. After harvest, ears were dried in a heated forced air dryer for at least five days and then shelled. Ears were either shelled separately or in bulk and then analyzed as described below. Many accessions in 2014 could not produce sufficient grain for analysis due to disease pressure or photoperiod sensitivity issues. In addition, some samples harvested were missing data on the replication to which they belonged. All missing replication samples were included with replication 1 for analyses. In 2015, a subset of 43 accessions was

grown in a randomized complete block design at the same location with three replications in 25' rows and 30" spacing. A list of the subset with major phenotypes and sample counts is in listed in Table 2.1. At least five individuals from each plot were selfed for a seed increase when possible. After harvest, samples were dried as previously described then shelled separately. A maximum of three ears per plot were chosen for analyses and averaged. The subset of accessions was initially chosen based on estimated grain yield in 2014. If it was known that an accessions would not be able to produce sufficient kernels for analysis, it was excluded. After grain yield, accessions were narrowed down based on their phenotypic stability. Many accessions were segregating for genes with large effects on anthocyanin composition and were avoided in the subset. Some accessions still unknowingly segregated for these genes in 2015. For example, two accessions segregated for *pr1* alleles and two other accessions segregated for the reduced acylation trait (See 2.3 Results and Discussion for a description of the trait). These accessions were still included, but only their genetically dominant phenotype was used for analyses. Four accessions contained both pericarp and aleurone pigmentation together. In all cases these lines were analyzed as aleurone-pigmented accessions.

2.3.2 Sample Preparation

Representative kernel samples of each accession were ground to a fine powder in a coffee grinder. Accessions knowingly segregating for *pr1* alleles were separated visually and analyzed separately. One gram of whole corn powder was weighed into a 15 mL conical centrifuge tube and extracted with 5 mL 2% (v/v) formic acid (ACS Reagent Grade) in distilled, 2 μm Millipore (Billerica, MA, USA) filtered water. Air in the centrifuge tube was replaced with argon (0 grade or purer) before extracting overnight (12-16 hours). Samples were kept in the dark, at room temperature, and were constantly rotated on a LabQuake (Thermo Fisher Scientific Inc.) test tube

rotator to evenly extract the powder. After extraction, samples were centrifuged and the supernatant was filtered through a 25 mm 0.45 µm Millex Millipore LCR PTFE syringe filter.

2.3.3 HPLC Analysis

A 20 µL aliquot of anthocyanin extract was separated within a Grace Prevail C_{18} 5 µm (250 mm x 4.6 mm) analytical column (W. R. Grace & Co., Columbia, MD, USA) maintained at 30.0 ˚C. A Hitachi L-7250 HPLC (Hitachi High Technologies America, Inc., Schaumburg, IL, USA) equipped with a Hitachi L-7400 ultraviolet-visible detector set to 520 nm was used to generate chromatograms. The mobile phase used 2% (v/v) formic acid as solvent A and acetonitrile as solvent B at a flow rate of 1 mL/min in the following linear gradient: 100% to 90% A for 3 min, 60% A at 30 min, then 100% A at 35 min. The column was allowed to equilibrate for 10 min before each sample with 100% A.

2.3.4 Total Anthocyanin Content (TAC)

Total anthocyanin content (TAC) was determined by summing peak areas integrated on the Hitachi HPLC Software Management 4.0 software. All measurements were calibrated using a bulk sample of commercially available Angelina's Gourmet Maize Morado (Swanson, CT, USA) as a relative external standard. To prepare MM standard, large quantities of MM would be ground in a coffee grinder then 1.0 g would be extracted as described above. TAC of each new batch of MM powder was quantified using C3G standard. C3G peak areas of concentrations ranging from 1 to 1000 µg/mL were used to calculate the linear slope of the standard in Excel (Microsoft Corp., Redmond, WA, USA). TAC of MM was set as 1000 mg anthocyanins per kg whole corn, based on the C3G standard calibration. MM was analyzed in replicates throughout every new run of samples. A run was considered new if the HPLC was turned off and a new set of samples were extracted. MM standard was a relative check to ensure consistency between

batches. To quantify the weight of anthocyanins in a sample per weight of whole corn, Equation 2.1 was used. In Equation 2.1, TAC is expressed in units of mg anthocyanins in C3G equivalents per kg whole corn relative to the MM concentration. For simplicity, units for TAC will be shortened to mg/kg from this point forward. In Equation 2.1, *MM* represents the maize morado concentration of 1000 mg anthocyanins per kg MM.

Equation 2.1: $$TAC = \frac{Total\ Peak\ Area\ Sample}{Total\ Peak\ Area\ MM} \times MM$$

2.3.5 Anthocyanin Identification

Cyanidin 3-glucoside (C3G), pelargonidin 3-glucoside (Pg3G), and peonidin 3-glucoside (Pn3G) standards were obtained from Phytolab GmbH & Co. (Vestenbergsgreuth, Germany) at 89% purity. Remaining compounds were presumed based on atomic masses obtained by LC-MS and by comparing elution orders in previous literature (Figure 2.4; de Pascual-Teresa and Sanchez-Ballesta, 2008; Cuevas Montilla et al., 2011; Lao and Giusti, 2015). In the gradient method used, peonidin 3-(6”-malonyl)glucoside (Pn3MG) coelutes with cyanidin 3-(3”,6”-dimalonyl)glucoside (C3DMG). Because of this, the proportion of peonidin or cyanidin in a sample can only be estimated. The proportion of condensed forms was calculated by summing peak areas before C3G. Major condensed form catechin-(4-8)-cyanidin-3,5-diglucoside resides in this region according to LC-MS results. The formation of this condensed form was always followed by the production of several pigments, which are tentatively identified as condensed forms (Figure 2.4c). Proportion of acylation was calculated by summing the peak areas of identified acylated compounds Pg3MG, Pn3MG, C3DMG, cyanidin 3-(6”-malonyl)glucoside (C3MG), pelargonidin 3-(3”,6”-dimalonyl)glucoside (Pg3DMG), and peonidin 3-(3”,6”-dimalonyl)glucoside (Pn3DMG). Concentrations of individual compounds in mg/kg were

calculated similarly to TAC in Equation 2.1. The only difference is the *Total Peak Area Sample* term is substituted with peak area of the compound.

2.3.6 Comparison to Acidified Methanol Extract

Acidified methanol is an acceptable alternative to aqueous formic acid extraction. To compare compounds extracted from the two methods, the following procedure was employed. Triplicate 2 g samples of ground MM were extracted in 2% (v/v) formic acid or methanol containing 0.01% HCl as described above. However, after centrifugation the supernatants were decanted into scintillation vials and placed in a freezer (-20 ˚C) until use. 10 mL of solvent was added to each pellet and the extraction process repeated until no visible color remained. In 2% (v/v) formic acid the process was repeated eight times and with acidified methanol, five times. When all extractions were completed, samples were filtered, diluted (2% [v/v] formic acid extracts only, acidified methanol extracts did not require dilution), and analyzed spectrophotometrically as described in Lee et al. (2005). With the completion of these analyses, samples from the acidified methanol extract replicates 1 and 2 were separately combined. The solvent was removed from the combined acidified methanol extracts on a rotary evaporator using water-aspirator vacuum and a bath temperature of 40 °C. The dried extracts were quantitatively removed from the evaporator flasks using 2% (v/v) formic acid and brought up to 100 mL in volumetric flasks. Samples of replicates 1 and 2 for the 2% (v/v) formic acid extracts were likewise separately combined and brought up to 100 mL. No solvent removal step was necessary for these samples. Portions of the combined extracts were filtered, diluted, and analyzed spectrophotometrically as before. Each extract from replicate 3 of the acidified methanol series were dried separately using a SpeedVac concentrator (Thermo Fisher Scientific Inc.) with the

centrifugation chamber at room temperature. When dry, all samples were dissolved in 2% (v/v) formic acid, filtered, and analyzed on the HPLC as described above.

2.3.7 Statistical Analyses

Linear regressions were calculated in Proc Reg in SAS Enterprise Edition Release 3.5 (SAS Institute Inc., Cary, NC, USA) with the generalized first-order linear model shown in Equation 2.2. $\boldsymbol{Y}_i$ refers to the response phenotype of interest; $\boldsymbol{\beta_0}$ refers to the intercept; $\sum_{k=1}^{p-1} \boldsymbol{\beta_k X_{ik}}$ refers to the summation of predictor variables, where *p* is the total number of parameters in the model; and $\boldsymbol{\varepsilon}_i$ refers to the error term.

Equation 2.2: $$\boldsymbol{Y_i} = \boldsymbol{\beta_0} + \sum_{k=1}^{p-1} \boldsymbol{\beta_k X_{ik}} + \boldsymbol{\varepsilon_i}$$

Coefficient of variation (CV) was calculated as $(\sigma / \bar{x}) \times 100$ where σ is the standard deviation and $\bar{x}$ refers to the average of the samples. Tukey Honest Significant Differences (HSDs) among the various anthocyanin compositional categories were calculated in Proc GLM of SAS. Correlations between compounds used the average concentrations from each accession and were calculated in Proc Corr in SAS. Principal component analysis (PCA) was performed in R using the princomp function (R Core Team, 2015). Peak areas for all known compounds or groups of compounds were converted to percentages of total peak area in the sample so principal components (PCs) could be calculated. PCA was performed on the variance covariance matrix since phenotypes were already approximately normalized when converted to percent data. Hierarchical clustering used Ward's minimum variance method in R (Ward, 1963; R Core Team, 2015). PCA plots were generated using the ggplot2 function (Wickham, 2009). Mixed model ANOVAs were calculated using Proc Mixed in SAS with method equal to Type 3. The model to calculate a single year ANOVA is shown in Equation 2.3. y_{ijk} is the response for the phenotype, μ is the grand mean, β_i is the random effect of the *i*th replication, γ_j is the random effect of the

*j*th genotype, $\beta\gamma_{ij}$ is the random interaction of the *j*th genotype in the *i*th replication, and $\varepsilon_{(ij)k}$ is the random error term.

$$\text{Equation 2.3: } y_{ijk} = \mu + \beta_i + \gamma_j + \beta\gamma_{ij} + \varepsilon_{(ij)k}$$

The model to calculate a multiple-year ANOVA is shown in Equation 2.4. y_{ijkl} is the response for the phenotype, μ is the grand mean, α_i is the random effect of the *i*th environment, $\beta_{j(i)}$ is the random effect of the *j*th replication within environment, γ_k is the random effect of the *k*th genotype, $\alpha\gamma_{ik}$ is the random interaction of the *k*th genotype in the *i*th environment, and $\varepsilon_{(ijk)l}$ is the random error term.

$$\text{Equation 2.4: } y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \varepsilon_{(ijk)l}$$

2.3.8 Heritability Calculations

Broad-sense heritability (H^2) was calculated using the method from Bernardo (2010) and adapted as shown in Equation 2.5. In Equation 2.5, σ^2_e is variance due to error, σ^2_{ge} is the variance due to the genotype by environment interaction, σ^2_g is the variance due to genotype, *r* is degrees of freedom for replications, and *e* is degrees of freedom for environments.

$$\text{Equation 2.5: } \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_e^2}{r \times e}}$$

All variance components were estimated using Proc Mixed in SAS with the model in Equation 2.4. Due to missing data, *r* and *e* were not exactly 2 and 3, respectively, as calculated conventionally. To obtain a better estimate, the expected mean square for genotype given in the ANOVA was solved for *r* and *e* as shown in equation 2.6.

$$\text{Equation 2.6: } \sigma_e^2 + r \times \sigma_{ge}^2 + r \times e \times \sigma_g^2$$

$$= \sigma_e^2 + 1.2772 \times \sigma_{ge}^2 + 3.6979 \times \sigma_g^2$$

$$r = 1.2772;\ e = 2.8953$$

2.4 Results and Discussion

2.4.1 Extraction and HPLC Method

Presented here is a simple and reproducible method for relative quantification of TAC in maize. Using bulk MM as a relative standard simplifies quantification, reduces run time, and reduces costs by eliminating the reliance purely on standards. The quantification method was designed to be as high-throughput as possible without losing reproducibility. Since MM was replicated every new batch, a good estimate of the repeatability of the protocol can be estimated. A subsample of nine extractions of MM from the same bulk powder replicated two to three times was chosen to test repeatability. The CV of the nine batches ranged from 0.62% to 6.06% with an overall average CV of 2.92% meaning the HPLC method has high consistency within batches of samples. Error can be introduced when a new bulk sample of MM is ground and goes uncalibrated. Every new batch of MM powder must be calibrated with C3G standard to ensure that relative quantification is accurate. The CV was also compared among five 2 g and five 1 g samples of MM powder extracted with a 1:5 dilution as described above. CV was similar (2% to 4%) with both extraction methods meaning both have highly repeatability. More whole corn powder should theoretically increase homogeneity, but a 1 g to 5 mL dilution was chosen so less of a sample had to be destroyed for analyses.

The extraction method here is reproducible, but cannot be considered an exhaustive extraction of anthocyanins. An exhaustive extraction method was not chosen because it would increase the technicality of the protocol, the time to complete the protocol, and the reagent costs. A high-throughput method was preferred due to the volume of samples analyzed. After nine 1:5 dilution extraction steps with five samples of MM powder, it was found that approximately 37% of extractable anthocyanins are consistently removed in the first extraction (Figure 2.5b). Future

work could be focused on developing a correction factor for TAC using samples with a range of TAC. Since the method was designed for relative quantification, the correction factor is not necessary for comparing among samples here.

Aqueous extraction utilized in this study was meant to more closely resemble commercial extraction procedures so results can be comparable. Acidified methanol is a common alternative to the extraction method used here. For this reason, the results of both solvent systems were compared to see if conclusions would be different in different solvent systems. Formic acid extraction was better at extracting condensed forms and acylated anthocyanins (Figure 2.5a). Preferentially extracting these compounds is advantageous because condensed forms and acylated anthocyanins are thought to be important pigments for TAC and stability. Knowing that certain compounds extract more efficiently in different solvents, compound proportions calculated here are relative to the extraction method and not absolute. Despite these limitations, this extraction method is efficient, reproducible, and ideal for the numerous samples that would be expected in a breeding program.

2.4.2 Categorization of Accessions

With the abundance of accessions in the survey, categories based on visual characteristics and compositional data were developed to make meaningful clusters. Of the 398 accessions collected, 167 were capable of producing anthocyanins in detectable amounts in the grain. Complete categorized anthocyanin producing accessions with anthocyanin yields are shown in Supplementary Table 2.9. A summary of Supplementary Table 2.9 is in Table 2.2. Phlobaphenes were the most common flavonoids outside of anthocyanins (n=166). Phlobaphenes are brick-red pigments that can be mistaken for anthocyanin coloration. These pigments are only extractable with non-aqueous solvents, which are not ideal for food and beverage systems (Sharma, et al.,

2012). A list of phlobaphene-producing accessions is in Supplementary Table 2.10. This list may not be complete since accessions that produced anthocyanins in the pericarp masked phlobaphene pigments. A few accessions (*n*=11) produced bronze pigments not detectable with the HPLC method used. Many of these accessions may contain homozygous recessive *Bronze* mutants (Marrs et al., 1995). A list of accessions producing bronze pigments is also in Supplementary Table 2.10.

Within the anthocyanin producing accessions, the first major category designed split accessions visually on the presence or absence of anthocyanins in the pericarp (Figure 2.1). Aleurone layers do not develop on the germ of the kernel, which makes them easy to distinguish from pericarp-pigmentation (Figure 2.1, 2.6a, 2.6b, and 2.6d). Phlobaphenes were easy to distinguish from anthocyanin pigmentation due to the distinct brick-red pigments for phlobaphenes in contrast to dark purple anthocyanin pigments, and due to the inability to extract pigments in phlobaphene lines. Many contained phlobaphenes with aleurone pigmentation. Any accessions that were unclear were sectioned roughly with a razorblade and viewed under a compound microscope. Aleurone-pigmented accessions had the highest representation in the survey. Over 82% (*n*=137) were capable of producing pigments in the aleurone (Table 2.2) A few (*n*=4) were found to produce anthocyanins in both kernel layers. Separating the two layers is difficult, so in all cases the accessions that produced anthocyanins in both layers were combined with aleurone-pigmented lines for analyses. The TAC of these four accessions ranged from 22.6 to 85.2 mg/kg, which is within the normal range of aleurone-pigmented accessions. The pericarp layers of these lines are relatively thin, which may explain why TAC is low compared to other pericarp-pigmented lines.

Within the category of kernel tissue, accessions could be further divided based on anthocyanin composition. Within pericarp-pigmented accessions, composition was divided on the presence or absence of condensed forms (Figure 2.4). Aleurone-pigmented lines could visually be separated on whether they produced blue or pink kernels (Figure 2.6). Chemically, the effect of this gene could be detected with the HPLC. In the pink aleurone category, Pg3G and its derivatives are the most abundant pigments (>50% of total pigments) while in the blue aleurone category, C3G and its derivatives are most abundant (Figure 2.4). Genetically, this is due to a homozygous recessive *pr1* gene that is unable to convert Pg3G precursors to C3G precursors (Sharma et al., 2011). Only three homozygous *pr1* recessive accessions produced anthocyanins in the pericarp. Puebla 403 (PI 485071) and Puebla 456 (PI 489081) produced both pericarp and aleurone pigmentation, while Apache Red (Siskiyou Seeds, Williams, OR, USA) produced anthocyanins exclusively in the pericarp. Apache Red was also unique because it could produce the pelargonidin condensed forms that are not detectable or in high abundance in other accessions.

In addition to these four compositional categories, a unique trait was discovered that had previously never been characterized in maize. Seven accessions within the collection produced markedly less acylated anthocyanins than most other accessions (Figure 2.4e). Across all categories, C3G ranged from 2.99% to 28.3% (Table 2.2). In lines with this unique trait, C3G was the dominant pigment and averaged 57.7% of total anthocyanins. Some acylated anthocyanins can be detected in these unique lines, but the average is only 7.6%, which is much lower than the average of 58.9% observed in other categories. This unique phenotype has been found in several diverse backgrounds that seem to have no relation. This trait will be referred to

as “reduced acylation”. The hypothesis is that reduced acylation is due to a partial loss-of-function anthocyanin acyltransferase (See Chapter 3).

It is important to note that some accessions belonged to more than one category. All the compositional categories presented here can be described by the actions of a few genes, so segregating between categories is common. Accessions belonging to more than one category were coded as unique accessions so accurate conclusions could be made about each category in terms of composition. For example, Apache Red was segregating for *pr1* alleles and the ability to produce condensed forms simultaneously, so it appears four times in Supplementary Table 2.10.

2.4.3 TAC Results

Across the whole survey, average TAC was 64.7 mg/kg with a maximum of 2560 mg/kg (Table 2.2). The highest performing accession in terms of TAC was a Peruvian landrace named Arequipa 204 (PI 571427) that had an average TAC of 1100 mg/kg. Andean purple corn landraces in general had high TAC, but very low grain yield. Landraces from the tropics are poorly adapted to the Midwest and are plagued with photoperiod sensitivity and disease susceptibility (White et al., 2012). Adapting these landraces to the corn belt region of the US may require backcrossing to Midwestern inbreds to improve grain yield. Within the aleurone-pigmented accessions, the highest total anthocyanin yielding varieties were homozygous recessive for *intensifier1* (*in1*). Stocks designated 707G and 707B were donated by the MGCSC and were differentiated by homozygous dominant and recessive *Pr1* alleles, respectively. 707G averaged 133 mg/kg and 707B 128 mg/kg, which are almost 1.5 times the TAC of the next best aleurone-pigmented accession. This demonstrates the importance of *in1* as an enhancer of TAC in the aleurone.

The highest-performing category in terms of TAC was pericarp-pigmented lines that could produce condensed forms (Table 2.2). These accessions produced on average 251 mg/kg, which is significantly greater than all other categories. Condensed forms may provide stability to compounds just as malonylation does in the cytosol and vacuole. In general, pericarp-pigmented accessions have a greater TAC potential. The average TAC of all aleurone-pigmented accessions was 29.0 mg/kg and 206 mg/kg for pericarp-pigmented accessions. This is in agreement with Ryu et al. (2013) and Abdel-Aal et al. (2006), who found the highest TAC in purple pericarp corn. The greater potential for anthocyanin production in pericarp may be due to the processes that form pericarp tissue. Pericarp contains as many as 5 to 22 cell layers due to the fusion of maternal tissues during development (Morohashi et al., 2012). Typically aleurone is only a single layer, but several accessions have been found that are capable of producing up to six layers (Wolf et al., 1972). Integrating multiple aleurone layers and the *in1* gene may be a route to increase TAC in aleurone.

Comparisons between categories found no statistical difference in TAC between blue and pink aleurone accessions ($p>0.05$), but on average, pink aleurone had lower TAC (22.6 mg/kg versus 30.8 mg/kg; Table 2.2). Blue aleurone accessions were not statistically different from pericarp-pigmented lines without condensed forms, but pink aleurone was significantly different. The most likely reason significance could not be established was because of the underrepresentation of pericarp-pigmented accessions in the survey. Only thirteen accessions were included in the pericarp-pigmentation without condensed form category, while 98 were included in blue aleurone category (Table 2.2).

2.4.4 Compositional Differences Between Compositional Categories

Proportion of known acylated compounds was lowest (7.6%) in the reduced acylation category as expected (Table 2.2). However, acylation was also significantly lower in the pericarp-pigmentation with condensed forms category. Average acylation ranged from 56.8% to 63.2% of total anthocyanins in the other categories, but only 35.7% in this category. One possible explanation may be due to the limited identification of condensed forms. González-Manzano et al. (2008) confirm that malonylglucoside anthocyanins can be conjugated to catechins and epicatechins. Many acylated anthocyanins may have been included within the condensed form calculation. With the results of this survey, it can be concluded that acylated anthocyanins are typically the most predominant pigments produced in maize. Despite preferential extraction of acylated anthocyanins with formic acid, the result found here is similar to other studies that analyzed blue and purple corn (Moreno et al., 2005; Abdel-Aal et al., 2006; Salinas-Moreno et al., 2012).

C3MG is the single most abundant compound on average in the whole collection. It was most abundant compound in the pericarp without condensed forms and blue aleurone categories with a difference of 3.98% and 16.4% from C3G, respectively. The fact that C3MG is in higher abundance than C3G is in contrast with another study that found C3G to be 31% to 51% of total anthocyanins in *Pr1* dominant lines (Abdel-Aal et al., 2006). In the collection here, the average proportion of C3G in each category besides reduced acylation ranged from 2.99% to 28.3% (Table 2.2). Abdel-Aal et al. (2006) used an acidified methanol extraction which may have inflated the C3G concentration within their samples and may explain the discrepancy between this study and theirs. The category with the lowest average of C3G was pink aleurone. Homozygous recessive *pr1* lines almost always produced low, but detectable amounts of C3G

although non-functional alleles of *pr1* should theoretically not be able to produce C3G. This may be evidence of an additional F3'H in maize expressed at lower levels than *Pr1,* or at least reduced function alleles of *pr1* (Larson et al., 1986).

Pink aleurone accessions were much more represented in this collection than any other pigmented maize collection indicating that *pr1* recessive alleles are common variations in maize germplasm. Selecting for these recessive alleles in breeding programs may provide a wider range of hues in the anthocyanin extracts that the food industry may be able to utilize. Pelargonidin-predominant extracts tend to be more orange to red-orange, while cyanidin-predominant extracts tended to be more red to purple (Figure 2.6). The pink aleurone category did not have significantly greater proportions of total acylation, but it did have greater proportions of Pg3DMG compared to all other categories. The proportion of dimalonyl anthocyanins in pink aleurone lines may be indicative of the rate of dimalonyltransferase activity in maize. Dimalonyltransferase activity is most likely a rate-limiting reaction since (6"-malonyl)glucosides are more common in maize. This is in agreement with observed reaction kinetics for the dimalonyltransferase in chrysanthemum (*Dm3MaT2*). 6" malonylation is more enzymatically favored than 3" malonylation in *Dm3MaT2* (Suzuki et al., 2004a). Since the proportion of C3DMG cannot be calculated directly because Pn3MG co-elutes, a prediction based on pink aleurone lines may be appropriate. A summary of models used to predict Pg3DMG concentration and percentage is in Table 2.3. The models were built in Proc Reg using Equation 2.2. The most parsimonious model with the highest R^2 was chosen to predict Pg3DMG. The Pg3MG model could predict Pg3DMG concentration with an R^2 of 0.83 (Table 2.4). Assuming the enzyme kinetics for Pg3G are the same for C3G, the concentration of C3DMG in a sample can be roughly estimated by substituting parameter estimates from Table 2.4 into the linear model to

form Equation 2.7. $\boldsymbol{X_i}$ refers to C3MG concentration in mg and $\boldsymbol{Y_i}$ refers to C3DMG concentration in mg.

Equation 2.7: $\boldsymbol{Y_i} = 0.45708 + 0.56901 \times \boldsymbol{X_i}$

For future work, if the concentration of C3DMG in a sample is of interest, the proportion of cyanidin and peonidin anthocyanins before and after acid or alkaline hydrolysis is an alternative way to calculate approximate C3DMG concentration (Moreno et al., 2005).

Peonidin has potential for adding stability to maize extracts and is therefore a pigment that must be investigated. It has been demonstrated that anthocyanins with fewer free hydroxyl groups on the B-ring tend to be more stable, but the results are somewhat unclear (Cabrita et al., 2000). Nevertheless, the category with the highest proportion of Pn3G was pericarp without condensed forms (Table 2.2). This category also had the highest average abundance of Pn3DMG with 4.5 mg/kg on average. Generally, peonidin was in low amounts on average in all samples.

Condensed form pigments in this study are much more prevalent than previously found. The initial characterization of condensed form pigments in maize found them to account for 0.3% to 3.2% of total anthocyanins (González-Manzano et al., 2008). In the condensed form category here, the average proportion of condensed forms in each accession ranges from 3.8% to 32.8%, with an overall average of 22.7%. The most abundant condensed form in maize was characterized in another study, but also confirmed here with MS (Figure 2.4; González-Manzano et al., 2008). In MM, this pigment alone is approximately 13 to 14% of total anthocyanins. The underrepresentation of condensed forms in most studies may be due to the utilization of acidified methanol as the choice solvent system (Figure 2.5a). Aqueous solvents should be utilized to accurately represent the concentration of these pigments since they have potentially important effects on TAC. Wide variation in condensed form content among accessions in the collection,

however, indicates there may also be genetic diversity for condensed form concentration that can be improved with breeding.

2.4.5 Correlation of Compounds

Concentrations of each compound significantly correlated with the TAC of the sample (Table 2.5). Most were moderate to strong correlations ($\rho>0.60$) with the exception of Pg3DMG ($\rho=0.32$). When separating the dataset by *Pr1* alleles, the correlation of Pg3DMG improved to 0.60 and 0.70 for dominant and homozygous recessive alleles of *pr1*, respectively. C3G was able to predict the TAC with a correlation of 0.93 in *Pr1* dominant lines, while Pg3G was able predict the TAC for homozygous *pr1* recessive lines with a correlation of 0.96. The malonylglucoside forms of the dominant anthocyanins were better predictors of TAC in both *Pr1* allele categories, with correlations of 0.97 for both homozygous dominant and recessive alleles.

2.4.6 Principal Component Analysis

Overall, the principal component analysis (PCA) is effective at explaining a large proportion of the variability in the dataset (Figure 2.7a). Principle component 1 (PC1) explained 60.8% of the total variance, and PC2 explained 21.2% of the total variability in the data set. Based on the loadings, PC1 primarily appears to be a contrast between the presence or absence of functional *Pr1* alleles. Cyanidin-based anthocyanins have positive PC1 loadings, while pelargonidin-based anthocyanins have negative PC1 loadings. PC2 primarily appears to be a contrast between condensed forms and C3G versus acylated anthocyanins. Condensed forms and C3G both had negative loadings, while C3MG, Pg3MG, and C3DMG/Pn3MG had positive loadings. Observations with high scores for PC2 likely had a higher proportion of acylated anthocyanins, while low scores on PC2 suggest few acylated forms and large amounts of C3G or condensed forms. Because of this most observations with high loadings for PC2 tended be in the

aleurone categories. More negative loadings for PC2 were indicative of pericarp with condensed forms and the reduced acylation category. Pericarp with condensed forms were intermediate in PC2 most likely due to the higher proportion of C3G on average than the blue aleurone category (Table 2.2). Additionally, the biplot shows good separation between observations and suggests there may be three natural groups within the data. A hierarchical clustering approach was used to cluster accessions with *a priori* knowledge so clusters could be unbiased. Ward's minimum variance method was used for hierarchical clustering (Ward, 1963). While dividing the dendrogram at the largest distance would produce two clusters in the dataset, dividing to make three clusters provides more separation and makes more meaningful clusters (Figure 2.7b). The first cluster comprises a bulk of the dataset and includes blue aleurone lines and pericarp without condensed forms. The second cluster is mainly a mixture of pericarp-pigmented accessions that produce condensed forms and reduced acylation accessions. The two lowest observations at PC2 were two reduced acylation accessions, primarily because C3G was the only pigment made in abundance in these accessions. If more reduced acylation accessions were present in the collection, this area of the biplot may be considered its own significant cluster. The third cluster consisted of accessions homozygous recessive for *pr1* and includes all pink aleurone accessions plus Apache Red. Results of the PCA show that variation in composition is largely explained by genetic factors that have a large effect on composition. Namely, *Pr1* and condensed form synthesis genes have large effects on composition. This provides evidence that the visual and compositional categories created for the accessions are sufficiently explaining the biology of the collection.

