A Study in Purple:
An Investigation into the Genetics of Anthocyanin Accumulation in *Cannabis sativa*

Carter Prillaman

A thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

(Plant Breeding and Plant Genetics)

At the

UNIVERSITY OF WISCONSIN-MADISON

2025

Committee members:

Dr. Shelby Ellison, Assistant Professor, Plant and Agroecosystem Sciences

Dr. Irwin Goldman, Professor, Plant and Agroecosystem Sciences

Dr. Patrick Masson, Professor, Genetics

Acknowledgements

I would like to first and foremost thank Dr. Shelby Ellison for her unwavering support, patience, and commitment to not only my project, but to my own personal success throughout my masters. Shelby has taught me most of what I know about how to be a good scientist and her advice, guidance, and willingness to tell me that I am wrong have been a source of confidence and growth for me. Shelby gave me the opportunity to prove to myself that I am capable, competent, and most of all, that I can improve at almost anything if I set my mind to it, and for this I am eternally grateful.

I would like to thank my undergraduate advisor, Dr. Joshua Puzey, for showing me that science is only a tool to understand the world and not a definitive truth. Josh was a source of advice, support, and much needed laughter during difficult years in my undergraduate career, and his contributions to my success are manifold.

I would like to thank my committee member Dr. Patrick Masson for his infectious excitement for science. Dr. Masson seemed just as excited to ask me about future directions in my research as he was teaching me about the genetics of self-incompatibility, and I am very grateful for his interest and advice throughout my masters.

I am grateful to my committee member Dr. Irwin Goldman for his advice throughout my master's tenure and for his willingness to step just outside of the boundaries of traditional science to ask me thought-provoking questions. His combination of depth of

knowledge and expertise in plant breeding with joy and interest in the bigger picture is an inspiration to me.

I would like to extend my deepest gratitude to my Plant Breeding and Plant Genetics cohort Avelina Gaston, Andrea Garberding, Heather Wodehouse, Hudson Koch, Anson Sui, Amparo Daza Castanon, and Tony Schuh for their support, kind words, knowledge, and company throughout my time. I have learned as much from the people around me as I have from experiences in the classroom and the lab.

I am immensely grateful to Phillip Alberti for his hard work and advice in many fields throughout my graduate career. Phil was instrumental in encouraging me to improve my organizational skills (sorely needed) and was a constant source of advice on how to continue improving. Phil not only helped me organize every greenhouse experiment that I ran, but he also taught me how to effectively and efficiently manage growing plants. I am very grateful to Phil for patiently and properly teaching me much that I thought I already knew.

I am extremely grateful to my fellow graduate students in the Ellison lab, Jillian McGhee, Rebecca Arias, Grace Connelly, and Sean Kim, who laid the foundation for this work. Their support, willingness to listen, and camaraderie were crucial to my success and to my ability to weather setbacks and truly enjoy successes.

I would like to extend a special thank you to the undergraduate Ellison lab members, Sirapop Kiatchaipipat (in particular), Ava Hoodjer, Ella Cain, Dylan Mahant, Chloe Zimmer, and Cheyenne Mattie for their hard work and willingness to learn. They helped me find joy in teaching and mentoring others.

I would also like to thank Ellison lab members Dr. Derrick Grunwald and Robert Witherell for teaching me all that I know about bench science, and Dr. Eliot Stanton and Dr. Ademola Aina for teaching me the rest that I know about coding and bioinformatics. Without their support, I would not have been able to complete much of the analyses for this thesis.

I am extremely grateful to the Wisconsin Crop Innovation Center molecular technologies department and cannabis plant transformation lab members, Dr. Mike Peterson, Dr. Cindy Wang, Dr. Robert Sears, Dr. Shawn Kaeppler, and Dr. Ray Collier for coaching me through learning advanced molecular techniques and tissue culture. Without their expertise and willingness to teach me, there would not be a third chapter of my thesis at all.

I would like to thank my partner Adrianne Kinsey, who has never stopped believing in my ability to succeed. I extend a big thank you to Jamie and Chloe Godfrey, who have relentlessly helped me survive the rigors of graduate school in more ways than I can count. I would like to finally thank my family for their love and encouragement as I left home and set out to find my place in the world. They have been there for me throughout my whole life, and it has not stopped now.