2.4.7 Heritability of Anthocyanin Content

A representative subset of the survey was planted in 2015 with three randomized replications to test the repeatability of anthocyanin production (Table 2.1). Accessions chosen for the subset were predicted to not segregate between anthocyanin compositional categories and were able to produce enough grain for HPLC analysis. Many accessions in the survey were poorly adapted to the Midwest growing environment. Table 2.1 shows the count of samples analyzed for each accession. Some had poor representation because in certain years, entire plots would be lost to disease pressures. Accessions with data from every year were still included in the analysis because they provide more degrees of freedom to test environments.

Although there were several accessions cut from the original collection, the subset still accurately represented the whole survey. Twelve of the forty-three accessions were homozygous *pr1* recessive, which is similar to the ratio of pink aleurone lines represented in the survey (Table 2.2). Two accessions in the subset were pericarp-pigmented lines without condensed forms. This ratio is also comparable to this category's representation in the whole survey. There were no lines with condensed forms in the subset because these accessions had generally low grain yield. The limitation of the entire survey, overall, was the limited representation of pericarp-pigmented accessions, which are of most importance for TAC. Conclusions drawn from this subset may more accurately describe the genetics of aleurone-pigmented lines rather then pericarp-pigmented lines. Future work will have to test repeatability in pericarp-pigmented crosses to see if the genotypic effects are similar to the subset.

Within 2015, the effect of replication for TAC was insignificant (p=0.70), meaning that TAC does not change drastically within the same environment like grain yield can (Table 2.7). The effect of replication for percentage of acylation was significant at p<0.1, but the proportion

of variance obtained from replication was small relative to genotype. In 2014, replication data was missing for several samples, so samples with missing data were included with replication 1. This will provide a poor test of within-replication variance in the analysis for all years, but since the variance attributed to this term is normally small or insignificant, it should have a negligible effect on genotypic variance. Although there is a poor test for replication, in the analysis for all years, replications were included in the model for consistency since it was considered significant in 2015 for at least one phenotype.

Results of this subset show high consistency in anthocyanin composition from environment to environment. The effect of environment on proportion of acylation and proportion of pelargonidin anthocyanins (Summed proportions of Pg3G, Pg3MG, and Pg3DMG), for example, was insignificant across all three environments. The average proportion of acylation ranged from 61.9% to 65.7% and average proportion of pelargonidin anthocyanins ranged from 26.7% to 30.1% across the three environments (Table 2.8). TAC was a little less consistent as some years were significantly different than others. The original source samples had the lowest average at 33.5 mg/kg, 2014 had the highest average at 62.6 mg/kg, and finally the average of 2015 was 40.1 mg/kg. Genotype by environment interactions were significant at $p<0.1$ for all compounds except Pg3MG. However, the variances contributed by genotype by environment effects are magnitudes lower than the variances due to genotypes.

H^2 for each compound or group of compounds is listed in Table 2.8. The lowest H^2 are from peonidin-based compounds Pn3G (H^2=0.746) and Pn3DMG (H^2=0.520). These two compounds generally had low or undetectable concentrations in most samples. The compound with the highest H^2 was C3G, with an H^2 of 0.967. TAC results were promising, with values of 0.911 and 0.917 for log-transformed TAC. High H^2 values mean most of the variability in

anthocyanin production is attributed to genetics and not environmental factors. Selecting for TAC or composition will be rapid, requiring few replications and locations to estimate the genetic potential of a new variety.

Results of this survey are consistent with Ryu et al. (2013) that found no outstanding differences between 48 US/Mexican landraces grown in Ohio and Arizona. Jing et al. (2007) found that purple corncobs grown in three locations around Lima, Peru did not vary significantly either, but across all locations in Peru, they were significantly different according to an ANOVA. Several factors varied across the locations in these two studies: precipitation, elevation, temperature, etc. Any combinations of these factors may have influenced anthocyanin production. From the data presented here and some support from the two other studies, it appears that the results of this study can extend to environments typical of the corn belt region of the US.

2.5 Conclusions

Maize is a diverse source of anthocyanins and has great potential as an economic source of natural colors. Here, 398 accessions were categorized based on the abundance of anthocyanins in either the pericarp or aleurone layers and on anthocyanin composition. This investigation represents the most comprehensive analysis of anthocyanin content in maize germplasm to date. Presented here is a simple and repeatable high-throughput method for analyzing anthocyanins in maize. Five categories of anthocyanin production were derived from the 167 anthocyanin-pigmented accessions in the survey. One category has previously never been described in maize; some accessions produced markedly less acylated anthocyanins than the typical anthocyanin-producing lines. The genetics behind this unique trait are currently being investigated (See Chapter 3). PCA was performed on compounds quantified by HPLC and compositional categories were confirmed by hierarchical clustering, confirming their efficacy. Since the main

goal of this study was to provide information to plant breeders, the category that should be of most interest for developing anthocyanin-rich hybrids is purple pericarp with condensed forms. Andean purple corn landraces may be a key to developing new purple corn hybrids, but issues with adaptation and grain yield need to be overcome for them to be more economical. To provide a broader range of hues in maize extracts for the food and beverage industry, *pr1* recessive alleles should be incorporated into these purple corn accessions. Breeding for TAC and anthocyanin composition will be straightforward, according to the H^2 results, since most of the phenotypic variance is controlled by genetic factors as opposed to environmental factors. TAC of maize on a whole kernel basis is much lower than TAC reported for fruit and vegetable juices, but maize has value-added benefits fruit and vegetable juices do not have (Wu et al., 2006). Current milling processes are able to concentrate pericarp fractions for a higher anthocyanin recovery (Somavat et al., 2016). Pericarp fractions can then be collected for anthocyanin extraction while the rest of the kernel can still be utilized for food, fuel, and feed. The addition of natural colors can add value to the specialty corn supply chain and potentially decrease natural colorant costs.

2.6 Acknowledgements

I would like to thank Laura Chatham for performing PCA and hierarchical clustering on the survey and her analysis of both. I want to thank her for also providing the biplot and dendrogram for Figure 2.7. In addition, I would like to thank Laura Chatham for the photos I used in Figures 2.1 and 2.6. Next I want to thank Dr. Megan West, Dr. Leslie West, and Dr. Kang-Mo Ku for designing the extraction and HPLC protocol and for analyzing the original source packets. In addition, thank you to Dr. Leslie West for performing the acidified methanol/formic acid comparison experiment and the repeated extraction experiment. Finally, I

want to thank Dr. Martin Bohn from the University of Illinois at Urbana-Champaign for assistance in calculating heritability.

2.7 Figures and Tables

Figure 2.1: Major anthocyanin-producing tissues in the maize kernel. *Left*: Cross-section of a maize kernel. Figure adapted from Coe (2001). *Right*: Microscopic sections of kernels demonstrating anthocyanin pigmentation in various kernel tissues. Black bars represent 100 microns.

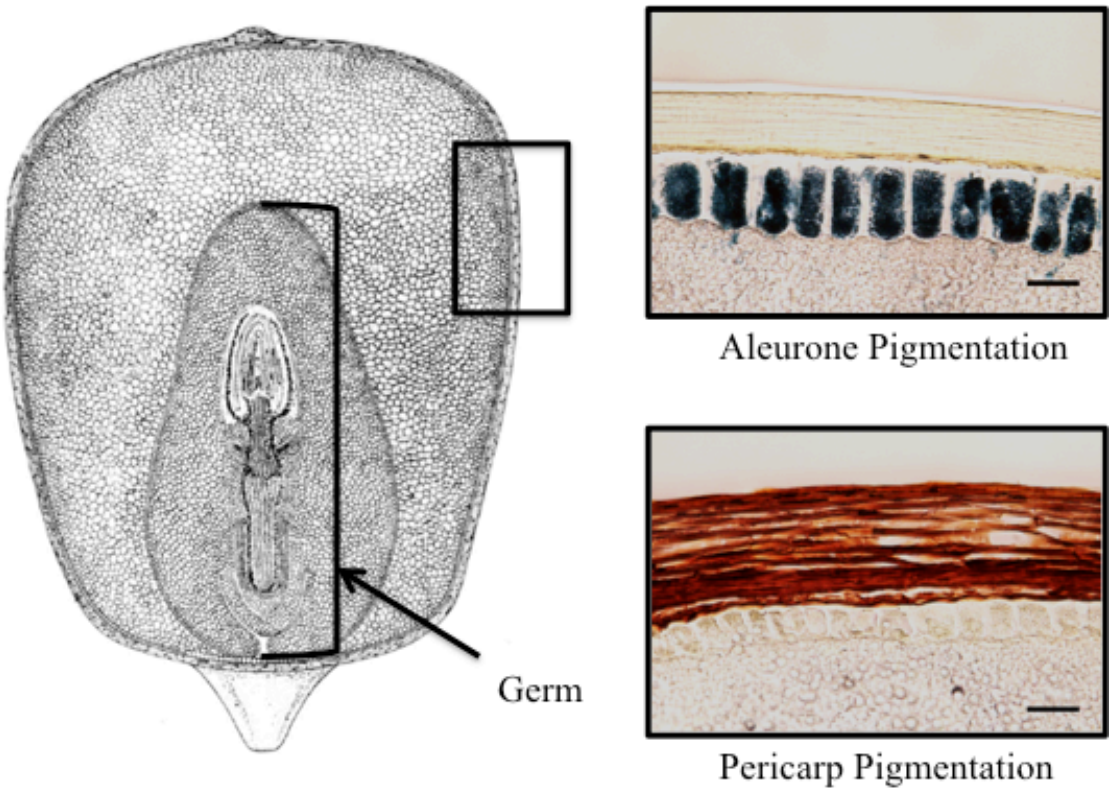


Figure 2.2: Structure of an anthocyanidin molecule with molecular nomenclature. R = H, Pelargonidin; R = OH, Cyanidin; R = OCH_3, Peonidin.

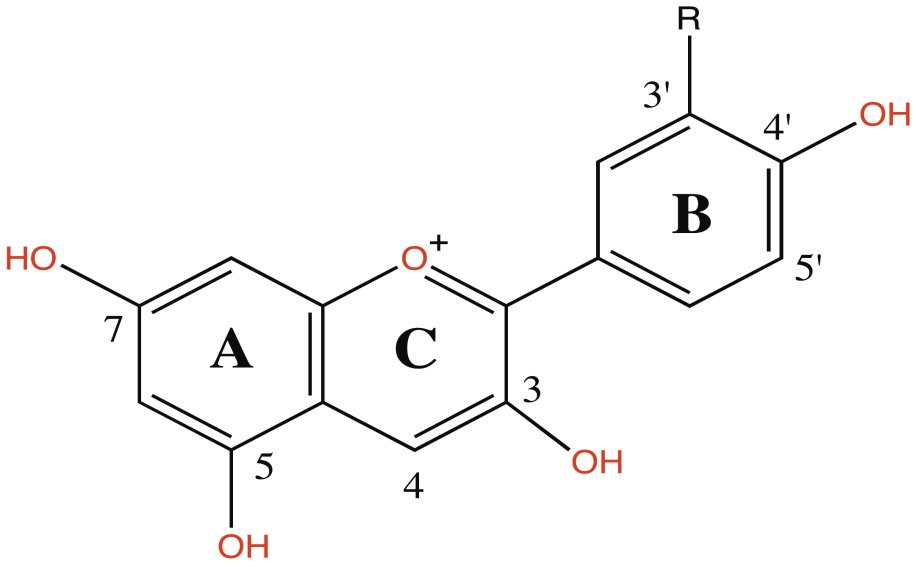


Figure 2.3: Anthocyanin biosynthetic pathway in maize with gene products.

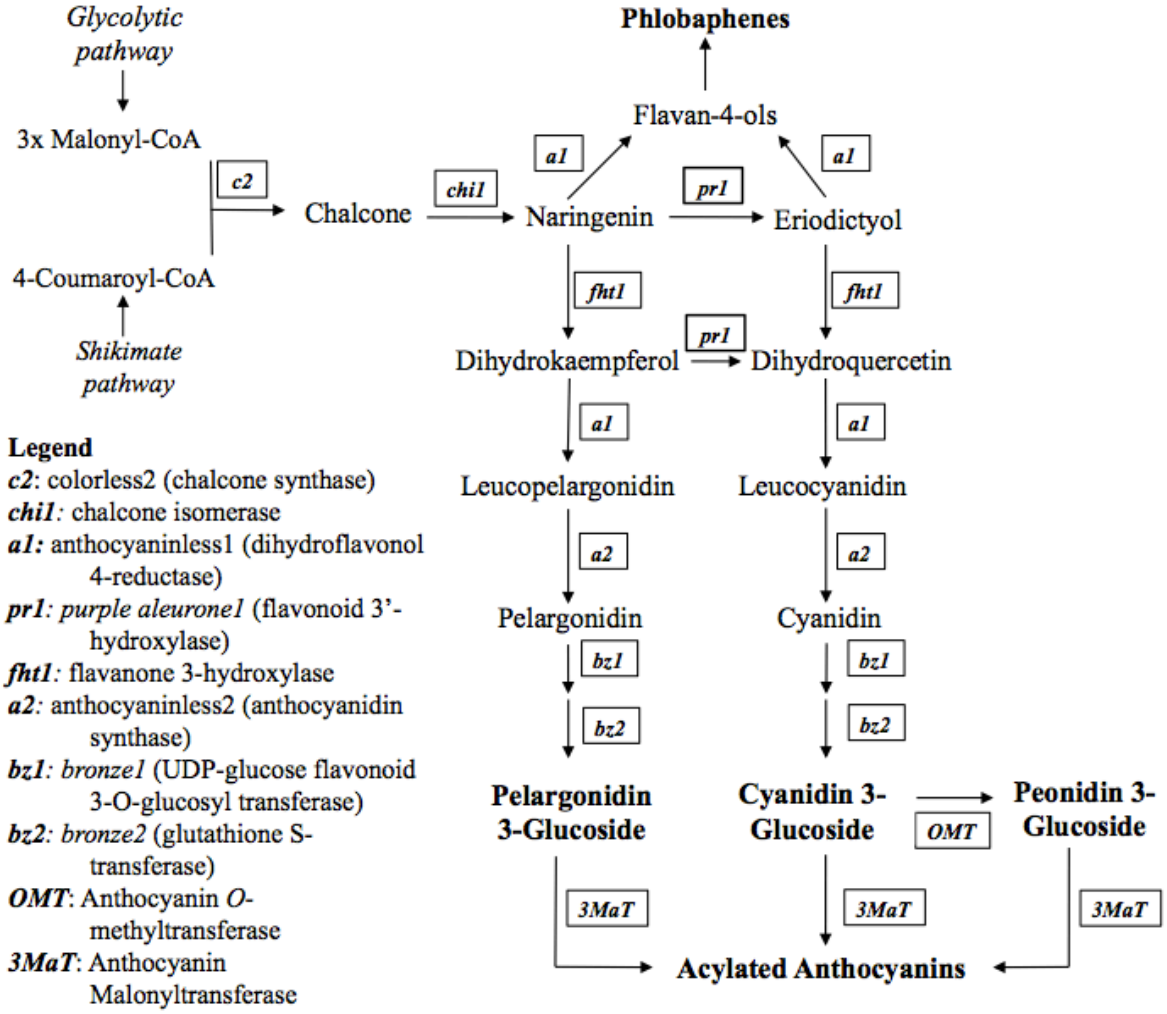
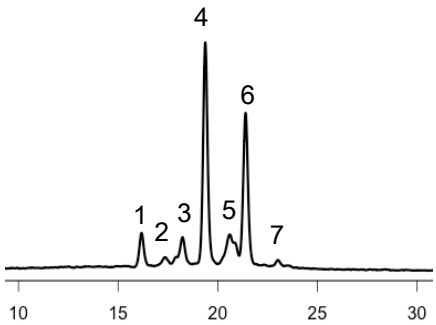
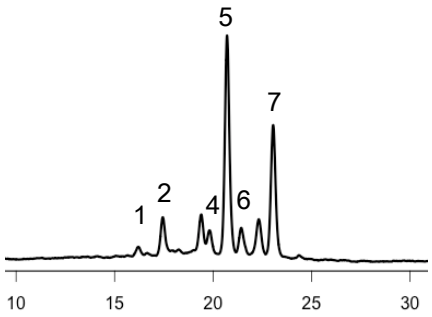


Figure 2.4: Representative chromatograms of each anthocyanin category

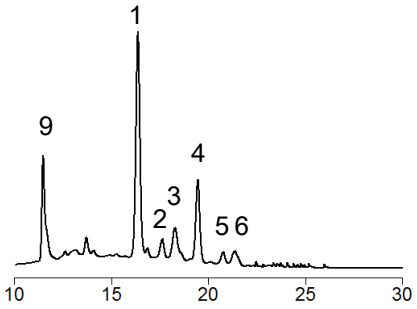
a) Blue aleurone



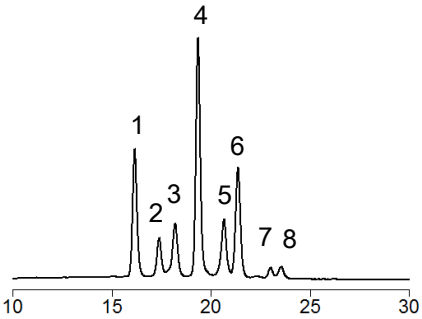
b) Pink aleurone



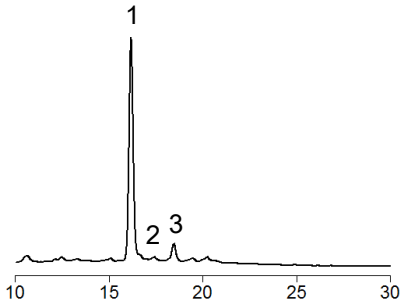
c) Pericarp with condensed forms



d) Pericarp without condensed forms



e) Reduced acylation



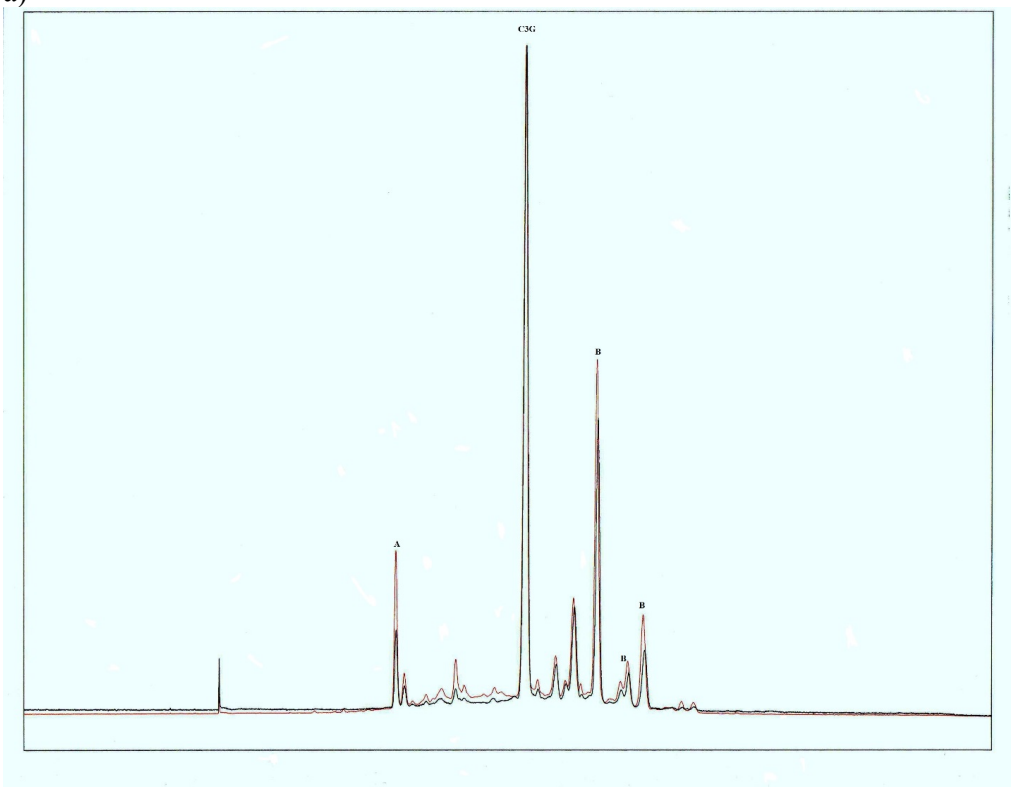
f) Legend for identified anthocyanin compounds

Label	Compound
1	Cyanidin 3-Glucoside
2	Pelargonidin 3-Glucoside
3	Peonidin 3-Glucoside
4	Cyanidin malonylglucoside
5	Pelargonidin malonylglucoside
6	Cyanidin dimalonylglucoside / Peonidin malonylglucoside
7	Pelargonidin dimalonylglucoside
8	Peonidin dimalonylglucoside
9	Catechin-4,8-Cyanidin-3,5-diglucoside

Note: The x-axis of chromatograms corresponds to retention time in minutes.

Figure 2.5: Comparison and efficiency of extraction methods. a) Combined 9 extractions with 2% (v/v) formic acid (red) compared with 6 extractions using acidified methanol (black); chromatograms normalized to largest peak (C3G). b) Percent of TAC obtained per extraction using the spectrophotometric method to determine TAC (Lee et al., 2005).

a)



b)

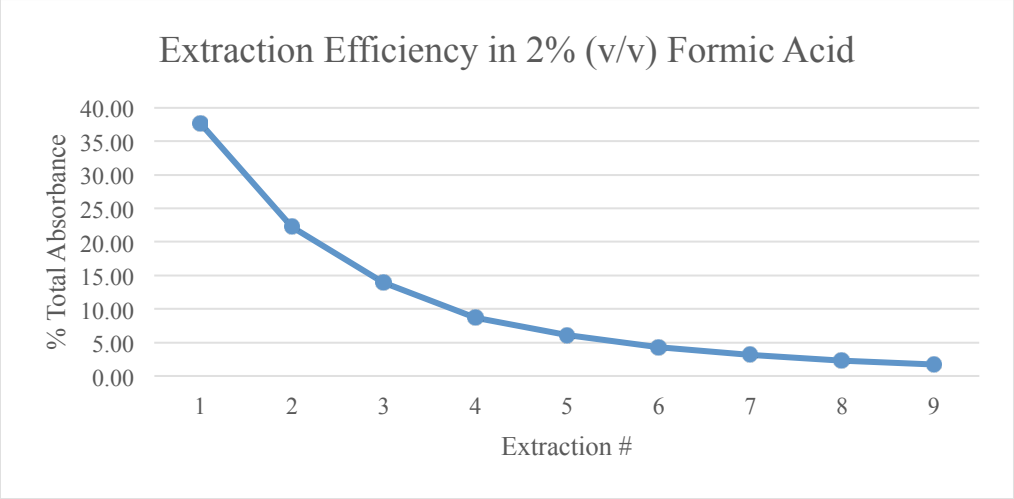
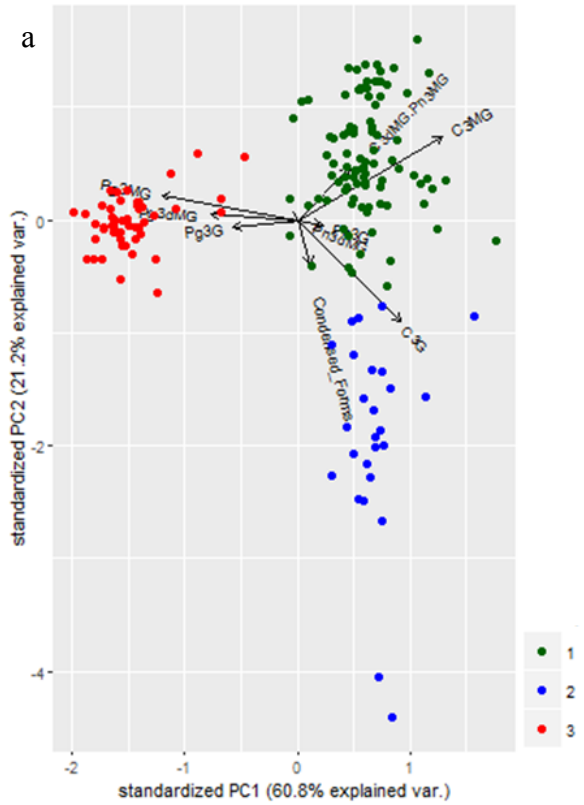


Figure 2.6: Representative kernel and anthocyanin extracts from a) blue aleurone, b) pink aleurone, c) pericarp pigmentation with condensed forms (*left*), pericarp pigmentation without condensed forms (*right*) and d) reduced acylation categories. Pericarp extracts appear similar.

Figure 2.7: PCA of all accessions in the collection with a) PC biplot and b) dendrogram produced from Ward's minimum variance method.



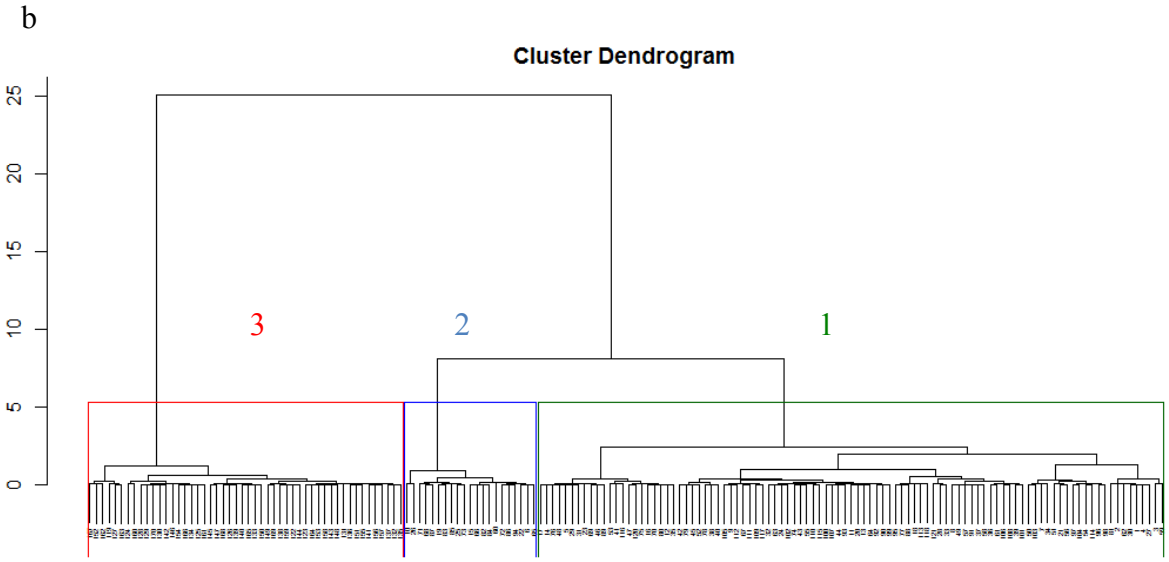


Table 2.1: Subset of accessions grown in 2015 labeled with major phenotypes

Accession	Pr1*	Tissue†	Acyl°	Count
Ames 6084 Z14-008	0	0	0	16
Ames 14276 Black Beauty	0	0	1	12
Ames 25207 Burford 2	0/1	0	0	4
Ames 27451 A632.75	0	0/1	0	15
BGEM-0085-N	0	0	0	13
Black Aztec	0	0	0/1	11
MGCSC 218GA	0	0	0	11
MGCSC 219AA	0	0	0	8
MGCSC 506B	1	0	0	12
MGCSC 707B	1	0	0	9
MGCSC 707G	0	0	1	14
MGCSC M142A	0	0	0	5
MGCSC X13l	0	1	0	14
MGCSC X19A	0	0	0	12
MGCSC X19EA	0	0	0	5
MGCSC Z433C	0	0	0	11
MGCSC Z433E	0	0	0	11
Siskiyou Seeds Hopi Blue Star	0	0	0	6
Ohio Blue Clarage	0	0	0	9
PI 213756 Fairfax Brown	0	0	0	11
PI 213791 SD RAINBOW	0	0	0	10
PI 217411 Tama Flint	1	0	0	11
PI 218175 Moencopi Pueblo	0	0	0	4
PI 340838 B-4541	0/1	0	0	9
PI 340841 B-15	0	0	0	11
PI 340846 RB-15	1	0	0	12
PI 340850 B-17	0	0	0	10
PI 340854 RB-17	1	0	0	11
PI 340855 B-18	0	0	0	12
PI 340857 B-22	0	0	0	12
PI 483476 Aguascalientes 27	0	0	0/1	10
PI 483517 Guanajuato 31	1	0	0	10
PI 483527 Guanajuato 98	1	0	0	7
PI 485071 Puebla 403	1	0/1	0	4
PI 489081 Puebla 456	1	0/1	0	13
PI 511613 Nicaragua 115	1	0	0	12
PI 553055 OC2	0	0	0	11
PI 553057 OC4	0	0	0	10
PI 596502 OC15	0	0	0	11
PI 596503 OC16	0	0	0	4
PI 596504 OC17	1	0	0	12
PI 596505 OC18	1	0	0	11
PI 596506 OC19	0	1	0	10

* 0 = *Pr1__*, 1 = *pr1pr1*

† 0 = aleurone, 1=pericarp

° 0 = normal profile, 1 = reduced acylation

Count refers to the number of HPLC samples

Table 2.2: Summary of Anthocyanin Production Categories

Categories	Count of Accessions	TAC (mg/kg)	*	C3G (%)	*	Pg3G (%)	*	Pn3G (%)	*	Acylation (%)	*
Blue Aleurone	98	30.81	BC	17.91	C	3.82	B	6.58	B	63.19	A
Pericarp Condensed	25	251.97	A	28.26	B	4.41	B	7.98	B	35.69	B
Pericarp Not Condensed	13	118.29	B	22.14	BC	5.18	B	11.69	A	56.84	A
Pink Aleurone	46	22.63	BC	2.99	D	14.88	A	1.11	C	62.84	A
Reduced Acylation	7	46.36	BC	57.23	A	2.33	B	7.40	B	7.57	C
Grand Total	**167**	**64.67**		**17.40**		**6.63**		**5.82**		**56.97**	

* Means with the same letter are not significantly different with a Tukey HSD $p>0.05$

Table 2.3: Model Selection for predicting Pg3DMG

Peak Area Pg3DMG

Number in Model	R-Square	Adjusted R-Square	Variables in Model
1	0.8263	0.8252	Peak Area Pg3MG
1	0.8263	0.8252	mg Pg3MG
1	0.6565	0.6545	Peak area Pg3G
2	0.8757	0.8743	Peak area Pg3MG, Pg3MG %
2	0.8757	0.8743	mg Pg3MG, Pg3MG %
2	0.8445	0.8427	Pg3G Pg3MG

mg Pg3DMG

Number in Model	R-Square	Adjusted R-Square	Variables in Model
1	0.8263	0.8252	Peak Area Pg3MG
1	0.8263	0.8252	mg Pg3MG
1	0.6565	0.6545	Peak area Pg3G
2	0.8757	0.8743	Peak area Pg3MG, Pg3MG %
2	0.8757	0.8743	mg Pg3MG, Pg3MG %
2	0.8445	0.8427	Pg3G Pg3MG

Pg3DMG %

Number in Model	R-Square	Adjusted R-Square	Variables in Model
1	0.3573	0.3535	Acylation %
1	0.3053	0.3013	Pg3G %
1	0.0553	0.0498	Pg3MG %
2	0.5217	0.5161	Acylation %, Pg3mg %
2	0.3948	0.3877	acylation percent_Pg3G
2	0.3653	0.3578	Pg3G %, Pg3mg %
3	0.5321	0.5238	Acylation %, Pg3G %, Pg3MG %
3	0.5281	0.5197	Peak Area Pg3MG, acylation %, Pg3MG %
3	0.5238	0.5154	Peak Area Pg3G, acylation %, Pg3MG %

Table 2.4: ANOVA for mg of Pg3DMG regressed with mg Pg3MG

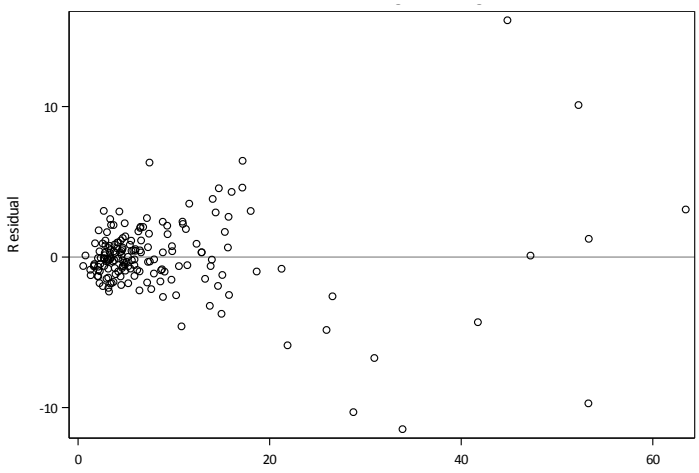
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	6117.09	6117.09	813.21	<.0001
Error	171	1286.28	7.52		
Corrected Total	172	7403.38			

Parameter Estimates

Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept	1	0.45708	0.27583	1.66	0.0993	-0.08738	1.00154
mg Pg3MG	1	0.56901	0.01995	28.52	<.0001	0.52962	0.60840

Root MSE	2.74
Dependent Mean	5.61
Coeff Var	48.92
R-Square	0.8263
Adj R-Sq	0.8252

Residual plot for mg Pg3DMG



mg Pg3DMG

Table 2.5: Correlations of compounds in varying *Pr1* backgrounds

***Pr1__* Accessions**									
	C3G	**Pg3G**	**Pn3G**	**C3MG**	**Pg3MG**	**C3DMG/ Pn3MG**	**Pg3DMG**	**Pn3DMG**	**TAC**
C3G	1	0.838	0.878	0.858	0.56	0.633	0.406	0.496	0.937
Pg3G	0.838	1	0.933	0.84	0.714	0.793	0.688	0.69	0.889
Pn3G	0.878	0.933	1	0.893	0.702	0.877	0.679	0.755	0.937
C3MG	0.858	0.84	0.893	1	0.843	0.86	0.633	0.759	0.972
Pg3MG	0.56	0.714	0.702	0.843	1	0.816	0.77	0.89	0.781
C3DMG/ Pn3MG	0.633	0.793	0.877	0.86	0.816	1	0.836	0.91	0.822
Pg3DMG	0.406	0.688	0.679	0.633	0.77	0.836	1	0.86	0.604
Pn3DMG	0.496	0.69	0.755	0.759	0.89	0.91	0.86	1	0.723
TAC	0.937	0.889	0.937	0.972	0.781	0.822	0.604	0.723	1

***pr1pr1* Accessions**									
	C3G	**Pg3G**	**Pn3G**	**C3MG**	**Pg3MG**	**C3DMG/ Pn3MG**	**Pg3DMG**	**Pn3DMG**	**TAC**
C3G	1	0.959	0.866	0.809	0.787	0.713	0.346	0.916	0.895
Pg3G	0.959	1	0.747	0.77	0.883	0.605	0.481	0.909	0.956
Pn3G	0.866	0.747	1	0.842	0.593	0.84	0.217	0.66	0.69
C3MG	0.809	0.77	0.842	1	0.791	0.888	0.59	0.571	0.825
Pg3MG	0.787	0.883	0.593	0.791	1	0.595	0.793	0.696	0.968
C3DMG/ Pn3MG	0.713	0.605	0.84	0.888	0.595	1	0.456	0.452	0.649
Pg3DMG	0.346	0.481	0.217	0.59	0.793	0.456	1	0.213	0.695
Pn3DMG	0.916	0.909	0.66	0.571	0.696	0.452	0.213	1	0.825
TAC	0.895	0.956	0.69	0.825	0.968	0.649	0.695	0.825	1

Table 2.6: PCA loadings and variability explained by the first two principal components

	PC1	PC2
Standard deviation	0.224	0.132
Proportion of Variance	0.608	0.212

Loadings:	PC1	PC2
Condensed_Forms		-0.297
C3G	0.406	-0.674
Pg3G	-0.252	
Pn3G		
C3MG	0.567	0.549
Pg3MG	-0.53	0.169
C3dMG.Pn3MG	0.207	0.349
Pg3dMG	-0.339	
Pn3dMG		

Table 2.7: Type 3 ANOVAs for 2015 and all environments for TAC and Percentage of Acylation

TAC 2015	DF	Sum of Squares	Mean Square	Expected Mean Square	Variance Components	F Value
Rep	2	99.28	49.64	Var(Residual)+33.5Var(Rep)	-2.69	0.36
Genotype	42	130433.00	3105.54	Var(Residual)+2.5476Var(Genotype)	1164.11	22.21 ***
Residual	65	9088.45	139.82	Var(Residual)	139.82	.

Acylation % 2015	DF	Sum of Squares	Mean Square	Expected Mean Square	Variance Components	F Value
Rep	2	173.48	86.74	Var(Residual)+33.5Var(Rep)	1.67	2.81*
Genotype	42	32583.00	775.80	Var(Residual)+2.5476Var(Genotype)	292.40	25.13***
Residual	65	2006.94	30.88	Var(Residual)	30.88	.

TAC All Env	DF	Sum of Squares	Mean Square	Expected Mean Square	Variance Components	F Value
Environment	2	8397.83	4198.92	Var(Residual) + 1.0644 Var(Year*Genotype) + 26.195 Var(Rep(Year)) + 43.724 Var(Year)	90.64	17.81***
Rep(Environment)	3	203.09	67.70	Var(Residual) + 24 Var(Rep(Year))	-6.09	0.32
Genotype	42	203740.00	4850.96	Var(Residual) + 1.2772 Var(Year*Genotype) + 3.6979 Var(Genotype)	1195.18	11.25***
Year* Environment	80	35729.00	446.62	Var(Residual) + 1.3667 Var(Year*Genotype)	170.38	2.09**
Residual	69	14749.00	213.75	Var(Residual)	213.75	.

Acylation % All Env	DF	Sum of Squares	Mean Square	Expected Mean Square	Variance Components	F Value
Environment	2	309.03	154.52	Var(Residual) + 1.0644 Var(Year*Genotype) + 26.195 Var(Rep(Year)) + 43.724 Var(Year)	1.96	2.24
Rep(Environment)	3	173.73	57.91	Var(Residual) + 24 Var(Rep(Year))	1.17	1.94
Genotype	42	48176.00	1147.04	Var(Residual) + 1.2772 Var(Year*Genotype) + 3.6979 Var(Genotype)	299.41	28.78***
Year* Environment	80	3244.76	40.56	Var(Residual) + 1.3667 Var(Year*Genotype)	7.88	1.36*
Residual	69	2055.10	29.78	Var(Residual)	29.78	.

*** $p<0.001$

** $p<0.01$

* $p<0.1$

Table 2.8: Average phenotypes per environment and heritability estimates with confidence intervals.

Compound(s)	Averages per Environment			H^2
	Source	2014	2015	
TAC mg/kg	33.54	62.63	40.09	0.9111
Acylation %	62.06	61.88	65.67	0.9653
C3G mg/kg	7.25	12.08	6.84	0.9673
C3MG mg/kg	9.75	13.70	10.29	0.9575
C3DMG + Pn3MG mg/kg	3.95	10.61	6.00	0.8408
Pg Anthocyanins %	28.26	26.68	30.09	0.9870
Pg3G mg/kg	1.91	2.53	2.02	0.8792
Pg3MG mg/kg	4.62	7.98	5.86	0.9015
Pg3DMG mg/kg	1.35	3.10	2.36	0.8720
Pn3G mg/kg	1.62	3.79	2.52	0.7459
Pn3DMG mg/kg	0.07	0.74	0.20	0.5196

2.8 Supplementary Tables

Table 2.9: Total list of anthocyanin pigmented accessions

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Blue Aleurone	**30.81**	**0.16**	**17.91**	**3.82**	**6.58**	**34.32**	**10.08**	**16.47**	**2.18**	**0.14**	**63.19**
Ames 27451 A632.75	85.20	0.00	9.00	3.19	8.54	33.58	8.30	30.10	1.36	1.64	74.98
MGCSC X19A	78.12	0.00	10.62	3.47	8.54	39.43	13.42	20.75	1.27	0.07	74.95
PI 340838 B-4541	73.98	0.48	17.97	7.28	4.24	29.62	16.80	12.78	4.34	0.02	63.55
NS ZM01-011	70.05	0.00	22.05	3.96	9.23	33.39	9.36	17.77	0.85	0.38	61.75
NS XX03-012	69.76	0.00	24.88	3.59	6.29	36.23	7.97	14.12	1.67	0.00	59.98
BGEM-0085-N	65.67	0.00	8.06	3.59	7.15	39.59	11.19	25.21	1.11	0.18	77.27
PI 596503 OC16	62.81	0.00	9.07	2.29	9.62	31.51	9.44	32.20	1.30	0.18	74.64
CIMMYT TLAX-732	61.81	0.00	12.71	3.35	4.16	28.45	4.03	16.56	5.20	0.00	54.23
MGCSC 219AA	60.88	0.00	22.39	3.54	4.50	36.40	9.88	8.59	3.45	0.64	58.95
PI 553057 OC4	59.38	0.00	7.72	3.66	7.27	40.70	6.19	23.20	0.46	0.10	70.64
NS XX03-013	57.86	0.00	40.76	3.33	7.21	31.01	5.57	8.52	0.27	0.00	45.37
CIMMYT MEXI13A	57.38	0.00	18.31	5.56	5.02	36.82	4.00	11.68	0.82	0.00	53.32
CIMMYT GUANGP22	56.67	0.00	15.39	4.60	5.42	35.11	13.47	15.25	2.68	0.00	66.51
MGCSC X19EA	52.85	2.62	37.92	2.78	13.35	18.64	2.65	7.62	0.64	0.15	29.69
CIMMYT MEX-950	49.64	0.00	23.20	6.65	4.78	23.36	4.35	14.96	3.13	0.00	45.79
PI 218175 Moencopi Pueblo	49.61	0.00	13.82	2.82	7.69	36.10	9.04	22.31	0.69	0.12	68.25
MGCSC Z433E	49.26	0.00	6.86	2.93	7.42	36.45	12.22	27.90	1.52	0.10	78.20
Ames 25207 Burford 2	46.77	0.00	6.91	4.12	6.37	30.84	18.93	20.20	7.13	0.08	77.18
CIMMYT PUE-1209	45.34	0.00	21.24	3.97	5.53	30.90	3.09	11.90	2.31	0.00	48.19
PI 213756 Fairfax Brown	45.21	0.00	9.40	3.65	6.69	42.67	7.75	19.23	0.30	0.07	70.01
CIMMYT MEX-974	45.00	0.00	23.44	6.93	4.03	25.92	3.80	10.23	3.22	0.00	43.18
CIMMYT MEX-767	43.54	0.00	16.01	4.54	3.50	26.91	14.14	10.98	8.10	0.00	60.13
Ames 4171 Squaw Corn	42.65	3.31	39.00	5.31	15.13	23.55	4.08	9.25	0.00	0.26	37.14
CIMMYT TLAX-672	42.22	0.00	14.24	3.72	5.03	35.16	3.32	14.73	1.33	0.00	54.54
SESE McCormack's Blue Giant	42.07	0.00	25.75	2.51	5.00	49.73	6.26	7.38	0.00	0.00	63.38
CIMMYT TLAX-759	40.87	0.00	16.85	5.36	8.90	35.13	8.89	13.21	0.74	0.00	57.96
CIMMYT TLAX-736	40.64	0.00	13.21	4.45	6.45	28.38	3.19	15.21	5.21	0.00	52.00
CIMMYT PUEB510-B	39.98	0.00	13.33	5.07	4.26	23.79	11.60	13.53	15.24	0.00	64.17
PI 483476 Aguascalientes 27	39.58	0.32	15.62	3.79	8.90	32.79	12.81	20.46	1.52	0.10	67.67
Ames 6090 Z03-015	38.23	0.00	17.99	3.30	6.75	31.30	8.06	22.48	2.40	0.16	64.39
NS ZM02-146	37.25	0.00	17.97	4.44	8.31	31.39	12.33	17.10	1.59	1.24	63.65
CIMMYT TLAX-650	37.22	0.00	16.26	5.04	13.77	29.87	3.39	13.10	2.76	0.00	49.11
CIMMYT HGO-524	37.01	0.00	24.49	7.00	5.22	29.08	7.22	12.22	1.27	0.00	49.79
CIMMYT MEXI639	36.64	0.00	17.69	4.09	7.31	30.64	2.75	13.89	2.20	0.00	49.49
MGCSC M142A	36.29	0.00	9.05	3.19	6.09	40.80	8.02	21.30	0.64	0.00	70.77

Table 2.9 Cont.

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Blue Aleurone Continued	**30.81**	**0.16**	**17.91**	**3.82**	**6.58**	**34.32**	**10.08**	**16.47**	**2.18**	**0.14**	**63.19**
CIMMYT TLAX-676	36.06	0.00	28.84	7.41	5.06	21.81	3.95	12.40	0.61	0.00	38.78
CIMMYT TLAX-783	35.91	0.00	14.49	3.48	7.32	28.42	12.23	17.27	3.67	0.00	61.59
PI 483517 Guanajuato 31	35.78	0.00	18.44	4.82	5.99	37.88	11.12	13.49	3.67	0.00	66.15
MGCSC M141A	35.64	0.00	16.27	4.54	13.13	29.72	10.45	22.18	0.83	2.19	65.37
PI 553060 OC7	35.19	2.86	36.90	2.82	17.79	7.64	8.49	1.99	0.00	0.00	18.12
CIMMYT MEX-807	33.89	0.00	20.28	6.05	5.73	34.58	4.11	13.57	0.56	0.00	52.83
CIMMYT MEX-821	33.71	0.00	12.63	3.76	3.86	31.13	13.30	15.21	4.18	0.00	63.83
Siskiyou Seeds Hopi Blue Star	33.59	1.16	16.92	2.42	8.88	30.43	11.48	23.43	1.47	0.58	67.38
Ames 6084 Z14-008	31.92	0.00	8.54	3.93	8.47	32.74	12.52	26.33	1.11	0.24	72.94
NS ZM02-147	31.40	0.00	19.10	6.87	5.76	33.21	12.89	13.46	0.75	0.00	60.31
NS ZM02-024	30.51	0.00	19.33	5.48	6.71	37.53	11.43	15.03	0.62	0.27	64.88
SSE Smoke Signals	29.17	0.00	8.55	3.87	4.23	29.73	18.58	18.80	11.27	0.00	78.38
MGCSC 218GA	28.71	0.00	13.38	3.51	9.36	34.44	13.32	19.93	3.41	0.09	71.19
PI 340855 B-18	27.52	0.00	11.42	3.09	7.15	40.12	11.24	19.91	1.39	0.02	72.69
PI 213787 Rainbow Flint	26.92	0.00	12.29	7.26	9.76	33.65	15.18	16.89	1.30	1.12	68.14
CIMMYT CHIH158-B	26.85	0.00	22.33	6.42	8.10	29.07	13.27	13.53	1.33	0.00	57.20
PI 619186 Coffee	26.23	0.00	14.15	1.89	3.37	46.24	12.87	18.22	0.00	0.00	77.33
PI 483303 SC PM	26.12	0.00	12.64	2.62	7.25	37.88	9.08	24.29	1.33	0.51	73.08
CIMMYT MAIZNEGRO	25.92	0.00	12.77	5.38	9.08	37.65	13.89	15.22	1.20	0.00	67.97
CIMMYT GUAT1130	25.55	0.00	19.79	4.69	3.98	33.92	12.26	11.30	1.01	0.00	58.49
PI 484092 Mihoacan 333	25.46	0.00	24.35	2.24	7.88	39.18	7.68	12.07	2.01	0.00	60.94
NS ZM02-178	23.85	0.00	22.90	5.00	4.26	41.82	12.85	10.53	0.56	0.00	65.76
PI 484505 Chihuahua 255	23.68	0.00	12.81	3.89	8.28	35.33	13.62	19.18	1.72	0.00	69.86
CIMMYT TLAX-629	23.04	0.00	12.98	4.86	4.16	33.13	13.57	10.63	1.01	0.00	58.34
NS ZM02-156B	22.67	0.00	20.74	5.20	6.02	31.41	13.12	12.36	2.15	0.00	59.04
PI 488987 Puebla 102	22.47	0.00	13.63	3.45	5.03	33.47	14.46	18.66	8.76	0.00	75.34
SESE Ohio Blue Clarage	22.12	0.00	8.31	2.09	7.98	36.62	8.87	25.91	0.77	0.02	72.19
PI 222309 N Dakota Colored Flour Bulk 5	20.21	0.00	20.68	2.70	4.75	32.83	13.57	12.27	1.43	0.00	60.10
PI 476866 Z03-001	19.41	0.00	15.93	0.97	5.46	45.76	8.57	22.27	0.62	0.00	77.23
MGCSC Z433C	18.31	0.00	8.43	2.45	7.83	40.04	9.08	26.41	0.20	0.04	75.76
PI 340841 B-15	18.17	0.00	7.10	2.37	7.47	36.74	8.83	26.87	5.12	0.00	77.56
PI 489081 Puebla 456	18.13	0.00	15.06	5.77	10.48	33.92	4.83	12.19	1.39	0.00	52.34
PI 213791 S. Dakota Rainbow	17.92	0.00	13.45	4.70	7.50	35.71	12.71	19.92	1.49	0.00	69.83
NSL 67059 ARIZONA 122	17.81	0.00	13.04	0.49	4.62	44.46	16.69	16.16	2.02	0.00	79.33
PI 217411 Tama Flint	17.66	0.00	17.69	4.30	7.13	30.77	14.99	16.33	1.50	0.00	63.58
NS ZM02-137	17.03	0.00	17.66	4.90	8.75	31.96	12.15	17.03	1.07	0.00	62.21
PI 488990 Puebla 148	16.89	0.00	22.82	1.01	8.51	45.07	5.34	10.59	2.60	0.00	63.60

Table 2.9 Cont.

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Blue Aleurone Continued	**30.81**	**0.16**	**17.91**	**3.82**	**6.58**	**34.32**	**10.08**	**16.47**	**2.18**	**0.14**	**63.19**
NS ZM02-141	16.71	0.00	19.50	5.35	8.46	35.93	11.14	12.45	0.84	0.00	60.35
SSE Black Aztec	15.53	0.00	36.37	2.60	6.29	25.34	9.37	9.35	1.51	0.00	45.58
PI 553055 OC2	15.12	0.00	7.46	3.02	8.07	34.47	14.11	29.75	1.14	0.00	79.47
CIMMYT TLAX-690	14.48	0.00	17.80	5.23	7.33	33.08	13.43	11.16	0.89	0.00	58.56
SSE Blue Jade	14.48	0.00	34.12	4.56	7.10	21.77	9.72	7.32	0.00	0.00	38.81
PI 553054 OC1	14.41	0.00	39.63	0.72	8.53	15.71	9.30	2.98	1.51	0.87	30.36
PI 340850 B-17	14.23	0.00	10.63	2.34	6.21	39.10	6.38	22.22	1.07	0.55	69.32
PI 317680 Black Beauty	13.78	0.00	16.89	3.12	5.22	39.12	13.01	19.44	0.24	0.00	71.81
White Buffalo Trading Co. Jerry Peterson Blue Dent	12.87	1.62	26.83	2.01	5.36	30.45	6.19	17.98	2.72	0.53	57.87
SESE Painted Mountain	11.81	0.00	9.18	10.15	3.65	23.06	27.51	14.40	8.04	0.51	73.53
PI 217478 Red Robin	11.55	0.00	19.06	4.62	11.77	41.57	7.36	13.83	1.40	0.40	64.56
Park Seeds Corn Fiesta Hybrid	10.91	0.00	17.44	3.46	3.86	38.82	10.28	15.85	1.01	0.00	65.96
PI 340857 B-22	10.52	0.00	13.27	2.64	4.95	32.63	13.09	21.05	2.82	0.03	69.61
PI 608600 CRC 47	10.23	0.00	19.45	3.11	2.95	45.12	8.68	13.99	1.95	0.00	69.74
NSL 91996 RAINBOW FLINT	8.66	0.00	9.48	1.40	2.66	32.35	19.53	22.01	9.85	0.00	83.74
PI 608602 CRC 48	8.30	0.00	17.26	1.32	1.97	34.55	13.59	29.97	0.00	0.00	78.11
SESE Pungo Creek Butcher	7.86	0.00	29.16	0.00	2.41	47.55	7.29	12.90	0.00	0.00	67.74
PI 445007 Magdalena 469	7.73	0.00	32.91	3.85	0.90	38.53	10.80	8.24	1.58	0.00	59.16
SESE Daymon Morgan's Kentucky Butcher	7.51	0.00	17.07	2.02	2.26	57.65	8.68	12.33	0.00	0.00	78.66
PI 596501 OC14	6.31	0.00	16.89	4.07	10.42	33.92	12.01	19.84	1.01	0.00	66.78
PI 596502 OC15	5.67	0.00	10.81	1.62	5.35	38.23	6.03	33.99	0.12	0.00	78.36
PI 532756 Masangu	5.64	0.00	20.06	3.39	3.83	44.15	10.37	2.43	0.00	0.00	56.95
PI 484768 San Luis Potosi 86	4.93	0.00	14.97	2.89	2.14	38.43	15.51	14.81	0.00	0.00	68.75
Ames 23683 Baby Blue	4.27	0.00	24.88	0.00	2.52	38.15	2.88	23.44	0.00	0.00	64.46
Ames 28291 Hi27 Goodman-Buckler	3.78	0.00	28.47	2.37	1.84	42.58	4.82	15.53	0.00	0.00	62.93
PI 490921 Nayarit 15	3.40	0.00	20.18	5.16	4.02	39.64	7.13	8.87	15.00	0.00	70.64
Pink Aleurone	**22.63**	**0.16**	**2.99**	**14.88**	**1.11**	**6.38**	**34.63**	**3.01**	**18.80**	**0.02**	**62.84**
MGCSC 707B	128.12	0.00	1.61	13.32	0.30	5.44	36.51	1.55	21.31	0.00	64.81
PI 340838 B-4541	64.79	0.00	1.16	13.20	0.09	2.90	48.99	0.54	18.80	0.00	71.24
CIMMYT NAYA191-B	45.44	0.00	7.62	17.66	6.64	17.68	25.52	8.34	4.72	0.25	56.52
MGCSC 506B	44.60	0.00	0.99	7.12	0.32	6.56	34.39	3.88	26.96	0.00	71.79
CIMMYT MEXI 323-#	41.55	0.00	4.16	14.23	3.98	10.90	24.15	7.03	18.85	0.00	60.93
CIMMYT PUEB,39-#	36.97	0.00	4.44	20.81	1.02	6.68	28.55	4.22	18.70	0.00	58.16
CIMMYT PUEB 149-#	36.86	1.02	6.30	26.08	4.22	6.26	25.54	4.06	13.50	0.00	49.36
PI 489081 Puebla 456	35.25	0.00	1.93	12.37	1.35	4.62	33.31	3.77	21.97	0.00	63.67
CIMMYT OAXA 121-#	31.44	0.00	2.69	25.29	0.30	5.84	33.30	1.06	15.37	0.00	55.56

Table 2.9 Cont.

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Pink Aleurone Continued	**22.63**	**0.16**	**2.99**	**14.88**	**1.11**	**6.38**	**34.63**	**3.01**	**18.80**	**0.02**	**62.84**
CIMMYT TLAX,252-#	30.64	0.00	3.81	26.01	0.36	6.91	35.87	1.19	13.13	0.22	57.32
PI 511613 Nicaragua 115	29.74	0.00	1.40	8.48	0.47	2.27	36.62	0.87	23.61	0.00	63.37
CIMMYT PUEB,252-#	29.17	0.00	2.99	21.76	0.48	6.10	31.88	3.91	16.59	0.00	58.49
Ames 25207 Burford 2	27.42	0.00	2.10	7.83	0.09	7.41	39.18	4.99	24.44	0.00	76.03
Ames 6084 Z14-008	26.95	4.52	2.24	5.75	0.41	4.93	28.64	6.12	28.44	0.00	68.13
PI 485071 Puebla 403	23.31	0.00	3.82	16.96	0.23	3.81	29.25	1.79	24.34	0.00	59.19
PI 273824 Tlexcala	23.16	0.00	1.83	16.83	0.00	5.29	41.09	2.36	13.21	0.00	61.95
CIMMYT GUAN 139-#	23.13	0.00	1.76	12.64	2.33	3.33	34.78	0.73	20.55	0.00	59.39
PI 483517 Guanajuato 31	19.90	0.00	2.32	14.33	0.61	3.77	35.51	1.71	18.86	0.00	59.85
CIMMYT CHIH441	18.17	0.00	6.13	8.78	1.46	9.21	33.20	5.97	20.34	0.00	68.71
CIMMYT OAXA,810-B	18.06	0.00	2.04	12.74	0.99	4.42	35.04	1.37	21.72	0.00	62.56
PI 596505 OC18	17.80	0.00	2.54	7.43	0.95	6.99	28.27	5.99	29.65	0.00	70.90
CIMMYT MEXI,244-#	17.13	0.00	2.85	12.73	0.81	5.55	28.47	3.75	21.59	0.00	59.37
CIMMYT QUER,94-B	17.07	0.00	1.57	16.81	1.53	4.60	37.04	0.82	16.06	0.00	58.52
CIMMYT JALI 633-#	17.03	0.00	3.38	6.70	0.85	12.38	29.16	10.88	22.75	0.00	75.18
CIMMYT QUER,94-#	17.00	0.00	1.18	8.47	0.69	3.23	41.49	1.07	28.44	0.00	74.23
SSE Smoke Signals	16.54	0.00	2.50	7.63	1.25	8.27	31.14	3.65	23.87	0.00	66.92
PI 596504 OC17	16.23	0.00	2.22	13.47	0.00	3.06	33.05	0.86	18.23	0.00	55.21
CIMMYT GUAN,336-#	16.05	0.00	1.59	8.36	1.69	1.93	32.99	3.36	31.03	0.00	69.31
CIMMYT SNLP,75-B	16.02	0.00	1.03	10.95	0.62	7.86	41.35	2.65	20.85	0.00	72.71
CIMMYT GUAN,336-B	15.42	0.00	1.42	14.25	3.21	4.07	36.45	3.94	11.80	0.00	56.26
CIMMYT DURA,215-#	14.61	0.00	2.98	10.30	1.31	5.51	38.61	1.73	20.90	0.00	66.75
PI 217411 Tama Flint	14.57	0.00	2.33	9.07	0.41	6.96	34.21	6.43	18.74	0.00	66.33
PI 483303 SC PM	14.18	0.00	0.87	6.64	0.00	5.92	39.43	0.00	26.00	0.00	71.35
PI 340854 RB-17	12.19	0.00	1.91	8.66	0.27	7.79	30.40	3.94	24.98	0.22	67.35
PI 489115 Puebla 498	10.04	0.00	6.62	28.63	1.45	6.32	31.92	1.71	14.31	0.00	54.25
Ames 8427 Cordoba 349	9.70	0.00	1.26	51.30	0.00	1.76	35.03	0.00	2.33	0.00	39.11
SESE Painted Mountain	9.42	0.00	6.09	14.79	5.31	18.15	37.05	3.65	10.76	0.00	69.61
PI 340846 RB-15	9.14	0.00	1.79	10.65	0.12	4.69	32.18	8.85	18.81	0.00	64.53
PI 483527 Guanajuato 98	8.65	0.35	1.99	9.87	0.09	2.16	33.54	4.20	35.50	0.00	75.40
PI 452064 Pinky Popcorn	8.40	0.00	3.59	9.90	0.00	0.00	34.63	0.00	24.69	0.00	59.32
PI 213787 Rainbow Flint	5.61	0.00	8.18	16.11	0.00	20.53	24.35	0.00	6.47	0.00	51.36
CIMMYT GUAN 438-B	5.51	0.00	3.29	13.78	3.67	6.27	44.23	3.51	15.74	0.00	69.76
PI 484768 San Luis Potosi 86	5.32	0.00	6.19	19.96	0.00	11.32	39.26	0.00	7.83	0.00	58.41
PI 490935 Puebla 219	4.88	0.00	2.56	16.88	0.00	11.85	44.86	0.00	7.74	0.00	64.44
CIMMYT GUAN 429-B	4.51	0.00	3.65	16.96	1.38	2.13	39.85	2.21	20.45	0.00	64.64
PI 490945 San Luis Potosi 22	3.32	0.00	2.81	32.01	0.00	0.00	42.55	0.00	0.00	0.00	42.55

Table 2.9 Cont.

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Pericarp Without Condensed Forms	**118.29**	**0**	**22.14**	**5.18**	**11.69**	**26.12**	**10.38**	**15.57**	**2.02**	**2.75**	**56.84**
PI 213730 Selection from Apache Red Cob	533.06	0.00	10.45	5.06	11.21	22.73	14.50	22.59	4.05	6.43	70.30
Double Red Sweet Corn	199.35	0.00	31.21	4.44	7.63	28.85	5.16	15.86	1.35	0.52	51.74
MGCSC X131	164.90	0.00	18.12	4.36	7.67	33.30	12.79	16.28	1.56	1.93	65.86
PI 435329 Kulli	164.51	0.00	28.12	5.04	14.70	20.89	8.56	15.05	1.13	2.61	48.23
Siskiyou Seeds Apache Red	149.81	0.00	16.59	3.42	9.94	21.23	13.74	16.07	2.50	2.45	56.00
PI 645583 Bronze Beauty	82.71	0.00	14.19	4.59	13.47	30.15	10.53	18.43	2.02	3.57	64.70
Siskiyou Seeds Apache Red	82.10	0.00	6.39	22.68	3.71	7.82	35.55	5.21	9.44	1.09	59.11
PI 596506 OC19	76.96	0.00	14.86	3.57	13.60	19.15	10.11	26.01	2.10	8.01	65.38
NSL 68325 ARIZONA 064	35.30	0.00	16.77	2.27	15.48	29.16	7.12	20.70	0.88	4.20	62.05
PI 484092 Mihoacan 333	17.69	0.00	22.74	4.77	9.63	33.85	8.16	16.05	1.17	1.62	60.85
PI 514702 Chococeno Negrito	13.90	0.00	32.16	0.84	2.94	46.61	4.65	7.88	0.00	0.00	59.14
NS ZM02-155	11.01	0.00	32.24	4.30	26.30	18.04	4.07	14.24	0.00	0.82	37.17
PI 485123 Chihuahua 171	6.50	0.00	43.96	2.05	15.65	27.75	0.00	8.08	0.00	2.52	38.34
Pericarp With Condensed Forms	**251.97**	**22.67**	**28.26**	**4.41**	**7.98**	**20.06**	**5.93**	**7.90**	**0.69**	**1.11**	**35.69**
PI 571427 Arequipa 204	1101.17	22.58	29.79	2.77	7.95	20.36	3.91	8.14	0.78	1.30	34.49
Kraft Proprietary	560.86	20.26	18.80	1.60	2.09	25.03	13.27	3.76	0.45	2.12	44.62
PI 514862 Arequipa 81	515.11	17.33	34.46	3.26	8.11	20.86	3.96	8.19	0.58	1.29	34.89
Kraft Proprietary	397.98	27.49	17.22	0.71	8.13	21.08	4.85	14.08	0.76	2.82	43.59
PI 514866 Arequipa 115	381.67	32.87	29.01	2.52	5.27	15.54	2.14	5.33	0.40	0.60	24.01
PI 390833 Maize Morado	380.80	31.22	33.45	5.30	8.47	14.45	2.88	3.59	0.14	0.16	21.23
PI 213730 Selection from Apache Red Cob	367.98	22.91	21.58	6.01	10.28	17.78	6.04	10.08	1.00	1.48	36.38
Ames 8488 Arequipa 35	358.47	24.00	38.69	4.48	10.11	13.67	2.22	4.18	0.76	0.00	20.83
Kraft Proprietary	334.10	21.27	19.20	1.44	5.87	24.31	5.67	14.44	0.53	2.04	46.98
PI 435329 Kulli	263.34	17.64	23.43	3.19	9.60	21.19	6.02	13.33	1.10	2.57	44.22
PI 571548 Lima 84	234.18	28.21	37.31	3.52	4.53	14.43	3.36	3.44	0.78	0.23	22.25
Kraft Proprietary	227.93	23.80	17.72	1.09	4.61	22.09	9.43	9.28	0.56	2.45	43.81
PI 514747 Ancash 14	225.47	24.52	35.29	3.25	9.40	16.55	2.55	5.76	0.00	0.61	25.46
Siskiyou Seeds Apache Red	220.50	20.92	5.28	23.20	1.59	3.81	25.96	1.51	4.81	2.35	38.45
PI 514746 Ancash 13	136.12	11.94	41.36	6.78	9.65	21.46	3.32	5.27	0.22	0.00	30.28
PI 273824 Morado	131.60	15.71	30.39	4.78	9.52	20.90	6.60	9.56	1.10	1.31	39.45
Siskiyou Seeds Apache Red	101.34	11.90	19.24	8.27	9.74	21.13	9.89	12.59	1.45	2.02	47.07
PI 514830 Ancash 460	82.51	26.03	43.10	2.23	6.71	15.91	1.71	3.89	0.00	0.00	21.50
PI 503660 Ancash 385	60.04	19.14	37.13	3.61	9.26	18.51	2.84	7.00	0.00	0.22	28.57
PI 574189 Ancash 383	56.30	20.62	30.77	7.00	10.26	19.47	5.53	5.75	0.29	0.18	31.22
PI 222297 Peruvian Red	46.69	3.82	13.04	2.65	13.70	32.53	8.02	21.41	0.83	3.50	66.29
CIMMYT GUAN 139-#	44.23	23.67	31.51	2.43	6.34	22.34	3.07	5.65	0.67	0.27	32.00

Table 2.9 Cont.

Name *	TAC mg/kg	Condensed forms %	C3G %	Pg3G %	Pn3G %	C3MG %	Pg3MG %	C3DMG/ Pn3MG %	Pg3DMG %	Pn3DMG %	Proportion Acylation
Pericarp With Condensed Forms Continued	**251.97**	**22.67**	**28.26**	**4.41**	**7.98**	**20.06**	**5.93**	**7.90**	**0.69**	**1.11**	**35.69**
PI 514750 Ancash 51	33.62	10.81	38.61	5.25	12.91	19.62	2.74	7.50	0.00	0.38	30.25
PI 484092 Michoacan 333	31.07	7.01	28.29	2.74	12.93	26.58	4.80	13.76	0.00	0.00	45.13
PI 503650 Ancash 49	6.19	21.90	31.77	2.18	2.40	31.98	7.41	0.00	0.00	0.00	39.38
Reduced Acylation	**46.36**	**0.04**	**57.23**	**2.33**	**7.40**	**4.55**	**2.57**	**0.36**	**0.08**	**0.01**	**7.57**
MGCSC 707G	132.61	6.66	72.03	1.53	3.44	1.45	1.02	0.08	0.01	0.03	2.57
PI 483476 Aguascalientes 27	51.54	0.00	44.92	4.38	11.38	0.72	0.00	0.00	0.00	0.00	0.72
Ames 14276 Black Beauty	49.39	3.34	66.75	3.66	5.15	1.59	1.77	0.68	0.05	0.04	4.14
MGCSC X19EA	31.75	5.13	51.86	1.31	11.20	9.09	4.70	1.75	0.00	0.00	15.55
PI 553060 OC7	26.53	0.00	42.21	0.79	14.01	8.94	4.49	0.00	0.00	0.00	13.43
White Buffalo Trading Co. Jerry Peterson Blue Dent	25.98	7.65	57.80	2.82	5.61	5.47	6.01	0.00	0.48	0.00	11.96
SSE Black Aztec	6.72	0.00	65.00	1.82	1.03	4.62	0.00	0.00	0.00	0.00	4.62
Grand Total	**64.67**	**0.12**	**17.40**	**6.63**	**5.82**	**23.97**	**15.25**	**11.40**	**5.94**	**0.41**	**56.97**

*BGEM = Iowa Genetic Enhancement of Maize (Ames, IA, USA), CIMMYT = International Maize and Wheat Improvement Center (El Batán, Edo Mex, Mexico), NS = Native Seeds/SEARCH (Tucson, AZ, USA), SESE = Southern Exposure Seed Exchange (Mineral, VA, USA), SSE = Seed Savers Exchange (Decorah, IA, USA)

Table 2.10: List of phlobaphene and bronze pigmented accessions in the collection

Phlobaphene Pigmented Accessions	
BGEM-0156-S	NCRPIS PI 172600 (IGDIR, KARS)
BGEM-0164-S	NCRPIS PI 175976 ITALYAN
BGEM-0188-S	NCRPIS PI 213730 Selection from Apache Red Cob
BGEM-0196-N	NCRPIS PI 213775 Fairfax Brown
BGEM-0247-N	NCRPIS PI 213776 Northwestern Dent
CIMMYT BRAZ MS050-B	NCRPIS PI 213787 Rainbow Flint
CIMMYT BRAZ,1185-#	NCRPIS PI 213788 Northwestern Dent
CIMMYT BRVI,138-#	NCRPIS PI 213791 SOUTH DAKOTA RAINBOW
CIMMYT BRVI,140-B	NCRPIS PI 213796 Nueta Sweet Corn
CIMMYT JALI 707-#	NCRPIS PI 214198 Northwestern Dent (Brandon Strain)
CIMMYT JALI,188-B	NCRPIS PI 214273 Northwestern Dent
CIMMYT URUG,177A-#	NCRPIS PI 217404 Argentine Popcorn
CIMMYT URUG,177A-B	NCRPIS PI 217460 King Philip
CIMMYT URUG,228A-#	NCRPIS PI 217462 Rainbow
CIMMYT URUG,254A-#	NCRPIS PI 217477 Red Robin
CIMMYT URUG,254A-B	NCRPIS PI 217478 Red Robin
CIMMYT URUG,325A-#	NCRPIS PI 217479 Red Field
CIMMYT URUG,325A-B	NCRPIS PI 217480 Northwestern Dent
Everwilde Farms Ruby Red	NCRPIS PI 218147 Mesita Pueblo
MGCSC 107B	NCRPIS PI 219885 Northwestern
MGCSC 6502D	NCRPIS PI 222297 Peruvian Red
NCRPIS Ames 10602 Lambayeque 1	NCRPIS PI 222319 Red Meadowbrook Reid
NCRPIS Ames 10603 Lambayeque 6	NCRPIS PI 251884 BELOYHARVE PSHENO
NCRPIS Ames 10626 Libertad 10	NCRPIS PI 340835 BR-1-1
NCRPIS Ames 10628 Libertad 14	NCRPIS PI 340837 R-4524
NCRPIS Ames 16950 PCM-WCG 910	NCRPIS PI 340838 B-4541
NCRPIS Ames 17352 PCM-WCG 1353	NCRPIS PI 340840 R-Strawberry Open Pollinated
NCRPIS Ames 21960 Red Popcorn	NCRPIS PI 340843 CA-15-2 DENT STERILE
NCRPIS Ames 22463 Martinique 3	NCRPIS PI 340844 R-15 DENT STERILE
NCRPIS Ames 25207 Burford 2	NCRPIS PI 340845 BR-15 DENT STERILE
NCRPIS Ames 2754 King Phillip's RI Flint	NCRPIS PI 340853 R-17 INBR.FR.SUPERGOLD
NCRPIS Ames 2758 Millersburg Red Sweet No. 2	NCRPIS PI 340856 R-18 INBR.FR.SUPERGOLD
NCRPIS Ames 2768 G 11866	NCRPIS PI 340859 R-22 INBR.FR.SUPERGOLD
NCRPIS Ames 2769 Dark Yellow King Philip	NCRPIS PI 340861 BR-28 INBR.FR.YEL.PEARL
NCRPIS Ames 2770 Golden Bantam Strain	NCRPIS PI 340862 R-28
NCRPIS Ames 28291 Hi27 Goodman-Buckler	NCRPIS PI 340863 R-28
NCRPIS Ames 30532 MR18 (Reventador) S6	NCRPIS PI 340866 R-53-1 INBR.FR.YELPEARL
NCRPIS Ames 4171 Squaw Corn	NCRPIS PI 340867 R-53-2 INBR.FR.YELPEARL
NCRPIS Ames 8458 Ancash 21	NCRPIS PI 340868 RB-53 INBR.FR.YELPEARL
NCRPIS NSL 30056 W703	NCRPIS PI 340869 R-PAH-1 IN.FR.TOM-THUMB
NCRPIS NSL 42706 Chile 314	NCRPIS PI 340870 R-PAH-2
NCRPIS NSL 42708 Chile 316	NCRPIS PI 340872 R-PAH-1
NCRPIS NSL 437868 AusTRCF 305772	NCRPIS PI 340873 R-RAH-2 IN.FR.TOM-THUMB
NCRPIS NSL 67055 ARIZONA 118	NCRPIS PI 340873 R-RAH-2 IN.FR.TOM-THUMB
NCRPIS NSL 67059 ARIZONA 122	NCRPIS PI 393726 VIII/26
NCRPIS NSL 68325 ARIZONA 064	NCRPIS PI 414184 STRAWBERRY DENT
NCRPIS NSL 91996 RAINBOW FLINT	NCRPIS PI 435329 Kulli
NCRPIS PI 170881 (CAYIROVA, KOCAELI)	NCRPIS PI 443827 Antioquia 401
NCRPIS PI 171916 (NIKSAR,TOKAT)	NCRPIS PI 443854 Antioquia 437

Table 2.10 Cont.

Phlobaphene Pigmented Accessions, Continued	
NCRPIS PI 444486 Cordoba 334	NCRPIS PI 515091 San Martin 114
NCRPIS PI 444492 Cordoba 342	NCRPIS PI 520691 439
NCRPIS PI 444543 Cundinamarca 366	NCRPIS PI 531469
NCRPIS PI 444785 Guajira 308	NCRPIS PI 531476 TIMARI VOROS LOFOGU
NCRPIS PI 445007 Magdalena 469	NCRPIS PI 531492 SZOREGI 100 NAPOS
NCRPIS PI 451692 Cargill N. Temp. Zone Coroico	NCRPIS PI 532310 7114
NCRPIS PI 452031 (Pod Corn, Ottumwa, Iowa)	NCRPIS PI 532312 7132
NCRPIS PI 452036 Brindle corn	NCRPIS PI 532329 7451
NCRPIS PI 452044 Red Dent	NCRPIS PI 542687 Kabl
NCRPIS PI 452062 Ames 1790	NCRPIS PI 553059 OC6
NCRPIS PI 483303 SC PM	NCRPIS PI 553061 OC8
NCRPIS PI 484506 Chihuahua 256	NCRPIS PI 553062 OC9
NCRPIS PI 485171 Ancash 453	NCRPIS PI 571814 Lambayeque 13
NCRPIS PI 485247 Apurimac 137	NCRPIS PI 571962 Ancash 188
NCRPIS PI 485329 Libertad 28	NCRPIS PI 572201 Piura 149
NCRPIS PI 488493 Madre de Dios 5	NCRPIS PI 596499 OC12
NCRPIS PI 488672 Ancash 622	NCRPIS PI 596500 OC13
NCRPIS PI 488674 Ancash 624	NCRPIS PI 608458 NEWTON RED
NCRPIS PI 490774 5491	NCRPIS PI 608599 Bloody Butcher
NCRPIS PI 490777 5507	NCRPIS PI 608643 Calico
NCRPIS PI 490921 Nayarit 15	NCRPIS PI 608654 Rainbow Flint
NCRPIS PI 503428 KABA-Langan	NCRPIS PI 614735 EPM6
NCRPIS PI 503716 Lambayeque 10	NCRPIS PI 619186 Coffee
NCRPIS PI 503729 Lambayeque 43	NCRPIS PI 619187 Red
NCRPIS PI 503739 Lambayeque 56	NS ZM02-141
NCRPIS PI 503773 Lambayeque 114	Park Seed Fiesta Hybrid
NCRPIS PI 503806 Piura 144	Seeds of Change Dakota Black Popcorn
NCRPIS PI 503810 Piura 148	SESE Bloody Butcher
NCRPIS PI 503837 Piura 180	SESE Daymon Morgan's Kentucky Butcher
NCRPIS PI 503844 Piura 196	SESE Floriani Red Flint
NCRPIS PI 503849 Piura 208	SESE Painted Mountain
NCRPIS PI 504296 Lambayeque 154	SESE Pungo Creek Butcher
NCRPIS PI 511498 Brazil 1185-AF	Siskiyou Seeds Apache Red

Bronze Pigmented Accessions	
MGCSC Ch1 2007-2721-5 X (Ch1) X	NCRPIS PI 340851 BR-17 INBR.FR.SUPERGOLD
NCRPIS Ames 21974 Parker's Flint	NCRPIS PI 420245 Chapalote
NCRPIS NSL 283388 Sinaloa 2	NCRPIS PI 485330 Libertad 34
NCRPIS PI 213742 Oklahoma Bronze	NCRPIS PI 553058 OC5
NCRPIS PI 213748 Concho Brown	Seed Savers Exchange Smoke Signals
NCRPIS PI 340845 BR-15 DENT STERILE	

Abbreviations Used: BGEM = Iowa Genetic Enhancement of Maize; CIMMYT = International Maize and Wheat Improvement Center, El Batán, Edo Mex, Mexico*;* MGCSC = Maize Genetics Cooperation Stock Center, Urbana, IL, USA*;* NCRPIS = North Central Regional Plant Introduction Station, Ames, IA, USA*;* NS = Native Seeds/SEARCH database, Tucson, AZ, USA*;* SESE = Southern Exposure Seed Exchange, Mineral, VA, USA.

2.9 References

Abdel-Aal, E.-S. M.; Young, J. C.; Rabalski, I. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J. Agric. Food Chem.* **2006**, *54* (13), 4696–4704.

Bernardo, R. N. *Breeding for Quantitative Traits in Plants*, 2nd ed.; Stemma Press: Woodbury, Minn, 2010.

Bhatla, S.; Pant, R. Isolation and Characterisation of Anthocyanin Pigment from Phosphorus-Deficient Maize Plants. *Curr. Sci.* **1977**, *46*, 700–702.

Cabrita, L.; Fossen, T.; Andersen, Ø. M. Colour and Stability of the Six Common Anthocyanidin 3-Glucosides in Aqueous Solutions. *Food Chem.* **2000**, *68* (1), 101–107.

Center for Food Safety and Applied Nutrition. Color Certification Reports - Report on the Certification of Color Additives: 4th Quarter, Fiscal Year 2014, July 1-September 30 http://www.fda.gov/ForIndustry/ColorAdditives/ColorCertification/ColorCertificationReports/ucm418381.htm (accessed Sep 5, 2016).

Cuevas Montilla, E.; Hillebrand, S.; Antezana, A.; Winterhalter, P. Soluble and Bound Phenolic Compounds in Different Bolivian Purple Corn (Zea Mays L.) Cultivars. *J. Agric. Food Chem.* **2011**, *59* (13), 7068–7074.

Deboo, G. B.; Albertsen, M. C.; Taylor, L. P. Flavanone 3-Hydroxylase Transcripts and Flavonol Accumulation Are Temporally Coordinate in Maize Anthers. *Plant J. Cell Mol. Biol.* **1995**, *7* (5), 703–713.

González-Manzano, S.; Pérez-Alonso, J. J.; Salinas-Moreno, Y.; Mateus, N.; Silva, A. M. S.; de Freitas, V.; Santos-Buelga, C. Flavanol–anthocyanin Pigments in Corn: NMR Characterisation and Presence in Different Purple Corn Varieties. *J. Food Compos. Anal.* **2008**, *21* (7), 521–526.

Grotewold, E.; Peterson, T. Isolation and Characterization of a Maize Gene Encoding Chalcone Flavonone Isomerase. *Mol. Gen. Genet.* **1994**, *242* (1), 1–8.

Grotewold, E.; Chamberlin, M.; Snook, M.; Siame, B.; Butler, L.; Swenson, J.; Maddock, S.; St Clair, G.; Bowen, B. Engineering Secondary Metabolism in Maize Cells by Ectopic Expression of Transcription Factors. *Plant Cell* **1998**, *10* (5), 721–740.

Han, Y.; Ding, T.; Su, B.; Jiang, H. Genome-Wide Identification, Characterization and Expression Analysis of the Chalcone Synthase Family in Maize. *Int. J. Mol. Sci.* **2016**, *17* (2), 161.

Hatier, J.-H. B.; Gould, K. S. Anthocyanin Function in Vegetative Organs. In *Anthocyanins*; Winefield, C., Davies, K., Gould, K., Eds.; Springer New York: New York, NY, 2008; pp 1–19.

He, J.; Giusti, M. M. Anthocyanins: Natural Colorants with Health-Promoting Properties. *Annu. Rev. Food Sci. Technol.* **2010**, *1*, 163–187.

Irani, N. G.; Hernandez, J. M.; Grotewold, E. Chapter Three Regulation of Anthocyanin Pigmentation. *Recent Adv. Phytochem.* **2003**, *37*, 59–78.

Jing, P.; Noriega, V.; Schwartz, S. J.; Giusti, M. M. Effects of Growing Conditions on Purple Corncob (*Zea Mays* L.) Anthocyanins. *J. Agric. Food Chem.* **2007**, *55* (21), 8625–8629.

Koes, R. E.; Quattrocchio, F.; Mol, J. N. The Flavonoid Biosynthetic Pathway in Plants: Function and Evolution. *BioEssays* **1994**, *16* (2), 123–132.

Lao, F.; Giusti, M. M. Quantification of Purple Corn (Zea Mays L.) Anthocyanins Using Spectrophotometric and HPLC Approaches: Method Comparison and Correlation. *Food Anal. Methods* **2015**, *9* (5), 1367–1380.

Larson, R.; Bussard, J.; Coe Jr, E. Gene-Dependent Flavonoid 3′-Hydroxylation in Maize. *Biochem. Genet.* **1986**, *24* (7–8), 615–624.

Lee, J.; Durst, R. W.; Wrolstad, R. E. Determination of Total Monomeric Anthocyanin Pigment Content of Fruit Juices, Beverages, Natural Colorants, and Wines by the pH Differential Method: Collaborative Study. *J. AOAC Int.* **2005**, *88* (5), 1269–1278.

Lev-Yadun, S.; Gould, K. S. Role of Anthocyanins in Plant Defence. In *Anthocyanins*; Winefield, C., Davies, K., Gould, K., Eds.; Springer New York: New York, NY, 2008; pp 22–28.

Marrs, K. A.; Alfenito, M. R.; Lloyd, A. M.; Walbot, V. A Glutathione S-Transferase Involved in Vacuolar Transfer Encoded by the Maize Gene Bronze-2. *Nature* **1995**, *375* (6530), 397–400.

Moreno, Y. S.; Sánchez, G. S.; Hernández, D. R.; Lobato, N. R. Characterization of Anthocyanin Extracts from Maize Kernels. *J. Chromatogr. Sci.* **2005**, *43* (9), 483–487.

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.; et al. A Genome-Wide Regulatory Framework Identifies Maize Pericarp Color1 Controlled Genes. *Plant Cell* **2012**, *24* (7), 2745–2764.

Nakayama, T.; Suzuki, H.; Nishino, T. Anthocyanin Acyltransferases: Specificities, Mechanism, Phylogenetics, and Applications. *J. Mol. Catal. B Enzym.* **2003**, *23* (2–6), 117–132.

de Pascual-Teresa, S.; Sanchez-Ballesta, M. T. Anthocyanins: From Plant to Health. *Phytochem. Rev.* **2008**, *7* (2), 281–299.

Petroni, K.; Pilu, R.; Tonelli, C. Anthocyanins in Corn: A Wealth of Genes for Human Health. *Planta* **2014**, *240* (5), 901–911.

R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

Ryu, S. H.; Werth, L.; Nelson, S.; Scheerens, J. C.; Pratt, R. C. Variation of Kernel Anthocyanin and Carotenoid Pigment Content in USA/Mexico Borderland Land Races of Maize. *Econ. Bot.* **2013**, *67* (2), 98–109.

Salinas-Moreno, Y.; Pérez-Alonso, J. J.; Vázquez-Carrillo, G.; Aragón-Cuevas, F.; Velázquez-Cardelas, G. A. Antocianinas Y Actividad Antioxidante En Maíces (Zea Mays L.) de Las Razas Chalqueño, Elotes Cónicos Y Bolita. *Agrociencia* **2012**, *46* (7), 693–706.

Sharma, M.; Cortes-Cruz, M.; Ahern, K. R.; McMullen, M.; Brutnell, T. P.; Chopra, S. Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. *Genetics* **2011**, *188* (1), 69–79.

Sharma, M.; Chai, C.; Morohashi, K.; Grotewold, E.; Snook, M. E.; Chopra, S. Expression of Flavonoid 3'-Hydroxylase Is Controlled by P1, the Regulator of 3-Deoxyflavonoid Biosynthesis in Maize. *BMC Plant Biol.* **2012**, *12* (1), 1.

Soave, C.; Salamini, F. The Role of Structural and Regulatory Genes in the Development of Maize Endosperm. *Dev. Genet.* **1984**, *5* (1), 1–25.

Somavat, P.; Li, Q.; de Mejia, E. G.; Liu, W.; Singh, V. Coproduct Yield Comparisons of Purple, Blue and Yellow Dent Corn for Various Milling Processes. *Ind. Crops Prod.* **2016**, *87*, 266–272.

Suzuki, H.; Nakayama, T.; Yonekura-Sakakibara, K.; Fukui, Y.; Nakamura, N.; Yamaguchi, M.; Tanaka, Y.; Kusumi, T.; Nishino, T. cDNA Cloning, Heterologous Expressions, and Functional Characterization of Malonyl-Coenzyme A:Anthocyanidin 3-O-Glucoside-6"-O-Malonyltransferase from Dahlia Flowers. *Plant Physiol.* **2002**, *130* (4), 2142–2151.

Suzuki, H.; Nakayama, T.; Yamaguchi, M.; Nishino, T. cDNA Cloning and Characterization of Two Dendranthema X Morifolium Anthocyanin Malonyltransferases with Different Functional Activities. *Plant Sci.* **2004**, *166* (1), 89–96.

Vivar-Quintana, A. M.; Santos-Buelga, C.; Francia-Aricha, E.; Rivas-Gonzalo, J. C. Formation of Anthocyanin-Derived Pigments in Experimental Red Wines. *Food Sci. Technol. Int.* **1999**, *5* (4), 347–352.

Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58* (301), 236.

White, W. G.; Vincent, M. L.; Moose, S. P.; Below, F. E. The Sugar, Biomass and Biofuel Potential of Temperate by Tropical Maize Hybrids. *GCB Bioenergy* **2012**, *4* (5), 496–508.

Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2009.

Wilmouth, R. C.; Turnbull, J. J.; Welford, R. W. D.; Clifton, I. J.; Prescott, A. G.; Schofield, C. J. Structure and Mechanism of Anthocyanidin Synthase from Arabidopsis Thaliana. *Structure* **2002**, *10* (1), 93–103.

Wolf, M. J.; Cutler, H. C.; Zuber, M. S.; Khoo, U. Maize with Multilayer Aleurone of High Protein Content. *Crop Sci.* **1972**, *12* (4), 440.

Wu, X.; Beecher, G. R.; Holden, J. M.; Haytowitz, D. B.; Gebhardt, S. E.; Prior, R. L. Concentrations of Anthocyanins in Common Foods in the United States and Estimation of Normal Consumption. *J. Agric. Food Chem.* **2006**, *54* (11), 4069–4075.

Zafra-Stone, S.; Yasmin, T.; Bagchi, M.; Chatterjee, A.; Vinson, J. A.; Bagchi, D. Berry Anthocyanins as Novel Antioxidants in Human Health and Disease Prevention. *Mol. Nutr. Food Res.* **2007**, *51* (6), 675–683.

Chapter 3: Discovery of the Anthocyanin Acyltransferase Involved with the Reduced Acylation Trait in Maize

3.1 Abstract

In a previous study analyzing the diversity of anthocyanin content in diverse maize accessions, a unique phenotype was discovered that had previously never been characterized in maize. Some accessions produced markedly less acylated anthocyanin pigments than normal accessions. In these accessions, cyanidin 3-glucoside was the major pigment, whereas acylated cyanidin 3-(6”-malonyl)glucoside is typically the most abundant pigment. It is hypothesized that this phenotype is due to a partial loss-of-function acyltransferase. To map the location of the mutant in the maize genome, a mapping population was developed from a mutant phenotype line crossed to B73, the reference genome. After HPLC analysis of kernels from 129 F_2 lines, the trait was determined to be due to a single locus. A genotyping-by-sequencing approach was used to discover SNPs to map the locus. The percentage of total anthocyanins modified by acylation was used as the response variable to model the most significant SNP markers. This study concludes that the most significant locus is at the end of chromosome 1 and corresponds to anthocyanin acyltransferase-like protein GRMZM2G387394. Using a UniformMu *Mu* transposon insertion in this gene, it was concluded the gene has anthocyanidin 3-*O*-glucoside 6”-*O*-malonyltransferase activity. Several control traits which correspond to known anthocyanin transcription factors segregating in the population were also confirmed in this study and their effect on total anthocyanin content (TAC) was characterized. It was shown that the candidate anthocyanin acyltransferase in maize has an estimated 34% reduction in TAC on average. This study shows that GBS is a reliable tool to discover new genes even with a small mapping population size.

3.2 Introduction

Anthocyanin are the colorful molecules responsible for most of the red, blue, pink, and purple colors exhibited in plants. The anthocyanins are a diverse class of secondary metabolites with over 635 unique compounds discovered to date (He and Giusti, 2010). The anthocyanin biosynthetic pathway is the most well-studied secondary metabolite pathway in the plant kingdom (Irani et al., 2003). In maize, all the essential genes in the anthocyanin biosynthetic pathway have been cloned and sequenced to date, with *Purple aleurone1* completing the core pathway (Figure 2.3; Sharma et al., 2011). However, the genes involved with increasing the diversity of anthocyanin compounds in maize have not all been discovered. For example, a major anthocyanin in maize, cyanidin 3-glucoside (C3G), can be methylated by an anthocyanin *O*-methyltransferase to produce peonidin 3-glucoside (Pn3G). However, the anthocyanin *O*-methyltransferase in maize has not been identified. Moreover, the most common modification to anthocyanins in maize is acylation. Acylation is the addition of an acyl group to the glycoside of an anthocyanin molecule. In the survey of anthocyanin production in maize, known acylated compounds accounted for an average of 59% of the total anthocyanins produced with a maximum of 84% (Supplementary Table 2.9). Despite the importance of acylated anthocyanins, the genes forming these pigments in maize have not been characterized.

Anthocyanin acyltransferases are the diverse family of genes that synthesize acylated anthocyanins. These genes belong to the BAHD (Benzylalcohol *O*-acetyltransferase, Anthocyanin *O*- hydroxycinnamoyltransferase, anthranilate *N*-Hydroxycinnamoyl/ benzoyltransferase, and Deacetylvindoline 4-*O*-acetyltransferase) superfamily of acyltransferases (D'Auria, 2006). Different types of acyltransferases depend on the acyl group substrate and site of acylation. The two classes of anthocyanin acyl group substrates are aromatic and aliphatic.

Aromatic acylation is the addition of gallic acid or hydroxycinnamoyl groups like sinapoyl, caffeoyl, or *p*-coumaroyl groups to the anthocyanin glycoside. Aliphatic acylation adds carboxylic acid groups like malonic, succinic, or acetic acid (Yonekura-Sakakibara et al., 2008). In maize, aliphatic acylation is most common modification with the predominant compounds being malonylglucosides. Succinylglucosides and caffeoylglucosides have also been reported in various maize accessions (Abdel-Aal et al., 2006; Salinas-Moreno et al., 2012).

Acyltransferase specificity is also determined by the site of acylation. Acyltransferases are specific for the position of glycoside attachment on the anthocyanin molecule and also the site of acylation on the glycoside. The primary site of acylation in maize is at the 6”-position of the 3-glucose. The secondary site of acylation occurs on the 3”-position to form anthocyanin 3-*O*-(3”,6”-dimalonyl)glucosides (Figure 3.1). Only one enzyme with the capability to produce these compounds has been characterized and sequenced: *Dendranthema* × *morifolium* anthocyanidin 3-*O*-3”,6”-*O*-dimalonyltransferase (*Dm3MaT2*) (Suzuki et al., 2004a). This suggests the candidate for an anthocyanin acyltransferase in maize is an anthocyanidin 3-*O*-glucoside-3”,6”-*O*-dimalonyltransferase (3MaT).

Acylated anthocyanins generally show greater stability than their corresponding non-acylated pigments. Aromatic acylation is regarded as the most stable type of acylation. Intramolecular stacking of aromatic acyl groups with anthocyanidins has been theorized to protect the anthocyanins from nucleophilic attack and to shift the color spectrum towards bluer hues (Goto and Kondo, 1991). Aliphatic acylation is still important, however. Aromatic and aliphatic acylation are both known to increase the stability of anthocyanin molecules with increased pH (Bakowska-Barczak, 2005). The color of anthocyanins is most stable in the flavilium state at a pH of 1.0. Increases in pH above 5.0 irreversibly shift the structural

equilibrium to colorless carbinol pseudobase and chalcone states. This effect is important because the subcellular locations of anthocyanins can have a pH high as 5.0 in the vacuoles and 7.0 in the cytosol (Suzuki et al., 2002). Acylated anthocyanins hinder the formation of colorless compounds under more basic conditions better than monoglucoside forms (Saito et al., 1988; Suzuki et al., 2002). Moreover, aliphatic acylation has been shown to increase anthocyanin concentration by enhancing solubility, protecting anthocyanins from enzymatic breakdown, and increasing uptake into the vacuoles (Suzuki et al., 2002; Nakayama et al., 2003; Zhao et al., 2011).

In the survey of anthocyanin production in diverse maize germplasm (See Chapter 2), a new trait was discovered that had previously never been characterized; certain accessions produced markedly less acylated anthocyanins than normal accessions (Figure 3.2). This reduction was on average 51.3% in the whole collection. The hypothesis is that this trait is due to a partial loss of function 3MaT in maize. To test the hypothesis, a mapping population was created by crossing B73, the maize reference genome with normal acyltransferase function, to reduced acylation genetic stock 707G from the Maize Genetics Cooperation Stock Center (MGCSC). 129 F_2 progeny from this cross were generated and phenotyped by high-performance liquid chromatograph (HPLC) to determine pigment content and composition. Next-generation sequencing (NGS) technology was utilized to genotype the progeny and develop molecular markers. NGS has revolutionized the way molecular markers are developed today. Now a single sequencing run can generate hundreds of billions of genotype reads at a very low cost. PCR- or fluorescent-based markers are no longer the standard for genotyping. Instead, single nucleotide polymorphisms (SNPs) are more widely used for correlating phenotypes with genotypic markers (He et al., 2014). In this study, a genotyping-by-sequencing (GBS) approach outlined by Elshire

et al. (2011) was used to discover SNPs in the mapping population. GBS is a form of reduced-representation sequencing that reduces the complexity of an organism's genome by only sequencing regions adjacent to restriction enzyme cut sites. GBS is advantageous for organisms with large genomes with highly repetitive DNA regions, like maize. Barcode nucleotide sequences and NGS primers are ligated to these restriction enzyme sites so numerous samples can be combined and differentiated in one sequencing run, which reduces costs (Elshire et al., 2011). Presented here is a demonstration of how GBS can be effectively used to discover candidate genes for highly heritable quantitative traits, using the reduced acylation trait as an example to discover a novel gene in maize involved with 3MaT activity.

3.3 Materials and Methods

3.3.1 Plant Materials

Many of the plant materials used in this study were from the original anthocyanin diversity panel from Chapter 2. The Maize Genetics Cooperation Stock Center (MGCSC) in Urbana, IL, USA also provided reduced acylation mutants M142X, M741I, and UniformMu stock UFMu-09775. The UniformMu stock was generated in a *Mu* transposon mutagenesis study (McCarty et al., 2005). The genetic stock 707G was chosen as a reduced acylation parent for the mapping population because it contained many phenotypes that contrasted with B73 (Figure 3.3). 707G was a genetic stock that originated from E. Coe's original *in1* mutant stocks. Selection for the mutant was indirect in this stock since the mutant has no visual phenotype. 707G was selfed for three generations before crossing to B73, meaning it is not fully inbred (M. Sachs, personal communication, November 9, 2016). This accession is known to contain recessive alleles *intensifier1* (*in1*), *Pericarp color1 (P1-ww),* and *glossy1* (*gl1*). *in1* is known to enhance total anthocyanin content in the aleurone when homozygous recessive (Burr et al., 1996), *P1-ww* lines

have white cobs and colorless pericarps if anthocyanins are not present (Grotewold et al., 1994), and *gl1* is a cuticular membrane mutation (Sturaro et al., 2005). B73 is also recessive for two well-known transcription factors that are required for aleurone pigmentation: *Colorless1* (*C1*) and *R1* (Andorf et al., 2010). Since these genes have all been characterized, they were used as control loci to verify the SNP dataset. The initial cross was made at the University of Illinois Vegetable Research Farm (40˚ 04′ 38.89″ N, 88˚ 14′ 26.18″ W) in the summer of 2014. F_1 kernels were taken to Buin, Maipo Province, Chile in the winter of 2015 where five plants were selfed to generate F_2 kernels. Only blue kernels were selected from the F_2 kernels to continue the mapping population. In the summer of 2015, 129 F_2 plants were selfed at the Vegetable Research Farm to generate $F_{2:3}$ kernels. F_2 plants were grown with 30" spacing in 25' rows. $F_{2:3}$ kernels from the 129 F_2 plants were phenotyped and genotyped as described below.

3.3.2 HPLC Analysis

Approximately 50 $F_{2:3}$ colored kernels from the F_2 plants were ground with a coffee grinder into a fine powder. Samples containing less than 50 kernels had approximately half of their colored kernels sampled. A 2.0 gram subsample of corn powder was added to 10 mL ACS reagent grade 2% (v/v) formic acid within a 15 mL centrifuge tube. The mixture was extracted overnight in the dark on an Innova 4000 incubator shaker (New Brunswick Scientific, Edison, NJ, USA) set at room temperature and at 200 RPM. Extracts were filtered through a 25 mm 0.45 µm Millex Millipore (EMD Millipore, Merck KGaA, Darmstadt, Germany) LCR PTFE Syringe Filter after centrifugation at approximately 7500 x *g* on a Beckman J2-21M floor centrifuge with a JA-20.1 rotor (Beckman Coulter, Inc, Brea, CA, USA). A 20 µL aliquot of extract was injected using a Hitachi L-7200 HPLC (Hitachi High Technologies, Inc. Schaumburg, IL, USA) equipped with a Grace Prevail (W. R. Grace & Co., Columbia, MD, USA) C_{18} 5 µm analytical column

(250mm x 4.6mm) and a Hitachi L-7455 Diode Array Detector quantifying absorbances at 520 nm. The column was heated to a constant 30.0 ˚C. The mobile phase consisted of 2% (v/v) formic acid and 100% HPLC grade acetonitrile (ACN) at a flow rate of 1 mL/min in the following linear gradient: 15% ACN at 0 minutes, 30% ACN at 10 minutes, and 15% ACN at 15 minutes. After each sample, the column was allowed to equilibrate with 15% ACN for 10 minutes. Samples were all run twice and averaged.

3.3.3 Identification of Anthocyanins

All known anthocyanins were previously identified in the survey that initially identified the mutant (See Chapter 2.2 Materials and Methods). Retention times are different in this study because a new mobile phase was used that reduced run time from 35 min to 15 min per sample (Figure 3.2). The order of elution for each absorbance peak was still indicative of the various compounds in the sample. One limitation of this method was that several unknown compounds would co-elute with Pn3G. One of these compounds may be an isomer of cyanidin 3-(6”-malonyl)glucoside (C3MG) based on mass spectrometry in other maize anthocyanin studies (Jing et al., 2007). All peaks found under Pn3G would be summed together to simplify quantification. In addition, an unidentified peak labeled ID #6 would appear next to pelargonidin 3-(6”-malonyl)glucoside (Figure 3.2a). This peak was considered as a separate acylated anthocyanin compound because it eluted with other known acylated anthocyanins. This peak may also be another isomer of unknown acylated anthocyanins.

3.3.4 Phenotypes

Total anthocyanin content (TAC) was quantified as shown in Equation 3.1.

Equation 3.1: $$TAC = \frac{Total\ Peak\ Area}{MM\ Peak\ Area} * MM$$

Total peak area in Equation 3.1 was calculated by summing the integrated absorbance peaks for each compound calculated in the Hitachi HPLC System Manager 4.0 software. *MM peak area* was the average total peak area of two or more replicates of the external standard, Angelina's Gourmet Maize Morado (MM; Swanson, CT, USA), analyzed throughout a batch of samples. The TAC of MM, *MM* in Equation 3.1, was assumed to be a constant 1000 mg/kg as described previously (See Chapter 2.2). All other anthocyanin compound phenotypes used integrated peak area for concentration and not MM relative. The phenotype chosen to describe reduced acylation mutants for QTL analysis was percentage of acylated anthocyanins of TAC. Acylation percentage was calculated by summing the integrated peak areas of all acylated compounds in the sample, then dividing that by the total peak area of the chromatogram, and multiplying by 100. A histogram of acylation percentage is included in Figure 3.4b. There were two distributions apparent in the histogram that represented mutants and wild types. Mutants were generally under 30% acylation while wild types were above 55% acylation. This rule was applied so reduced acylation could be coded as a binary phenotype. Three intermediate samples could not be defined unequivocally as mutants or wild-types with these parameters and were excluded from ANOVAs, but not QTL analyses. Many samples expressed speckled aleurone pigmentation due to the incomplete expression of *R1* in aleurone receiving one *R1* allele from a pollen parent in a female *r1r1* endosperm (Figure 3.3d; East and Hayes, 1911). The presence/absence of speckling was used as a covariate to measure the effect of *R1* on TAC. The presence of homozygous recessive *C1* in a kernel also produces colorless kernels, so the presence of the recessive allele was made into a covariate (Figure 3.3c). To define a sample as having recessive alleles of *C1*, the proportion of colorless kernels per ear was visually estimated as either 50%, 75%, or full color. If a sample had approximately 50% colored kernels, it was

defined as having recessive alleles of *C1*. If the sample contained 75% uncolored kernels without speckling, it was also defined as having recessive alleles of *C1*.

3.3.5 DNA Extraction

Tissue from six to eight germinating $F_{2:3}$ seedlings was collected into 1.2 mL microtiter tubes from the developing leaf whorl one week after planting. $F_{2:3}$ kernels were planted in Fafard Germination Mix (Sun Gro Horticulture, Agawam, MA, USA) and kept from direct sunlight to induce etiolation and reduce chlorophyll production. The six to eight kernels used to represent F_2 plants were chosen to emulate the sample from which they came in terms of pigmentation. Anthocyanin-pigmented to anthocyanin-less kernels were mixed in the ratio visually estimated for that sample. In the cases where *C1* and *R1* together were segregating recessive alleles, 50% would be expected to be yellow, so three to four yellow kernels were sampled for genotyping. In the case where only one of the transcription factors was segregating, one to two yellow kernels were sampled. Sampling blue and yellow kernels was necessary to ensure *C1* and *R1* alleles were present so the *C1* and *R1* loci could be mapped accurately. In a few cases, representative samples could not be collected from germinating kernels due to poor germination rates. For these, tissue samples were collected from the flag leaf of the tassel on the F_2 plant. All tissue was lyophilized before extraction and ground using a 2010 Geno/Grinder (SPEX SamplePrep, Metuchen, NJ, USA) using 17/32" chrome steel beads. The extraction protocol was a CTAB protocol optimized for 96-well plate extractions (Doyle and Doyle, 1990). DNA quantification was completed using the PicoGreen assay (Molecular Probes, Eugene, OR).

3.3.6 GBS Library Construction

The GBS protocol outlined by Elshire et al. (2011) was used for library construction with a double restriction enzyme modification from Poland et al. (2012). Restriction enzymes chosen

for library construction were HinP1I and PstI. In addition to the 129 samples from the reduced acylation mapping population, 201 other samples from an independent experiment were multiplexed during sequencing for 330 total samples. Libraries were read in one sequencing lane on an Illumina HiSeq2500 (San Diego, CA, USA) at the Roy J. Carver Biotechnology Center at the University of Illinois in Urbana, IL, USA.

3.3.7 SNP Discovery Pipeline

SNPs in the population were called using the TASSEL 5.0 GBS v2 pipeline (Bradbury et al., 2007). Bowtie 2 was used to align fragments to *Zea mays* B73 reference genome version 3 (Youens-Clark et al., 2011; Langmead and Salzberg, 2012). Minor allele frequency (MAF) was set to 0.05. This resulted in 4,660,000 total SNPs for all 330 samples. Missing data ranged from 13.2% to 99.7% per sample. Samples with greater than 80% missing data were excluded. This changed the dataset to 328 samples, with one sample lost from the reduced acylation population. Next, insertion-deletions were removed in TASSEL 5.0 to be imported into Beagle 4.1. The options defined in Beagle 4.1 were a window length of 200 and a step size of 50 (Browning and Browning, 2016). After imputation, the reduced acylation population was separated from the independent mapping population within TASSEL 5.0. A MAF of 0.05 was applied once again to the imputed SNPs resulting in a total of 1,031,936 SNPs or 8062 SNPs per sample for the reduced acylation population. Markers per chromosome ranged from 584 for chromosome 10 to 1212 for chromosome 1. Since 707G was a genetic stock and not inbred, multiple alleles were present in a few loci. Third alleles were present at 308 sites and fourth alleles for 116 sites, which is not typical of biparental populations. Third alleles can be explained by a heterozygous parent, but fourth alleles are most likely due to sequencing error or misalignment to the reference genome. Alleles were converted in TASSEL 5.0 to numerical genotypes 1 versus 0 for major

versus minor or 0.5 for heterozygotes based on allele frequency in the population. This means third and fourth alleles were converted to 0s if homozygous or 0.5 for heterozygous. Sites with third and fourth alleles will not affect the reduced acylation mutation or other control loci since control loci were homozygous for several generations in 707G and linkage blocks surrounding these loci should be large after three generations of selfing. Numerical genotypes were multiplied by two to remove decimals and imported into R to perform QTL analyses.

3.3.8 Statistical Analysis

ANOVAs were calculated using either Proc Reg or Proc Mixed in SAS 9.4 (SAS institute Inc., Cary, NC, USA) depending on the outputs needed. *C1, P1-wr, R1,* and reduced acylation mutants were coded as 0 or 1 depending on if they the sample contained dominant or homozygous recessive alleles, respectively, based on the phenotypes previously described. *In1* genotypes were inferred from GBS marker data. Three samples were deleted from analyses since they were intermediate in acylation percentage and could not be determined as mutants (See above). The significance of the *C1, In1, P1-wr, R1*, and reduced acylation mutant genotypes in explaining TAC was calculated using an ANOVA in Proc Mixed with method equal to Type 3 and the model in Equation 3.2. Parameter estimates for significant genes from Equation 3.2 were calculated in Proc Reg with the model in Equation 3.3. Variance components were estimated in Proc Mixed with a Type 3 ANOVA using the same model in Equation 3.3 and with gene effects set to random.

Equation 3.2: $$\boldsymbol{TAC = \beta_0 + \beta_1 In1_i + \beta_2 R1_i + \beta_3 C1_i + \beta_4 P1_i + \beta_5 mutant_i + \varepsilon_i}$$

Equation 3.3: $$\boldsymbol{Phenotype = \beta_0 + \beta_1 In1_i + \beta_2 R1_i + \beta_3 C1_i + \beta_4 mutant_i + \varepsilon_i}$$

All QTL analyses were run using the linear model (lm) function in R (R Core Team, 2015). The three samples deleted for the ANOVAs were included for QTL analyses since they will help

strengthen the QTL and since acylation percentage is unrelated to the control loci. A stepwise QTL analysis was used to find significant markers involved with the phenotype of interest. The first model run would assume a single QTL. The most significant marker from a single-QTL model would then be used as a covariate (**Q**) and the process repeated until no more markers were significant. The general model for stepwise QTL analysis is shown in Equation 3.4. $\sum_{k=2}^{p} \boldsymbol{\beta_k Q_{ki}}$ represents the summation of the *k*th significant marker as a covariate for *p* covariates, while $\boldsymbol{\beta_1 SNP_i}$ represents the numerical genotypes at a SNP site.

Equation 3.4: $$\boldsymbol{Phenotype = \beta_0 + \beta_1 SNP_i + \sum_{k=2}^{p} \beta_k Q_{ki} + \varepsilon_i}$$

The threshold for significant marker *p*-values in Equation 3.4 was determined by a Bonferroni correction of $\alpha = 0.05/n$ where *n* is the number of SNPs (*n*=8062) and 0.05 is 95% confidence. All plots were also generated in R (R Core Team, 2015).

3.4 Results

3.4.1 Inheritance of the Trait

Ten accessions have been found to date that exhibit the reduced acylation trait. Seven were from the survey of anthocyanin production (Chapter 2): Ames 14276, PI 483476, PI 553060, Jerry Peterson Blue Dent (White Buffalo Trading Company, San Jose, CA, USA), Black Aztec Sweet Corn (Seed Savers Exchange, Decorah, IA, USA), MGCSC 707G, and MGCSC X19EA. These accessions are from diverse backgrounds and are seemingly unrelated. The other three were tested for the reduced acylation mutation in the summer of 2016: PI 213729, MGCSC M142X, and MGCSC M741I. Pedigree analysis determined that all the MGCSC stocks share a similar, albeit distant, relationship to E. Coe's *in1* genetic stocks (E. Coe and P. Stinard, personal communication, May 5, 2016). PI 553060's lineage traces back to the Ornamental Corn

Improvement Project (Widrlechner and Dragula, 1992). The reduced acylation trait has no visual phenotype, so selection for this trait is apparently indirect.

To test whether reduced acylation was due to environment, 707G, Ames 14276, and Jerry Peterson Blue Dent were selfed for at least four seasons and analyzed on the HPLC. The phenotype was exhibited in these lines every season. Reciprocal crosses made with reduced acylation lines returned normal acylation abundance meaning the trait is recessive to normal acylation. Several reduced acylation lines were crossed with each other to determine if the trait was due to the same locus in every accession. It was concluded that the phenotype in Ames 14276, Jerry Peterson Blue Dent, and MGCSC stocks 707G, M142X, M741I, and X19EA are alleles at the same locus. Future crosses will be made to confirm allelism for the other four accessions.

3.4.2 Phenotypes

In the mapping population, 129 F_2 individuals generated kernels for phenotyping and sequencing. Twenty individuals visually phenotyped were homozygous *C1C1C1*/*R1R1R1* and exhibited full anthocyanin coloration. Thirty-seven visually phenotyped were *C1C1C1*/*R1r1_*, thirty-four *C1c1_*/*R1R1R1*, and thirty-eight *C1c1_*/*R1r1_* (Table 3.1). Genotypes were determined by the visual markers mentioned in 3.2 Materials and Methods (Figure 3.3). The *P1-wr* gene that confers red cobs was present in 95 samples. The histogram of acylation percentage was used to classify mutants based on mutant and wild type distributions (Figure 3.4b). Three samples could not unequivocally be called mutants without sequencing resulting in 29 of 126 reduced acylation mutants. Using a χ^2 test, the null hypothesis for a single locus could not be rejected with high confidence (p=0.607).

The overall population mean for TAC was 96.2 mg/kg (Table 3.1). The distribution of TAC was skewed right indicating that *in1* had a profound affect on TAC in this population (Figure 3.4a). Since a new HPLC method was used in this study, the coefficient of variation (CV) was calculated to determine repeatability. CV was calculated by dividing the standard deviation of TAC for the replicates by the sample mean. CV ranged from 0.02 to 30% with an overall average of 3.09% confirming the method is acceptably reproducible. To see which genes had an effect on TAC, a linear regression was modeled using Equation 3.2. All variables except *P1-wr* were significant at $p<0.05$ (Table 3.2a). *in1* raises TAC an estimated 103.7 mg/kg while the mutant, *r1,* and *c1* lower TAC 23.6 mg/kg, 20.8 mg/kg, and 27.2 mg/kg, respectively. The lowest TAC sample was 11.84 mg/kg in a *C1c1/R1r1/In1_* mutant plant and the highest was 369.54 mg/kg in a *C1C1/R1R1/in1in1* wild-type plant. The TAC reported here is 291% greater than the average TAC of parent 707G, which contained an average 126.9 mg/kg in 2015.

3.4.3 Identifying Candidate Genes

Fortunately, several anthocyanin acyltransferases have been characterized to date in other plant species, so candidate anthocyanin acyltransferase sequences could be discovered using the B73 reference genome (Unno et al., 2007). Since the most likely candidate for an acyltransferase in maize should have 3MaT activity, the dimalonyltransferase in chrysanthemum, *Dm3MaT2,* was used to find candidate sequences in BlastP (Suzuki et al., 2004a). The BlastP search found seventeen candidates in maize with 25% or greater identity and 15% or greater query coverage. The seventeen candidates were aligned in COBALT using defaults (Papadopoulos and Agarwala, 2007). The candidates all showed close homology to the three motifs that are present in all anthocyanin acyltransferases: HXXXD (motif 1), DFGWG (motif 2), and NYFGNC (motif 3) (Suzuki et al., 2002). Motif 3 is an anthocyanin acyltransferase specific motif. Protein sequences

were filtered on whether they contained motif 3 present in chrysanthemum: XYF/LGNC (Suzuki et al., 2004a). This filter resulted in nine candidate sequences (Table 3.3). Among these candidates, four were annotated as predicted anthocyanin acyltransferases. One was even annotated as an anthocyanidin 3-*O*-glucoside 6"-*O*-malonyltransferase, which is a strong candidate for a maize anthocyanin acyltransferase. The remaining five candidates were uncharacterized proteins or transferase family proteins.

3.4.4 QTL Analyses

Several control traits were mapped to assess the accuracy and robustness of the SNP dataset before the reduced acylation trait was mapped. The control loci, *C1*, *R1*, *P1-wr*, and *In1,* segregating in the population have been previously mapped, sequenced, and characterized (Paz-Ares et al., 1987; Dellaporta et al., 1988; Grotewold et al., 1994; Burr et al., 1996). The log-transformation of TAC was used as the phenotype to map *In1*, since the distribution was skewed (Figure 3.4a). Figure 3.5 shows the Manhattan plots generated from the four control traits. The most significant SNP correlations always corresponded to the known location of each gene. The most significant marker for *In1,* with a –log10 *p*-value of 21.36, was only 134 bp away from the actual gene. The known location of the other control loci was within a reasonable distance from the most significant SNP marker. *P1-wr*, *C1,* and *R1* were 642 Kb, 1.7 Mb, and 637 Kb away, respectively (Andorf et al., 2010).

The phenotype chosen to discover markers associated with the reduced acylation trait was percentage of acylation, since proportions are not affected by TAC. The most significant marker with a –log10 *p*-value of 18.7 was at position 298,042,277 of chromosome 1 (Figure 3.6). The distribution of percentage of acylation per allele at the most significant marker is shown in Figure 3.7. One major problem with GBS is the inability to call heterozygotes with low coverage

sequencing (Swarts et al., 2014). Some samples were called homozygotes because the SNP site did not have a high enough read depth to find both major and minor SNP states. The Bonferroni correction established a 22.3 Mb interval of significant markers from positions 279,188,074 to 301,476,924 (the end of chromosome 1). The genomic region search function in Phytomine (https://phytozome.jgi.doe.gov/phytomine/genomicRegionSearch.do, accessed November 9, 2016) has 1029 gene models in this interval on chromosome 1 (Youens-Clark et al., 2011; Goodstein et al., 2012). The QTL for the reduced acylation trait spanned only one of the candidate genes from Table 3.3. The candidate gene designated GRMZM2G387394 in B73 reference genome version 3 is at position 300,173,138 of chromosome 1 (Figure 3.6b).

TAC is a trait of interest to breeders looking to make high-anthocyanin yielding varieties. To find QTL associated with TAC other than the ones known, a model was fit with *In1* genotypes as a covariate and both log-transformed TAC and raw TAC as responses. No significant QTL could be distinguished from both models at the significance threshold established by the Bonferroni correction, so the *p*-value was relaxed to 1×10^{-4} in accordance with Li et al. (2016). With untransformed TAC, a QTL on chromosome 7 was most significant at position 144,238,875 (Figure 3.8b). Near this SNP is *MYB152,* a transcription factor that is demonstrated to correlate with *PAL* expression (Agarwal et al., 2016). When both *In1* and the second QTL on chromosome 7 were considered, a QTL at position 137,169,800 of chromosome 9 became significant at $p<1\times10^{-4}$ (Figure 3.8c). No obvious gene candidates for this QTL could be found at this time.

In addition to TAC, QTL analyses were run on individual compounds distinguished by HPLC. All compounds besides Pn3G and Pg3DMG had two QTL corresponding to the acyltransferase candidate and the *In1* gene (Figure 3.9). Pn3G and Pg3DMG had one significant

QTL at the site of *In1*. When Pn3G was modeled with *In1* as a covariate, a QTL spanning *R1* became significant. The same was not true for Pg3DMG.

Variance components for each known TAC-altering gene were estimated for all traits to determine the phenotypic variation attributed to each gene (Table 3.4). *In1* explained most of the variance (66.5%) for TAC in this population, while *R1*, *C1*, and the mutation accounted for 2.29%, 4.42%, and 2.94% of the total variance, respectively. In general, *In1* explained the highest proportion of variance for all compounds and ranged from 28.5% to 58.6%. The lowest proportion of variance that could be detected as a significant QTL was 5.27% for *R1* with Pn3G. Control loci could be detected when they reached a significance of $p<0.001$ in the ANOVA.

3.4.5 QTL Simulation

The dataset was sampled and simulated to determine the lowest number of samples possible to confirm a QTL. The two traits assessed were the presence of speckled kernels (*R1* phenotype) and percentage of acylation. 74 of the 128 genotyped samples contained speckled kernels while 29 of the 128 genotyped samples were determined to be reduced acylation mutants. Subsamples of 20, 40, 60, 80, and 100 genotypes were selected at random from the dataset. Sampling was repeated 100 times for each subsample of genotypes. The ability to detect QTL was arbitrarily defined as the most significant SNP above the Bonferoni significance threshold ($p<6.203\times10^{-6}$) being only 10 Mb away from the known locus. The summary of the simulation is presented in Table 3.5. For speckled kernels, forty samples was sufficient to accurately detect the *R1* locus 100% of the time within 10 Mb. The average distance from the *R1* locus for the 40 samples was 2.15 Mb away. 23% of the 20-genotype simulations significantly pinpointed the locus within 10 Mb. In the 20-genotype subset 44% accurately pinpointed the locus within 10 Mb, while 25% were considered false positives. The average *p*-value for the significant markers

ranged from a –log10 *p*-value of 10.99 to 61. The high *p*-values associated with the 20-genotype subsamples may due to ascertainment bias. For percentage of acylation, the QTL could be detected 99% of the time with sixty samples. The average distance from the candidate acyltransferase for sixty samples was 2.40 Mb. In 31% of the simulations, the reduced acylation QTL could be mapped to the correct location with only twenty genotypes. One simulation was able to establish significance with only three of twenty reduced acylation mutants. The average significance for percentage of a acylation ranged from 6.46 to 15.22 –log10 *p*-value. The reduced acylation simulation had more false positives then the *R1* locus. For the 40-genotype and 20-genotype subset, 6% and 27% of the significant markers were considered false positives, respectively.

3.4.6 UniformMu Insertion

To determine if the candidate acyltransferase on chromosome 1 was associated with the formation of acylated anthocyanins, a UniformMu stock with a *Mu* transposon insertion within GRMZM2G387394 was crossed to reduced acylation mutant 707G (McCarty et al., 2005). All of the *Mu* insertion lines obtained were heterozygous for the insertion and would theoretically confer the mutant phenotype in 50% of the crossed kernels. Single kernels (~0.2 g) were ground, extracted overnight in 5 mL 2% (v/v) formic acid, and analyzed on the HPLC to test for the presence of the reduced acylation phenotype. The UniformMu line was able to maintain the reduced acylation phenotype in 14 of the 25 kernels analyzed (χ^2 test, H_0=50%, p=0.5485), meaning that GRMZM2G387394 is required for normal anthocyanin acyltransferase function in maize and directly involved with the reduced acylation trait (Figure 3.10).

3.4.7 Anthocyanin Acyltransferase Candidate to *Dm3MaT2* alignment

The sequence of the candidate anthocyanin acyltransferase was aligned to *Dm3MaT2* (Figure 3.11). The sequences only shared 28% identity between each other, which is not uncommon among BAHD family members (St-Pierre and De Luca, 2000). There were several amino acid residues in *Dm3MaT2* that Unno et al. (2007) found to be important for 3MaT activity: Thr-35, Leu-37, Ala-38, Gln-51, Tyr-411, Ala-413, and Lys-419. The candidate gene has a deletion up to position 79 on *Dm3MaT2*, indicating that amino acids Thr-35, Leu-37, Ala-38, and Gln-51 are not required for anthocyanin acyltransferase activity in maize. In the same study, the crystal protein structure of the related homolog *Dm3MaT3* demonstrated that these four amino acids directly interacted with the acyl-acceptor binding site. It is unclear how the maize anthocyanin acyltransferase is able to bind to the anthocyanin molecule if this domain is missing. The three remaining amino acids had no identical counterparts in maize either. Tyr-411 was replaced with Val-429 in maize and Ala-413 with Met-338. Those two substitutions have similar hydrophobic properties and may be similar enough for anthocyanin acyltransferase activity. Finally, Lys-419 was replaced with Thr-344 in maize. Without performing enzyme feeding assays, it is difficult to predict whether the candidate acyltransferase in this study has 3MaT activity. Due to the UniformMu knockout, it is possible to conclude that the candidate gene at least has anthocyanidin 3-*O*-glucoside 6”-*O*-malonyltransferase activity.

3.4.8 Phylogenetic Analysis

At least nineteen flavonoid acyltransferases have been characterized to date (Unno et al., 2007; Kim et al., 2009). All characterized flavonoid acyltransferases are from eudicot species with the exception of one flavonoid malonyltransferase in *Oryza sativa* L. (Kim et al., 2009). The neighbor-joining phylogenetic tree of sixteen known flavonoid acyltransferases with the

maize candidate acyltransferase is shown in Figure 3.12. Monocots formed their own clade, as hypothesized, and the phylogenetic tree otherwise matched previously reported phylogenies (Unno et al., 2007). The monocot clade is more closely related to the other 3MaTs than the 5-*O*-glucoside or aromatic anthocyanin acyltransferases. The monocot phylogeny also resembles malonyl-CoA:anthocyanin 5-glucoside 4'''-*O*-malonyltransferase from *Salvia splendens* more than any other anthocyanin acyltransferase. This unique 4'''-*O*-malonyltransferase is an unusually divergent anthocyanin acyltransferase and is more closely related to acetyl-CoA:benzylalcohol acetyltransferase than anthocyanin acyltransferases (Suzuki et al., 2004b). In the alignment among all acyltransferase sequences, it was clear that the candidate anthocyanin acyltransferase in maize was unique in regards to the deletion in the first 70 to 93 amino acids (depending on the protein being aligned). Rice and maize show altered amino acid preference for motif 2, which is otherwise highly conserved in the eudicot flavonoid acyltransferases. Motif 2 in maize and rice is DFGFG, while in eudicots, this motif was always DFGWG for flavonoid acyltransferases. Motif 2 is not strictly conserved among all BAHD superfamily members however, as black cottonwood and *Arabidopsis* also contain BAHD members with a DFGFG motif (Yu et al., 2009). More anthocyanin acyltransferase members in monocots would need to be characterized to confirm phylogenetic amino acid specificity of motif 2 and other sites.

3.5 Discussion

3.5.1 Phenotypes

In the mapping population, TAC was significantly lower (estimated –23.6 mg/kg) when the reduced acylation trait was present (Table 3.2b). This finding agrees with the hypothesis that aliphatic acylation has a function in protecting anthocyanins in the cytosol and increasing uptake into the vacuoles (Suzuki et al., 2002; Zhao et al., 2011). Therefore, increasing the proportion of

acylated anthocyanins could have a positive influence on increasing TAC. In addition to the reduced acylation mutant, *In1*, *C1,* and *R1* all have a significant effect on TAC (Table 3.2a). The quantifiable change in TAC associated with alleles of each of these anthocyanin regulators has to the best of our knowledge never been studied. Here only heterozygous or homozygous dominant genotypes were sampled for *C1* and *R1*, which will bias the results towards higher TAC. Despite this, significant differences could be found among the combinations of each gene regulator. *In1* had the most profound impact on TAC with an estimated increase of 103.7 mg/kg for homozygous recessive *in1* genotypes (Table 3.2b). The effect of *In1* may be inexact, given the genotypes of *In1* were inferred from low-coverage SNP sites where heterozygotes may have been miscalled (Swarts et al., 2014). The high significance (–log10 *p*-value of 21.36) and relative proximity (134 bp away) of the most significant marker for log-transformed TAC and the approximate 1:3 Mendelian ratio of homozygous recessive *in1* genotypes to dominant *In1* genotypes provide evidence that inferred *In1* genotypes may have been accurately called. Outside of *In1*, *R1* had the second highest effect of the control genes. This may partially be due to *R1* heterozygotes producing speckled kernels, while *C1* heterozygotes produced full color that could not easily be differentiated visually from homozygotes (Figure 3.3).

3.5.2 GBS Reliability

GBS is a simple and robust method for reduced representation sequencing. The ability to multiplex 384 samples per sequencing run greatly reduces costs and allows for all reactions to be completed in 96-well plate format from DNA extraction to library construction. In this study, 330 samples from two separate mapping populations were analyzed in one lane of an Illumina HiSeq2500 (San Diego, CA, USA). The 129 samples from this mapping population resulted in 128 genotypes with 8062 SNPs per genotype. SNPs were more prevalent in areas where

recombination occurred more frequently, but generally, the entire genome was represented. The average distance between SNPs in the final dataset was 255 Kb. 95% of the SNPs were under 1 Mb apart and 17.8% of those were under 1 Kb apart.

A problem with many GBS datasets is the amount of missing data due to low-coverage sequencing. In the initial dataset with 330 samples, two samples were discarded that had missing data of over 80%. These samples most likely failed in the PCR or ligation steps of GBS library preparation. Once these were removed, the proportion of missing sites in the GBS library was 36.32%. The amount of missing data has implications on how imputation fills in SNPs. Imputation has been shown to increase power to detect QTL despite the uncertainty in variant calling for low-depth sequencing (Fragoso et al., 2016). Imputation is not the only issue with missing data. Shallow read depth of SNP sites can result in heterozygote markers being called incorrectly. Therefore, phenotypes may not correlate with SNPs even though flanking markers are highly significant. This issue is apparent in Manhattan plots. In the largest control QTL, which was for log-transformed TAC, 155 SNPs were significant with the Bonferroni significance threshold out of an area that spanned 269 SNPs. *P1-wr* was worse with 91 of 222 significant SNPs. In contrast, non-significant markers could become significant if the right combination of markers were present. False positives should be controlled by the conservative Bonferroni adjustment, but even in *P1-wr*, a single marker on a different chromosome was significant (Figure 3.5). Overall, the effect of heterozygote miscalling is seemingly negligible for this study since the most significant QTL for each control loci aligned correctly and only a single QTL was found for the reduced acylation trait.

The number of SNPs per sample in this study was much less than potential SNPs presented in other studies utilizing GBS or the Illumina MaizeSNP50 Beadchip (Glaubitz et al.,

2014). Separating the mapping populations or running the samples on more than one sequencing lane would have increased the potential number of SNPs and reduced heterozygote miscalling, but would have also increased costs. For the purposes of this study, a higher SNP density was not necessary since there were known candidate genes and the trait was due to a single locus. QTL simulations showed that QTL could be confirmed with as little as forty to sixty samples depending on the proportion of the recessive allele (Table 3.5). The power to detect QTL is best when the dominant/recessive traits are in nearly even proportions. The high power of detection presented here is in direct contrast to a previous study that could only establish significance for a single-locus trait at a chromosome-wide scale with 91 samples (Heffelfinger et al., 2014).

3.5.3 QTL Analyses

Using a Bonferroni threshold resulted in numerous significant markers for some traits. Distances between significant markers sometimes spanned a few hundred or thousands of genes, which may encompass several candidates. In this study, the QTL for reduced acylation only spanned a single candidate, so the width was acceptable. In the future, a new mapping approach based on linkage between markers, like bin-mapping, may be necessary to establish LOD thresholds to differentiate candidates within a single QTL (Chen et al., 2014). Recombination rate in this population is partially to blame for the large number of significant markers. With more recombination events, smaller areas in the genome will remain correlated to the trait due to the breakage of linkage blocks. More recombination events can be introduced with a greater population size. For example, in a study that analyzed F_2 progeny for grain yield traits, the *R1* gene was also mapped as a control, but used 611 progeny. *R1* was fine-mapped to a ~700 Kb confidence interval, as opposed to the 27.0 Mb interval between significant markers presented here (Chen et al., 2014). In addition to increasing population size, intermating the F_2 genotypes

could have been a way to increase recombination events. This scheme was utilized to develop the intermated B73 x Mo17 (IBM) population in maize that is used extensively for mapping traits (Lee et al., 2002; Eichten et al., 2011). Intermating requires additional generations, which is one reason it was not utilized here.

Despite the low SNP coverage relative to other GBS studies, the mapping population was able to map traits with one to three loci. Three QTL could be distinguished for TAC in this mapping population when the significance threshold was relaxed to 1×10^{-4} (Figure 3.8). Interestingly, the two QTL apart from *In1* were not *C1*, *R1*, or the acyltransferase as would be predicted. The phenotypic variance attributed to *In1* was 66.5% in this population, meaning that very little could be explained by other genetic factors (Table 3.4). Future work aimed at increasing TAC would have to expand the mapping population to confirm the presence of these QTL. Other quantitative traits determined by HPLC also confirmed the reliability of using GBS for detecting candidate genes. *In1* and the acyltransferase were both significant loci in explaining the abundance of most anthocyanin compounds. Genes of large effect can be discerned easily even in small populations with low SNP coverage. As evidenced with Pg3DMG, the power to detect a loci is in part determined by the genotypic variance or heritability of the trait. 59.2% of the total variance in Pg3DMG was due to error not attributed to the four control loci (Table 3.4). Only 26 samples had detectable amounts of Pg3DMG and these 26 had relatively high TAC. The large error variance was probably due to the inability to detect the compound at low concentrations. The highest proportion of variability explained by control loci was 79.5% for C3G. This means that C3G composition is highly heritable and most of the variation in C3G content was not due to environmental or analytical error and most likely due to genetic factors. This is in agreement with heritability for C3G calculated in Chapter 2 (Table 2.8).

More than three QTL could not be found in this mapping population since *In1* and the acyltransferase mutant explained most of the variability in most traits. The minimum proportion of variance that produced a QTL in this study was 5.3% (Table 3.4). More than three QTL have the potential to be significant in a GBS study if the trait is highly heritable and due to many genes with smaller effects. Chen et al. (2014) were able to distinguish seven QTL correlated with tassel branch number in a large population of 692 F_2 plants. The proportion of variance explained by the QTL in this trait ranged from 2.0% to 6.3%, which was the typical proportion of variance of *R1* and *C1* in the reduced acylation population. Moreover, another study utilizing GBS for quantitative traits found 35 QTL for plant height, ear height, and internode number with phenotypic variances ranging from 2.6% to 15.68% with 314 recombinant inbred lines (Zhou et al., 2016). Plant height, ear height, and tassel branch number are highly heritable traits just like anthocyanin composition was found to be in Chapter 2 of this paper (Table 2.8).

3.6 Conclusions

Results of this study demonstrate the effectiveness of utilizing GBS for discovering new genes. GBS is a simple and reliable protocol streamlined for 96-well plate format to maximize efficiency. GBS was able to discover a new gene in maize involved with the production of acylated anthocyanins, which are typically the most abundant forms of anthocyanins in maize. This gene, designated GRMZM2G387394 is a predicted anthocyanin acyltransferase protein. Knockout mutants of this gene returned the reduced acylation phenotype, meaning this acyltransferase is responsible for synthesizing a majority of the acylated anthocyanins in maize. Whether or not this gene has dimalonyltransferase activity is yet to be determined. Future work will be aimed at expressing the gene *in vitro* and *in vivo*. Alignments to other flavonoid acyltransferases revealed this gene is a unique member that requires more investigation. Results

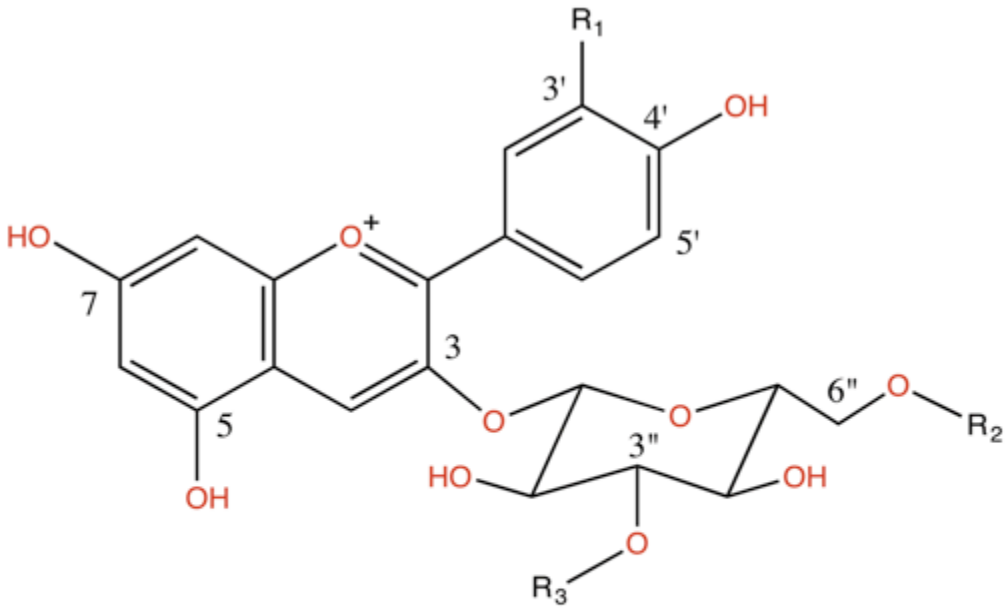
of this study expand the current knowledge of anthocyanin acyltransferases. This study is the first report of an anthocyanidin malonyltransferase in a monocot species. As confirmed in this study, anthocyanin acyltransferases are important for increasing TAC. Altering the proportion of acylated anthocyanins in maize has the potential increase the stability of anthocyanins in food and beverage systems.

3.7 Acknowledgements

I would like to thank Dr. Patrick Brown from the University of Illinois Urbana-Champaign for letting me use his laboratory and equipment and also for his guidance throughout the construction of my GBS library. I would also like to thank Brown Lab members Shilpa Manjunatha and Pradeep Hirannaiah for walking me through DNA extractions and GBS library construction. Thank you again to Pradeep and Dr. Brown for providing the code I used for the TASSEL 5.0 GBS v2 pipeline, Bowtie 2, and Beagle 4.1. Finally, I would like to thank Laura Chatham for the photographs of the kernels I used in Figure 3.3c and 3.3d.

3.8 Figures and Tables

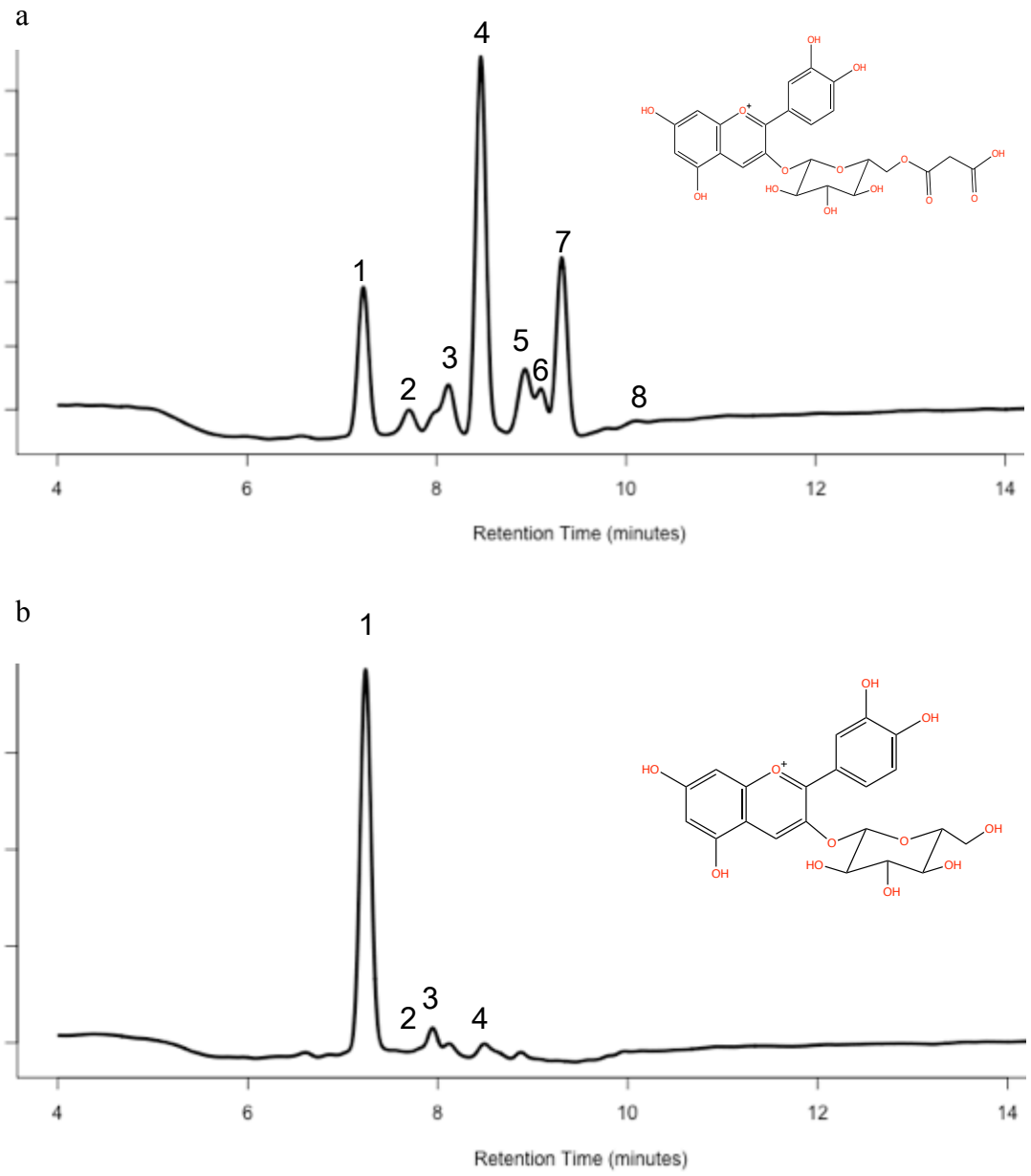
Figure 3.1: Structure of an anthocyanin molecule in maize.



ID #	Identified Compounds	R_1	R_2	R_3
2	Pelargonidin 3-Glucoside (Pg3G)	H	H	H
1	Cyanidin 3-Glucoside (C3G)	OH	H	H
3	Peonidin 3-Glucoside (Pn3G)	CH_3	H	H
5	Pelargonidin 3-(6”-malonyl)glucoside (Pg3MG)	H	malonyl	H
4	Cyanidin 3-(6”-malonyl)glucoside (C3MG)	OH	malonyl	H
7*	Peonidin 3-(6”-malonyl)glucoside (Pn3MG)	CH_3	malonyl	H
8	Pelargonidin 3-(3”,6”-dimalonyl)glucoside (Pg3DMG)	H	malonyl	malonyl
7*	Cyanidin 3-(3”,6”-dimalonyl)glucoside (C3DMG)	OH	malonyl	malonyl
N.D.	Peonidin 3-(3”,6”-dimalonyl)glucoside (Pn3DMG)	CH_3	malonyl	malonyl

N.D. = Not detected
* Pigments co-elute and cannot be separated.

Figure 3.2: Comparison of wild type (a) and mutant profiles (b) with major pigments inset.



Note: Peak labels correspond to the "ID #" listed in Figure 3.1.

Figure 3.3: Control phenotypes in the mapping population: A) *P1-ww* (top) and *P1-wr* (bottom); B) *In1* (left) and *in1* (right) in a *pr1* red kernel background (Photo courtesy of maizegdb.org); C) *C1*/*c1* segregating ear with yellow and blue kernels; and D) *R1*/*r1* segregating ear demonstrating the speckled phenotype.

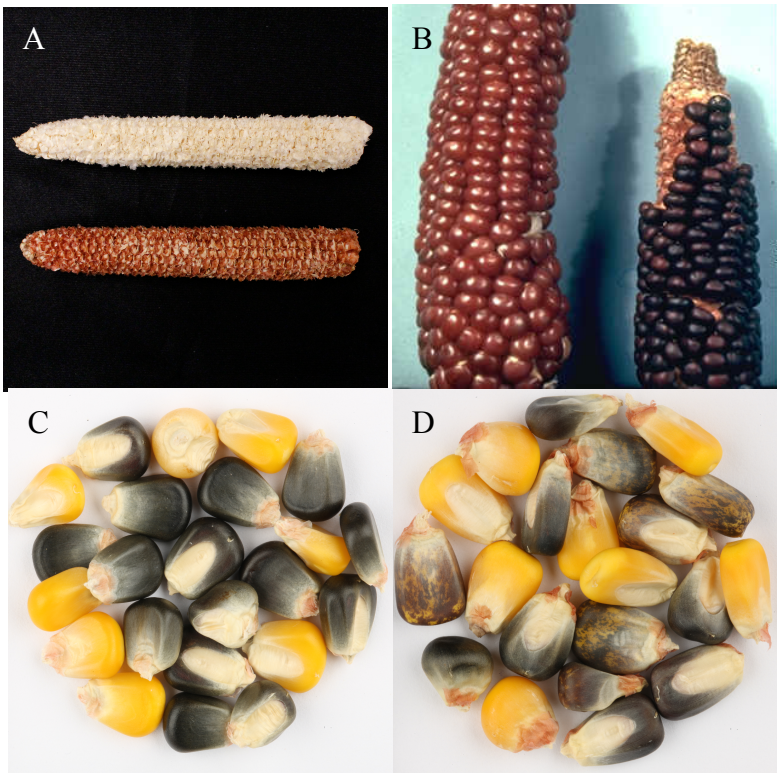


Figure 3.4: Histograms of two important phenotypes in the reduced acylation mapping population: a) TAC and b) Percentage of acylation.

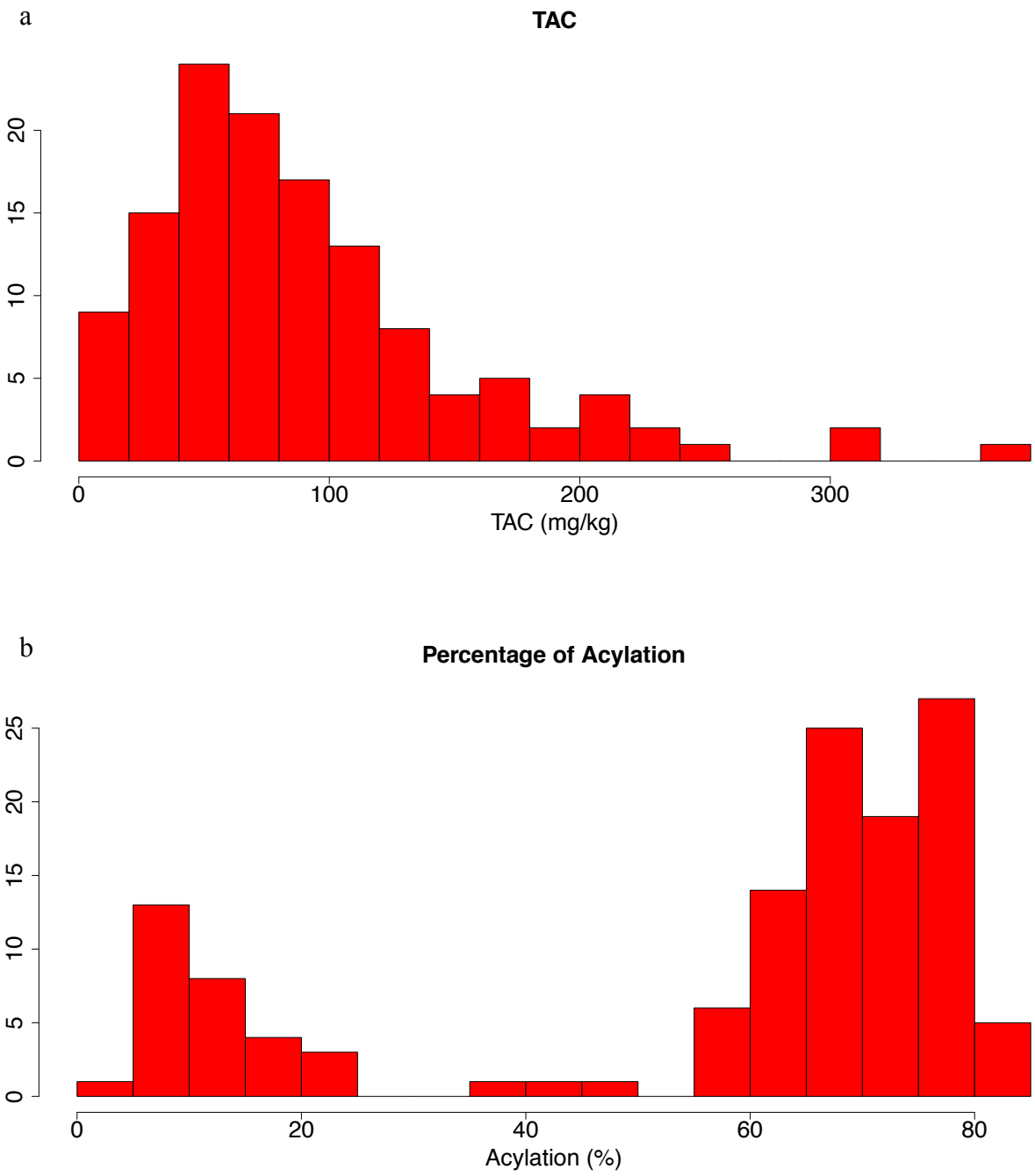


Figure 3.5: Control Trait QTL. Red dashed lines refer to known gene loci (Andorf et al., 2010).

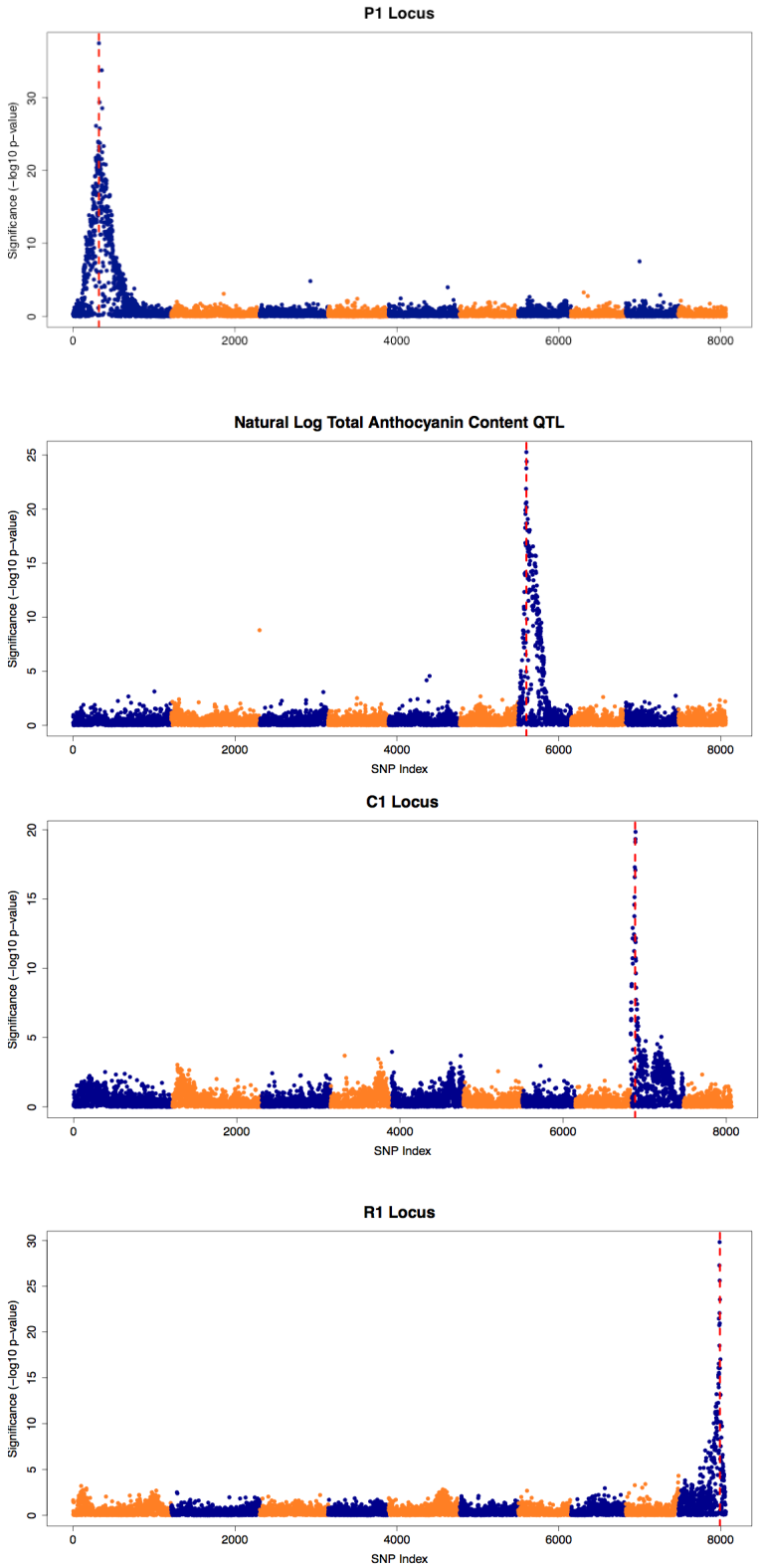
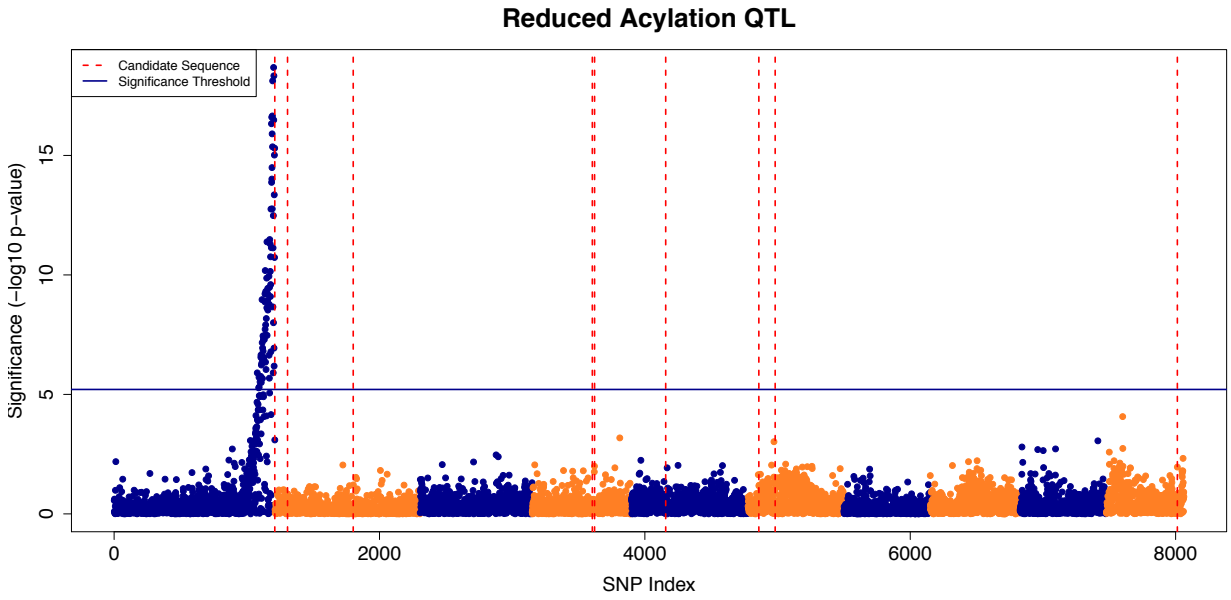


Figure 3.6: Reduced Acylation QTL. a) Genome-wide Manhattan plot with acyltransferase candidate genes marked with a red dashed line and the Bonferroni significance threshold with a solid blue line (–log10 *p*>5.2074). b) Manhattan plot for chromosome 1 plotted by physical SNP position with GRMZM2G387394 marked with a red dashed line.

a



b

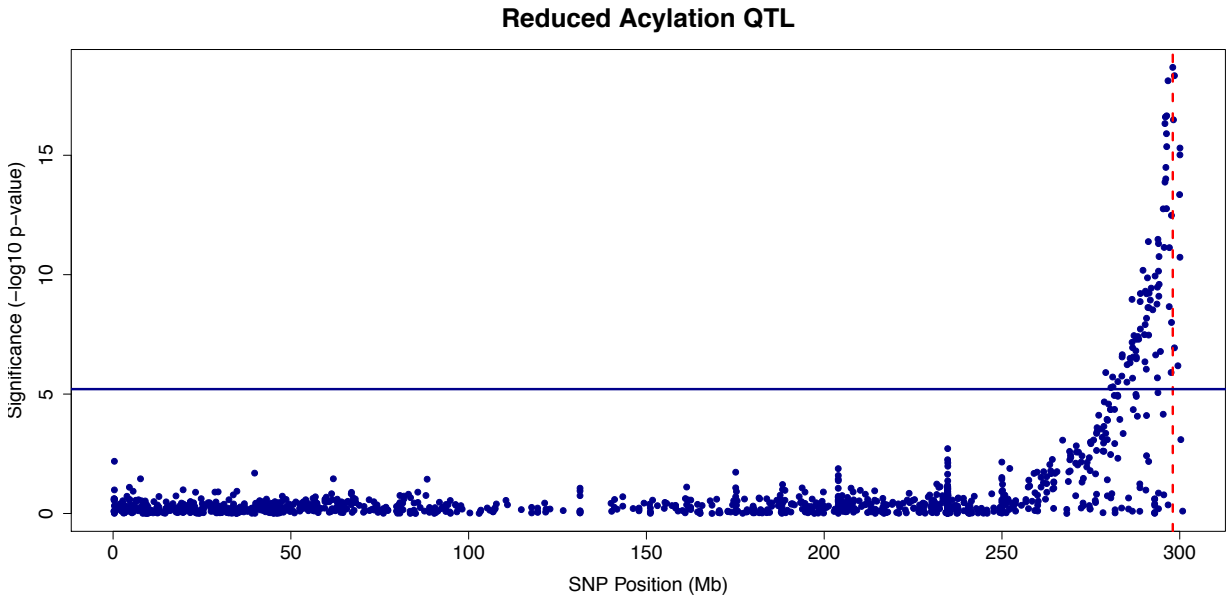


Figure 3.7: Box and whisker plot of percentage of acylation per genotypic marker

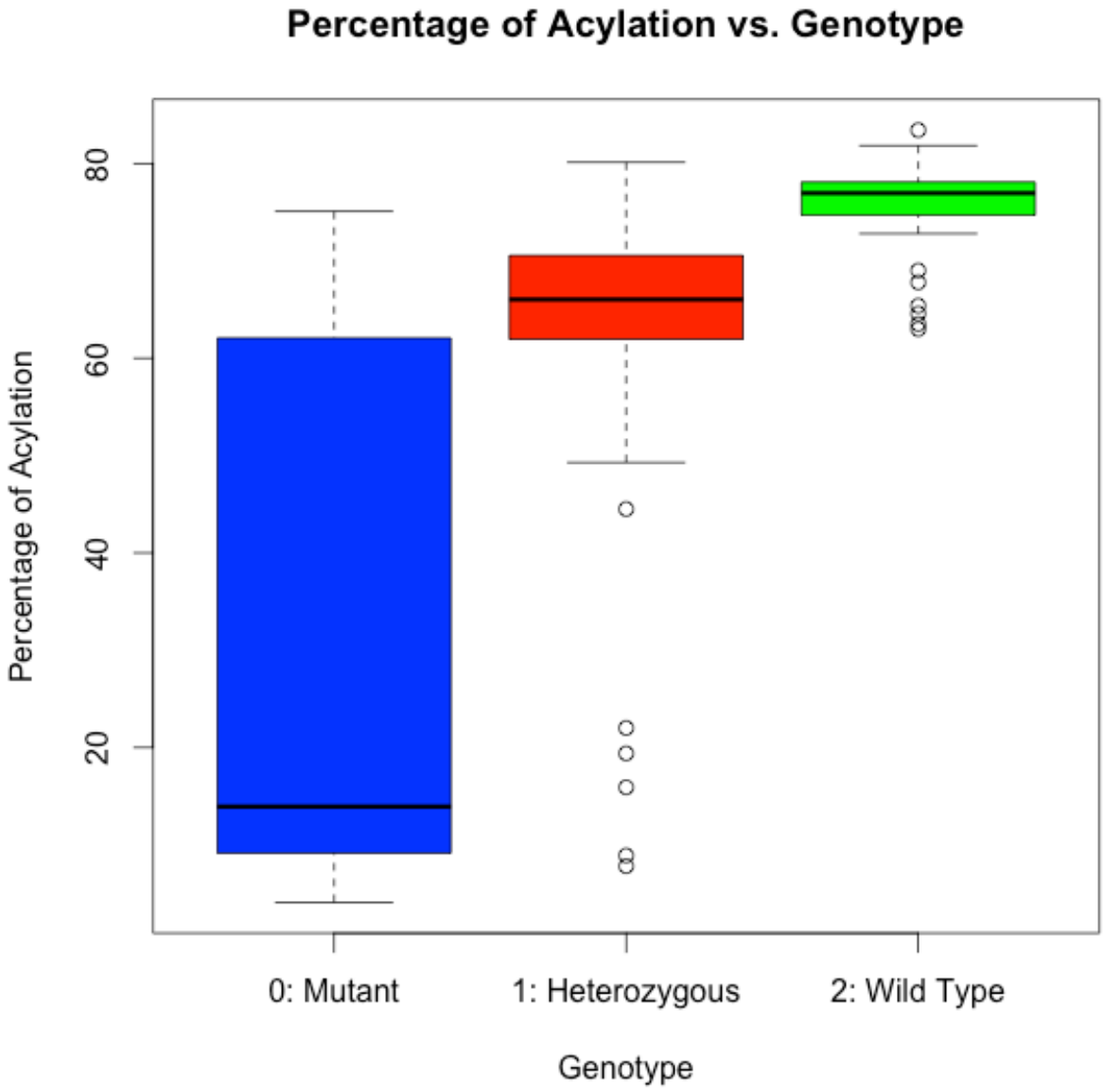
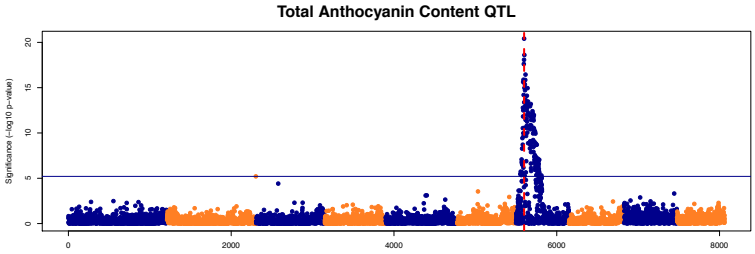
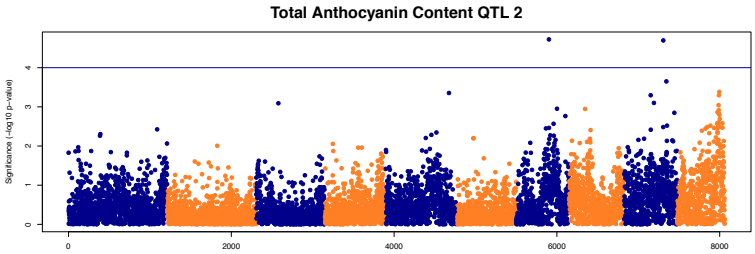


Figure 3.8: TAC Manhattan plots for step-wise QTL analysis. a) Untransformed TAC QTL with a Bonferroni significance threshold. b) TAC QTL with *in1* as a covariate and a $p<1\times10^{-4}$ significance threshold. c) TAC QTL with *in1* and S7_144238875 covariate and a $p<1\times10^{-4}$ significance threshold. d) TAC QTL with *in1,* S7_144238875, and S9_137169800 covariate and a $p<1\times10^{-4}$ significance threshold.

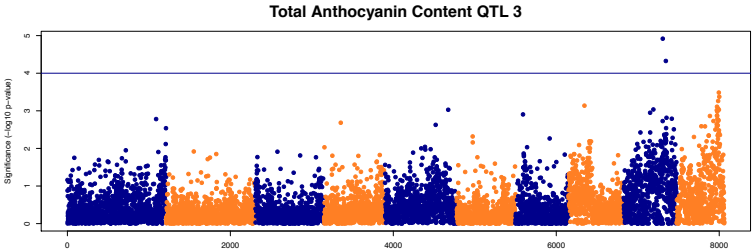
a



b



c



d

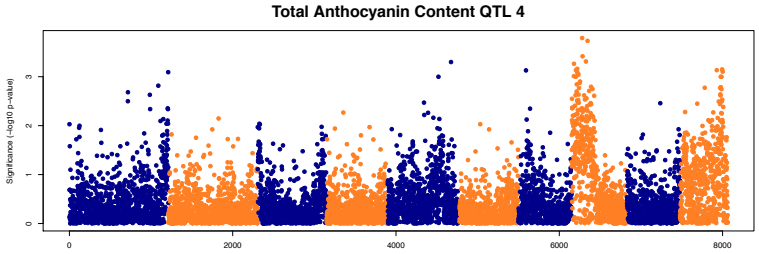


Figure 3.9: Manhattan plot for C3MG demonstrating significant QTL for *in1* and the acyltransferase candidate.

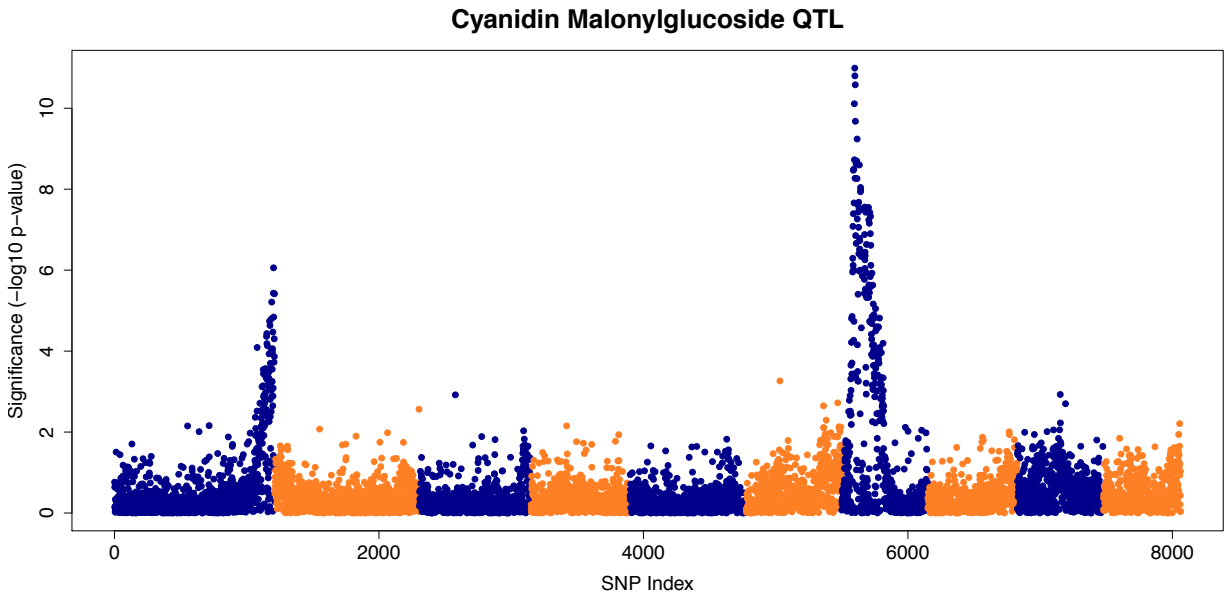


Figure 3.10: Chromatograms of *Mu* insertion crosses: a) Kernel without the *Mu* insertion demonstrating a wild-type anthocyanin profile; and b) Kernel with the *Mu* insertion returning the mutant phenotype. Peak labels correspond to ID # in Figure 3.1.

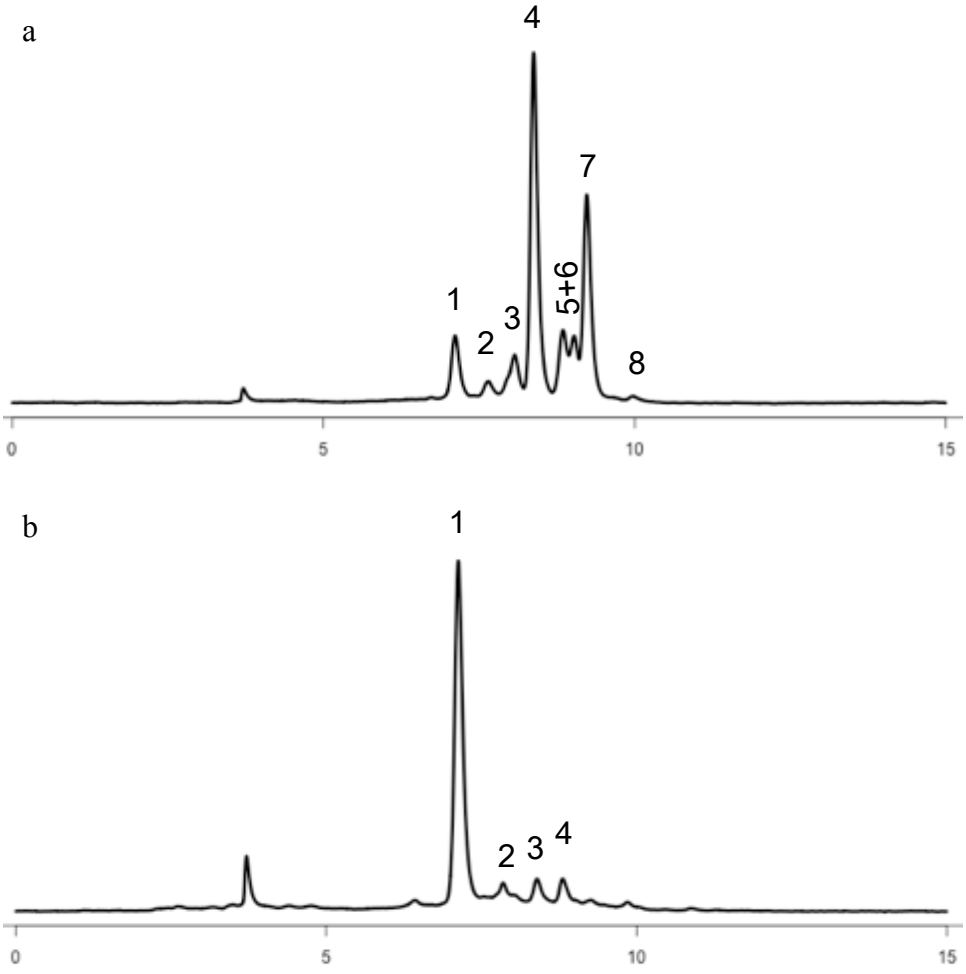
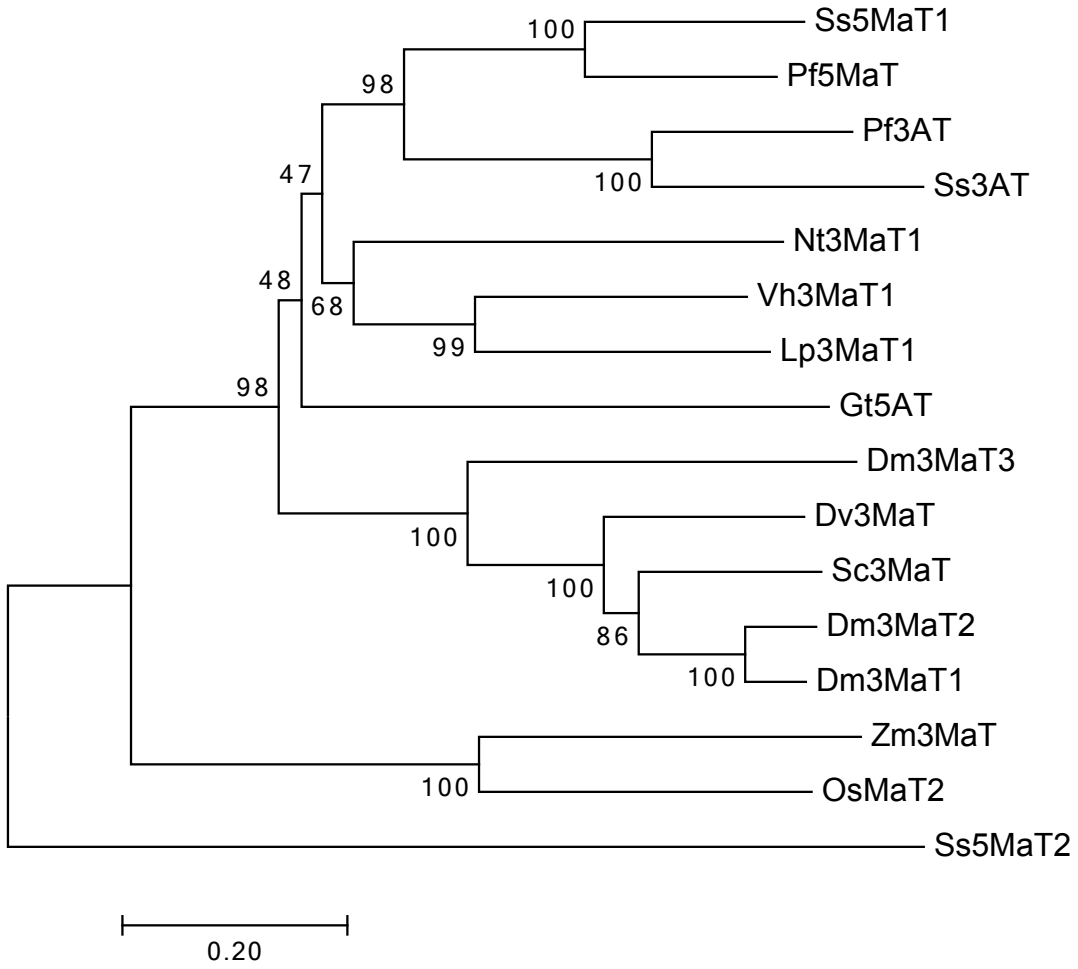


Figure 3.11: Amino acid alignment of *Dm3MaT2* with a predicted 3MaT in maize. Anthocyanin acyltransferase motif sequences are boxed. Amino acids highlighted in red are conserved. Amino acids marked with (*) are important for 3MaT2 activity according to Unno et al. (2007). Sequences were aligned in COBALT (Papadopoulos and Agarwala, 2007).

Figure 3.12: Evolutionary relationships of flavonoid acyltransferases



Neighbor-Joining phylogenetic tree of GenBank Accessions AAL50566.1 (Ss5MaT1), AAL50565.1 (Pf5MaT), Q9MBC1.1 (Pf3AT), AAR28757.1 (Ss3AT), BAD93691.1 (Nt3MaT1), AAS77402.1 (Vh3MaT1), AAS77404.1 (Lp3MaT1), BAA74428.1 (Gt5AT), BAF50706.1 (Dm3MaT3), Q8GSN8.1 (Dv3MaT), AAO38058.1 (Sc3MaT), AAQ63616.1 (Dm3MaT2), AAQ63615.1 (Dm3MaT1), NP_001148286.1 (Zm3MaT), NP_001046855.1 (OsMaT2), and AAR26385.1 (Ss5MaT2) (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 10,000 replicates is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site (Zuckerkandl and Pauling, 1965). All positions with less than 75% site coverage were eliminated. That is, fewer than 25% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 444 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

Table 3.1: Summary of phenotypes in the mapping population

	Allele	**Mutant**	***In1***	***R1***	***C1***	***P1***	**Total**
TAC (mg/kg)	Dominant	103.40	71.17	112.70	107.52	98.66	**96.19**
	Recessive	67.98	178.80	83.92	87.23	89.30	
Acylation (%)	Dominant	70.79	56.56	58.28	55.32	57.12	**56.92**
	Recessive	12.13	58.27	55.98	58.26	56.37	
Recessive Count		**29**	**30**	**75**	**72**	**34**	**129**

Table 3.2: Type 3 ANOVA for TAC and parameter estimates. a) The Type 3 ANOVA for TAC was calculated in Proc Mixed with all control loci set to random. b) Proc Reg calculation of parameter estimates with 95% confidence intervals for control loci that were significant in the Type 3 ANOVA of 3.2a.

a

Gene	DF	Mean Square	Expected Mean Square	Error	F-value	*p*-value
In1	1	225443	Var(Residual) + 41.672 Var(in1)	MSE	115.98	<.0001
C1	1	22390	Var(Residual) + 59.301 Var(C1)	MSE	11.52	0.0009
R1	1	12811	Var(Residual) + 59.62 Var(R1)	MSE	6.59	0.0115
Mutant	1	12220	Var(Residual) + 43.799 Var(mutant)	MSE	6.29	0.0135
P1	1	103.944161	Var(Residual) + 44.981 Var(P1)	MSE	0.05	0.8175
Residual	120	1943.788768	Var(Residual)	.	.	.

b

Gene	DF	Parameter Estimate	Standard Error	t Value	*p*-value	95% Confidence Limits	
Intercept	1	105.02519	8.47145	12.4	<.0001	88.25372	121.79666
In1	1	103.699	9.52151	10.89	<.0001	84.84867	122.54934
C1	1	-27.22009	7.98706	-3.41	0.0009	-43.03258	-11.40761
R1	1	-20.78432	8.0401	-2.59	0.0109	-36.70182	-4.86681
Mutant	1	-23.6494	9.38356	-2.52	0.013	-42.22664	-5.07217

Table 3.3: Candidate Anthocyanin Acyltransferases in Maize

Gene Annotation	**Gene Identifier (B73 RefGen_v3)**	**Chr**	**Start Position**	**End Position**
Transferase	GRMZM2G387394	1	300173138	300174772
Malonyl-CoA:anthocyanin 5-*O*-glucoside-6'''-*O*-malonyltransferase-like	GRMZM2G341253 / GRMZM2G436404	2	165106367	165108162
Anthocyanin 5-aromatic acyltransferase	LOC100281179	4	170843898	170845760
Malonyl-CoA:anthocyanidin 5-*O*-glucoside-6'''-*O*-malonyltransferase-like	GRMZM2G316787	4	174072855	174074677
Uncharacterized LOC100274392	GRMZM2G382785	5	50372211	50374054
Uncharacterized protein LOC100272672	LOC100272672	6	19122956	19124662
Anthocyanidin 3-*O*-glucoside 6''-*O*-acyltransferase-like	GRMZM2G095340	6	72337117	72341167
Transferase family protein	GRMZM5G800407	10	142328397	142330129

Table 3.4: Proportion of variance attributed to each control loci.

	In1 (%)	C1 (%)	R1 (%)	Mutant (%)	Residual (%)
C3G	35.71***	1.96*	0.28	41.52***	20.53
C3MG	35.67***	2.13*	1.55*	37.41***	23.25
Pg3DMG	29.68***	1.17	3.09*	6.81*	59.24
Pg3G	35.77***	1.09	1.16	37.73***	24.26
Pg3MG	41.49***	2.61*	1.69*	24.88***	29.33
Pn3G	58.57***	5.77**	5.27**	1.79	28.61
Pn3MG/C3DMG	28.52***	1.82*	2.59*	42.34***	24.72
TAC	66.49***	4.22**	2.29*	2.94*	24.05
Peak ID #6	37.70***	1.92*	0.2	26.73***	33.46

*** $p<0.0001$ significance
** $p<0.001$ significance
* $p<0.05$ significance

Table 3.5: QTL Simulation Study

A) *R1* Simulation

Sample Size	Significant Markers	Within 10 Mb*	False Positives	Avg. Proportion Recessive *	Min. Proportion Recessive *	Avg. *p*-value (-log10) *	Avg. Distance From Known Locus (bp) *
20	69	44	25%	55.58%	30.00%	61.00	8,646,266
40	100	100	0%	58.25%	42.50%	10.99	2,146,341
60	100	100	0%	57.78%	45.00%	14.80	1,294,528
80	100	100	0%	58.06%	51.25%	19.11	1,167,359
100	100	100	0%	57.81%	52.00%	23.78	1,144,927

B) Reduced Acylation Simulation

Sample Size	Significant Markers	Within 10 Mb*	False Positives	Avg. Proportion Recessive *	Min. Proportion Recessive *	Avg. *p*-value (-log10) *	Avg. Distance From Known Locus (bp) *
20	58	31	27%	30.00%	15.00%	6.46	5,751,593
40	99	93	6%	25.00%	12.50%	7.88	2,990,774
60	100	99	1%	23.33%	16.67%	10.48	2,395,972
80	100	100	0%	22.50%	13.75%	12.75	2,172,614
100	100	100	0%	23.00%	15.00%	15.22	2,174,669

* Only if significant and in the right location
Significance with a Bonferroni correction at -log10(p)≥5.2074

3.9 References

Abdel-Aal, E.-S. M.; Young, J. C.; Rabalski, I. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J. Agric. Food Chem.* **2006**, *54* (13), 4696–4704.

Agarwal, T.; Grotewold, E.; Doseff, A. I.; Gray, J. MYB31/MYB42 Syntelogs Exhibit Divergent Regulation of Phenylpropanoid Genes in Maize, Sorghum and Rice. *Sci. Rep.* **2016**, *6*, 28502.

Andorf, C. M.; Lawrence, C. J.; Harper, L. C.; Schaeffer, M. L.; Campbell, D. A.; Sen, T. Z. The Locus Lookup Tool at MaizeGDB: Identification of Genomic Regions in Maize by Integrating Sequence Information with Physical and Genetic Maps. *Bioinformatics* **2010**, *26* (3), 434–436.

Bakowska-Barczak, A. Acylated Anthocyanins as Stable, Natural Food Colorants - a Review. *Pol. J. Food Nutr. Sci.* **2005**, *14/15* (2), 107–116.

Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y.; Buckler, E. S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics* **2007**, *23* (19), 2633–2635.

Browning, B. L.; Browning, S. R. Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* **2016**, *98* (1), 116–126.

Burr, F. A.; Burr, B.; Scheffler, B. E.; Blewitt, M.; Wienand, U.; Matz, E. C. The Maize Repressor-like Gene Intensifier1 Shares Homology with the R1/B1 Multigene Family of Transcription Factors and Exhibits Missplicing. *Plant Cell* **1996**, *8* (8), 1249–1259.

Chen, Z.; Wang, B.; Dong, X.; Liu, H.; Ren, L.; Chen, J.; Hauck, A.; Song, W.; Lai, J. An Ultra-High Density Bin-Map for Rapid QTL Mapping for Tassel and Ear Architecture in a Large F2 Maize Population. *BMC Genomics* **2014**, *15* (1), 433.

D'Auria, J. C. Acyltransferases in Plants: A Good Time to Be BAHD. *Curr. Opin. Plant Biol.* **2006**, *9* (3), 331–340.

Dellaporta, S. L.; Greenblatt, I.; Kermicle, J. L.; Hicks, J. B.; Wessler, S. R. Molecular Cloning of the Maize R-Nj Allele by Transposon Tagging with Ac. In *Chromosome structure and function*; Springer, 1988; pp 263–282.

Doyle, J. J.; Doyle, J. L. Isolation of Plant DNA from Fresh Tissue. *Focus* **1990**, *12*, 13–15.

East, E. M. (Edward M.; Hayes, H. K. *Inheritance in Maize*; New Haven, Conn. : Connecticut Agricultural Experiment Station, 1911.

Eichten, S. R.; Foerster, J. M.; Leon, N. de; Kai, Y.; Yeh, C.-T.; Liu, S.; Jeddeloh, J. A.; Schnable, P. S.; Kaeppler, S. M.; Springer, N. M. B73-Mo17 Near-Isogenic Lines Demonstrate Dispersed Structural Variation in Maize. *Plant Physiol.* **2011**, *156* (4), 1679–1690.

Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S.; Mitchell, S. E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6* (5), e19379.

Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **1985**, *39* (4), 783–791.

Fragoso, C. A.; Heffelfinger, C.; Zhao, H.; Dellaporta, S. L. Imputing Genotypes in Biallelic Populations from Low-Coverage Sequence Data. *Genetics* **2016**, *202* (2), 487–495.

Glaubitz, J. C.; Casstevens, T. M.; Lu, F.; Harriman, J.; Elshire, R. J.; Sun, Q.; Buckler, E. S. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **2014**, *9* (2), e90346.

Goodstein, D. M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R. D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40* (Database issue), D1178–D1186.

Goto, T.; Kondo, T. Structure and Molecular Stacking of Anthocyanins—Flower Color Variation. *Angew. Chem. Int. Ed. Engl.* **1991**, *30* (1), 17–33.

Grotewold, E.; Drummond, B. J.; Bowen, B.; Peterson, T. The Myb-Homologous P Gene Controls Phlobaphene Pigmentation in Maize Floral Organs by Directly Activating a Flavonoid Biosynthetic Gene Subset. *Cell* **1994**, *76* (3), 543–553.

He, J.; Giusti, M. M. Anthocyanins: Natural Colorants with Health-Promoting Properties. *Annu. Rev. Food Sci. Technol.* **2010**, *1*, 163–187.

He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.; Li, Z. Genotyping-by-Sequencing (GBS), an Ultimate Marker-Assisted Selection (MAS) Tool to Accelerate Plant Breeding. *Front. Plant Sci.* **2014**, *5*.

Heffelfinger, C.; Fragoso, C. A.; Moreno, M. A.; Overton, J. D.; Mottinger, J. P.; Zhao, H.; Tohme, J.; Dellaporta, S. L. Flexible and Scalable Genotyping-by-Sequencing Strategies for Population Studies. *BMC Genomics* **2014**, *15* (1), 1.

Irani, N. G.; Hernandez, J. M.; Grotewold, E. Chapter Three Regulation of Anthocyanin Pigmentation. *Recent Adv. Phytochem.* **2003**, *37*, 59–78.

Jing, P.; Noriega, V.; Schwartz, S. J.; Giusti, M. M. Effects of Growing Conditions on Purple Corncob (*Zea Mays* L.) Anthocyanins. *J. Agric. Food Chem.* **2007**, *55* (21), 8625–8629.

Kim, D. H.; Kim, S. K.; Kim, J.-H.; Kim, B.-G.; Ahn, J.-H. Molecular Characterization of Flavonoid Malonyltransferase from Oryza Sativa. *Plant Physiol. Biochem.* **2009**, *47* (11–12), 991–997.

Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, msw054.

Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357–359.

Lee, M.; Sharopova, N.; Beavis, W. D.; Grant, D.; Katt, M.; Blair, D.; Hallauer, A. Expanding the Genetic Map of Maize with the Intermated B73 X Mo17 (IBM) Population. *Plant Mol. Biol.* **2002**, *48* (5–6), 453–461.

Li, X.; Zhou, Z.; Ding, J.; Wu, Y.; Zhou, B.; Wang, R.; Ma, J.; Wang, S.; Zhang, X.; Xia, Z.; et al. Combined Linkage and Association Mapping Reveals QTL and Candidate Genes for Plant and Ear Height in Maize. *Plant Genet. Genomics* **2016**, 833.

McCarty, D. R.; Settles, A. M.; Suzuki, M.; Tan, B. C.; Latshaw, S.; Porch, T.; Robin, K.; Baier, J.; Avigne, W.; Lai, J.; et al. Steady-State Transposon Mutagenesis in Inbred Maize. *Plant J. Cell Mol. Biol.* **2005**, *44* (1), 52–61.

Nakayama, T.; Suzuki, H.; Nishino, T. Anthocyanin Acyltransferases: Specificities, Mechanism, Phylogenetics, and Applications. *J. Mol. Catal. B Enzym.* **2003**, *23* (2–6), 117–132.

Papadopoulos, J. S.; Agarwala, R. COBALT: Constraint-Based Alignment Tool for Multiple Protein Sequences. *Bioinforma. Oxf. Engl.* **2007**, *23* (9), 1073–1079.

Paz-Ares, J.; Ghosal, D.; Wienand, U.; Peterson, P. A.; Saedler, H. The Regulatory c1 Locus of Zea Mays Encodes a Protein with Homology to Myb Proto-Oncogene Products and with Structural Similarities to Transcriptional Activators. *EMBO J.* **1987**, *6* (12), 3553–3558.

Poland, J. A.; Brown, P. J.; Sorrells, M. E.; Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE* **2012**, *7* (2), e32253.

R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

Saito, N.; Toki, K.; Honda, T.; Kawase, K. Cyanidin 3-Malonylglucuronylglucoside in Bellis and Cyanidin 3-Malonylglucoside in Dendranthema. *Phytochemistry* **1988**, *27* (9), 2963–2966.

Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4* (4), 406–425.

Salinas-Moreno, Y.; Pérez-Alonso, J. J.; Vázquez-Carrillo, G.; Aragón-Cuevas, F.; Velázquez-Cardelas, G. A. Antocianinas Y Actividad Antioxidante En Maíces (Zea Mays L.) de Las Razas Chalqueño, Elotes Cónicos Y Bolita. *Agrociencia* **2012**, *46* (7), 693–706.

Sharma, M.; Cortes-Cruz, M.; Ahern, K. R.; McMullen, M.; Brutnell, T. P.; Chopra, S. Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. *Genetics* **2011**, *188* (1), 69–79.

St-Pierre, B.; De Luca, V. Evolution of Acyltransferase Genes: Origin and Diversification of the BAHD Superfamily of Acyltransferases Involved in Secondary Metabolism. In *Recent Advances in Phytochemistry*; Elsevier, 2000; Vol. 34, pp 285–315.

Sturaro, M.; Hartings, H.; Schmelzer, E.; Velasco, R.; Salamini, F.; Motto, M. Cloning and Characterization of GLOSSY1, a Maize Gene Involved in Cuticle Membrane and Wax Production. *Plant Physiol.* **2005**, *138* (1), 478–489.

Suzuki, H.; Nakayama, T.; Yonekura-Sakakibara, K.; Fukui, Y.; Nakamura, N.; Yamaguchi, M.; Tanaka, Y.; Kusumi, T.; Nishino, T. cDNA Cloning, Heterologous Expressions, and Functional Characterization of Malonyl-Coenzyme A:Anthocyanidin 3-O-Glucoside-6"-O-Malonyltransferase from Dahlia Flowers. *Plant Physiol.* **2002**, *130* (4), 2142–2151.

Suzuki, H.; Nakayama, T.; Yamaguchi, M.; Nishino, T. cDNA Cloning and Characterization of Two Dendranthema X Morifolium Anthocyanin Malonyltransferases with Different Functional Activities. *Plant Sci.* **2004a**, *166* (1), 89–96.

Suzuki, H.; Sawada, S. 'ya; Watanabe, K.; Nagae, S.; Yamaguchi, M.; Nakayama, T.; Nishino, T. Identification and Characterization of a Novel Anthocyanin Malonyltransferase from Scarlet Sage (Salvia Splendens) Flowers: An Enzyme That Is Phylogenetically Separated from Other Anthocyanin Acyltransferases. *Plant J.* **2004b**, *38* (6), 994–1003.

Swarts, K.; Li, H.; Romero Navarro, J. A.; An, D.; Romay, M. C.; Hearne, S.; Acharya, C.; Glaubitz, J. C.; Mitchell, S.; Elshire, R. J.; et al. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. *Plant Genome* **2014**, *7* (3), 0.

Unno, H.; Ichimaida, F.; Suzuki, H.; Takahashi, S.; Tanaka, Y.; Saito, A.; Nishino, T.; Kusunoki, M.; Nakayama, T. Structural and Mutational Studies of Anthocyanin Malonyltransferases Establish the Features of BAHD Enzyme Catalysis. *J. Biol. Chem.* **2007**, *282* (21), 15812–15822.

Widrlechner, M. P.; Dragula, S. K. Eleven Ornamental Corn Inbreds: Lines OC1 through OC11. *HortScience* **1992**, *27* (12), 1338–1339.

Yonekura-Sakakibara, K.; Nakayama, T.; Yamazaki, M.; Saito, K. Modification and Stabilization of Anthocyanins. In *Anthocyanins*; Winefield, C., Davies, K., Gould, K., Eds.; Springer New York: New York, NY, 2008; pp 169–190.

Youens-Clark, K.; Buckler, E.; Casstevens, T.; Chen, C.; DeClerck, G.; Derwent, P.; Dharmawardhana, P.; Jaiswal, P.; Kersey, P.; Karthikeyan, A. S.; et al. Gramene

Database in 2010: Updates and Extensions. *Nucleic Acids Res.* **2011**, *39* (suppl 1), D1085–D1094.

Yu, X.-H.; Gou, J.-Y.; Liu, C.-J. BAHD Superfamily of Acyl-CoA Dependent Acyltransferases in Populus and Arabidopsis: Bioinformatics and Gene Expression. *Plant Mol. Biol.* **2009**, *70* (4), 421–442.

Zhao, J.; Huhman, D.; Shadle, G.; He, X.-Z.; Sumner, L. W.; Tang, Y.; Dixon, R. A. MATE2 Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in Medicago Truncatula. *Plant Cell* **2011**, *23* (4), 1536–1555.

Zhou, Z.; Zhang, C.; Zhou, Y.; Hao, Z.; Wang, Z.; Zeng, X.; Di, H.; Li, M.; Zhang, D.; Yong, H.; et al. Genetic Dissection of Maize Plant Architecture with an Ultra-High Density Bin Map Based on Recombinant Inbred Lines. *BMC Genomics* **2016**, *17*, 1–15.

Zuckerkandl, E.; Pauling, L. Evolutionary Divergence and Convergence in Proteins. *Evol. Genes Proteins* **1965**, *97*, 97–166.

Chapter 4: Summary and Future Prospects

Maize has great potential as an economic source of natural colors. Although the concentration in the kernel is not as high as some fruit and vegetable extracts, maize has value-added benefits these other plant species do not (Wu et al., 2006). A specialty corn supply chain is already in place that uses purple corn varieties for the purpose of extracting anthocyanins for use in food and beverages. Commercial milling processes are equipped to remove anthocyanin-rich pericarp fractions to concentrate anthocyanin-producing tissue. Once pericarp is recovered, the rest of the kernel can still be utilized for food, fuel, and feed. In this way, integrating anthocyanins into new maize hybrids will potentially reduce natural color costs and offer a value-added coproduct to the corn supply chain. This and future research focuses on developing maize hybrids adapted to Midwest growing conditions to provide a more economic source of natural colors.

In Chapter 2, a large collection of pigmented maize was collected to survey the diversity of anthocyanin production in maize germplasm. It was found that pericarp-pigmented accessions capable of producing condensed forms were generally the highest-performing lines in terms of total anthocyanin content (TAC). Andean purple corn varieties are currently being crossed to high grain-yielding inbreds to increase TAC in Midwest-adapted germplasm. Another result of Chapter 2 shows that anthocyanin production tended to have high heritability with 90.8% of variation in TAC attributed to genetics. This supports the feasibility of breeding for increased TAC in maize. Future work should be focused on developing populations of pericarp-pigmented accessions to test genetic variation in pericarp pigment production. Heritability will be re-tested in the future with either $F_{2:3}$ families, RILs, or hybrids of these pericarp-pigmented populations by growing them in several locations over several years. Genotyping these populations with

genotyping-by-sequencing (GBS) could also help discover QTL associated with high TAC for marker-assisted selection. Genomic selection of anthocyanin production would also be an interesting route for improvement of TAC (Heffner et al., 2009).

Another route to improve TAC in maize would be to increase the volume of anthocyanin-producing tissue. This can be accomplished by increasing pericarp thickness or introducing multiple aleurone layers. Pericarp thickness is a quantitative trait controlled by a few genes with pleiotropic effects, while the inheritance of multiple aleurone layers is unknown but most likely attributed to only a few genes (Wolf et al., 1972; Choe, 2010). Our lab has developed a mapping population between Midwestern inbred Mo17 and Peruvian landrace San Martin 105. The landrace has a thin pericarp, but a thick aleurone due to at least four aleurone layers. Mo17 is a typical Midwest inbred with a thick pericarp and single aleurone layer. Future work will determine QTL associated with thick pericarps and multiple aleurone layer formation so marker-assisted selection could increase these tissues.

The anthocyanin biosynthetic pathway is the most well-studied plant secondary metabolite pathway in plants due to its conspicuous nature and wide variation (Irani et al., 2003). Although it is a well understood pathway, there is still much to learn, especially in maize. In Chapter 2, a few unique accessions were found that could not produce abundant amounts of acylated anthocyanins, which are otherwise the most prominent forms of anthocyanins in maize. Chapter 3 presented a genetic characterization of this trait. A reduced acylation mutant crossed to B73 was used to create a segregating mapping population and was genotyped using GBS technology. The investigation lead to the discovery of a candidate anthocyanin acyltransferase in the maize genome. A UniformMu (McCarty et al., 2005) *Mu* transposon knockout of this candidate gene provided by the Maize Genetics Cooperation Stock Center confirmed that this

gene is sufficient to produce the reduced acylation phenotype when non-functional, confirming anthocyanin acyltransferase function in this gene. The next steps to characterize this gene would be to perform an enzyme feeding assay in which anthocyanins and acyl-donor substrates are added in lieu of a functional copy of the candidate acyltransferase in maize. Alternatively, the gene could be inserted into a plant host that does not produce acylated anthocyanins. In another study, petunia was already demonstrated as a good model for anthocyanin acyltransferase activity (Suzuki et al., 2002). Results of the enzyme feeding assay or Petunia transformation would unequivocally conclude that the candidate anthocyanin acyltransferase can produce all acylated anthocyanins in maize.

Although the anthocyanin peonidin is not a predominant compound in maize, it still may have implications on stability in food and beverage systems (Cabrita et al., 2000). Future work will focus on exploring the diversity in peonidin production in the collection in depth. A major obstacle working with peonidin is that the HPLC method utilized co-elutes peonidin 3-malonylglucoside with cyanidin 3-dimalonylglucoisde. A solution around this was to model pelargonidin 3-dimalonylglucoside formation based on the abundance of other compounds in homozygous *pr1* recessive lines. The validity of this model has not been tested, but more samples would help strengthen the model. A way to quantify peonidin content would be to perform acid hydrolysis or saponification on *Pr1* dominant lines. These two techniques break down anthocyanins into their anthocyanidin or 3-glucoside forms, respectively, to simplify pigment quantification (Moreno et al., 2005; Abdel-Aal et al., 2006). In addition, altering the mobile phase or changing chromatographic columns might be an alternative means of separating the two pigments. Once peonidin diversity is more adequately described, future studies could focus on developing mapping populations to discover QTL associated with peonidin content.

The goal would be that anthocyanin *O*-methyltransferases in maize can be discovered just as the anthocyanin acyltransferase was discovered in Chapter 3 by utilizing GBS.

In addition to peonidin production, condensed form synthesis is another interesting trait for which the genetic basis is unknown. Currently, our lab is performing a GBS study on the genetic basis for anthocyanin diversity in Apache Red. Apache Red segregates for the ability to produce condensed forms, which makes it the perfect accession for a condensed form synthesis study.

Overall the anthocyanin pathway is well-studied, but the genetics behind increasing TAC need to be investigated more thoroughly. Populations of pericarp-pigmented lines crossed to Midwestern inbreds will help explore the heritability of TAC and provide a means to investigate QTL associated with increased TAC. Increasing the volume of anthocyanin-producing tissue can help increase TAC. The genetics of increased pericarp thickness and multiple aleurone layer formation need to be investigated with the mapping population already in place. Finally, although all anthocyanin biosynthetic genes have been characterized, the genes compounding the diversity in anthocyanin production need to be characterized. These genes include the candidate maize anthocyanin acyltransferase from Chapter 3, maize anthocyanin *O*-methyltransferases, and condensed form synthesis genes. Finding genes that increase TAC and stability will increase maize's potential as an economic source of natural colors.

4.1 References

Abdel-Aal, E.-S. M.; Young, J. C.; Rabalski, I. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J. Agric. Food Chem.* **2006**, *54* (13), 4696–4704.

Cabrita, L.; Fossen, T.; Andersen, Ø. M. Colour and Stability of the Six Common Anthocyanidin 3-Glucosides in Aqueous Solutions. *Food Chem.* **2000**, *68* (1), 101–107.

Choe, E. Marker Assisted Selection and Breeding for Desirable Thinner Pericarp Thickness and Ear Traits in Fresh Market Waxy Corn Germplasm, University of Illinois at Urbana-Champaign, 2010.

Heffner, E. L.; Sorrells, M. E.; Jannink, J.-L. Genomic Selection for Crop Improvement. *Crop Sci.* **2009**, *49* (1), 1–12.

Irani, N. G.; Hernandez, J. M.; Grotewold, E. Chapter Three Regulation of Anthocyanin Pigmentation. *Recent Adv. Phytochem.* **2003**, *37*, 59–78.

McCarty, D. R.; Settles, A. M.; Suzuki, M.; Tan, B. C.; Latshaw, S.; Porch, T.; Robin, K.; Baier, J.; Avigne, W.; Lai, J.; et al. Steady-State Transposon Mutagenesis in Inbred Maize. *Plant J. Cell Mol. Biol.* **2005**, *44* (1), 52–61.

Moreno, Y. S.; Sánchez, G. S.; Hernández, D. R.; Lobato, N. R. Characterization of Anthocyanin Extracts from Maize Kernels. *J. Chromatogr. Sci.* **2005**, *43* (9), 483–487.

Suzuki, H.; Nakayama, T.; Yonekura-Sakakibara, K.; Fukui, Y.; Nakamura, N.; Yamaguchi, M.; Tanaka, Y.; Kusumi, T.; Nishino, T. cDNA Cloning, Heterologous Expressions, and Functional Characterization of Malonyl-Coenzyme A:Anthocyanidin 3-O-Glucoside-6"-O-Malonyltransferase from Dahlia Flowers. *Plant Physiol.* **2002**, *130* (4), 2142–2151.

Wolf, M. J.; Cutler, H. C.; Zuber, M. S.; Khoo, U. Maize with Multilayer Aleurone of High Protein Content. *Crop Sci.* **1972**, *12* (4), 440.

Wu, X.; Beecher, G. R.; Holden, J. M.; Haytowitz, D. B.; Gebhardt, S. E.; Prior, R. L. Concentrations of Anthocyanins in Common Foods in the United States and Estimation of Normal Consumption. *J. Agric. Food Chem.* **2006**, *54* (11), 4069–4075.