Abstract

Cannabis sativa L. is an emerging crop with a variety of uses for its nutritious grains, long bast fibers, and significant public interest in its secondary metabolites. Anthocyanins are a class of secondary metabolites which provide visible purple and pink pigments in plants, driving consumer preferences in the floral cannabis market and enhancing abiotic stress tolerance. Identifying and characterizing the candidate gene(s) responsible for anthocyanin accumulation in *C. sativa* will create new opportunities for developing cultivars with heightened levels of these compounds. Prior mapping studies implicated a region on chromosome 6 putatively influencing leaf and flower pigmentation traits in two independent populations (S. Kim, Unpublished). To help identify potential candidate genes responsible for anthocyanin biosynthesis in leaf, flower, and stem tissues, and validate previous marker associations in *C. sativa*, an F_1 population replicated from prior studies was used to identify significant SNP marker clusters associated with each trait using a genome wide association study (GWAS). This analysis validated the presence of a significant marker cluster on chromosome 6 from 66.75Mb-75.77Mb associated with flower anthocyanin, and containing the putative subgroup 6 R2R3-MYB transcription factor *CsMYB78*. A new significant marker cluster on chromosome 1associated with stem and leaf anthocyanin was also revealed from 85.08Mb-87.92Mb. Following this, *CsMYB78* and *CsMYB33*, previously reported candidate genes from Kundan et al. (2022) were transformed into *C. sativa* plants at the Wisconsin Crop Innovation Center (WCIC). Initial results of generated T_0 plants show a counterintuitive pattern of pigmentation with *CsMYB78* failing to induce strong anthocyanin accumulation and *CsMYB33* inducing strong accumulation across tissues. The results of these

experiments will be invaluable in future breeding efforts focused on creating and maintaining purple cannabis cultivars, as well as in advancing our understanding of transcriptional regulation mechanisms of cannabis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

GWAS: Genome Wide Association Study

WCIC: Wisconsin Crop Innovation Center

THC: Tetrahydrocannabidiol

CBD: Cannabidiol

THCA: Tetrahydrocannabidiolic Acid

CBDA: Cannabidiolic Acid

THCA: Tetrahydrocannabinolic acid

CBDA: Cannabidiolic acid

CBG: Cannabigerol

CBGA: Cannabigerolic acid

OA: Olivetolic Acid

GPP: Geranyl pyrophosphate

NCBI: National Center for Biotechnology Information

PAL: Phenylalanine ammonia lyase

C4H: Cinnamate 4-hydroxylase

4CL: Cinnamic acid-4-hydroxylase

CHS: Chalcone synthase

CHI: Chalcone isomerase

F3H: Flavanone-3-hydroxylase

F3’H: Flavanone 3'-hydroxylase

F3’5’H: Flavanone 3'5'-hydroxylase

DFR: Dihydroflavonol 4-reductase

ANS: Anthocyanidin synthetase

VOC: Volatile-organic-compounds

QTL: quantitative trait loci

MBW: MYB-bHLH-WDR

UVB: Ultraviolet-B

EBGs: Early biosynthesis anthocyanin genes

LBGs: Late anthocyanin biosynthesis genes

OP: Open-pollinated

UGT: Uridine diphosphate-dependent glucosyltransferase

OMT: O-methyltransferase

ROS: Reactive oxygen species

GBS: Genotyping-by-Sequencing

TE: Transposable Element

JA: Jasmonic Acid

SA: Salicylic Acid

ETH: Ethylene

CK: Cytokinin

GL3: GLABRA3

ANL2: ANTHOCYANINLESS2

TF: Transcription Factor

MLM: Mixed Linear Model

FarmCPU: Fixed and random model Circulating Probability Unification

QQ: Quantile-Quantile

NEB: New England Biolabs

BAM: Binary Alignment Map

PVE: Phenotypic variation explained

UTR: Untranslated region

GUS: β-Glucuronidase

GA*A*NTRY: Gene Assembly in Agrobacterium by Nucleic acid Transfer using Recombinase technologY

ddPCR: digital droplet PCR

CDS: Coding Sequence

DTT: Dithiothreitol

ATP: Adenosine Triphosphate

LB: Luria Broth

OD: Optical Density

Chapter 1: Literature Review

Origins and Historic Use of Cannabis

Cannabis sativa is among the oldest and most versatile crops utilized by humans. Thought to be native to the Tibetan Plateau, Northern China, Eastern China, or the Hindu Kush mountains, some recent evidence has suggested a divergence from the *Parasponia* clade approximately 53 MYA (Rull, 2022; Smart, 2015; Clark and Merlin, 2013; Jin, 2019; McPartland et al., 2023). Divergence from its closest relative species *Humulus lupus* occurred approximately 25 MYA, possibly corresponding with the rise of the Tibetan Plateau (Jin, 2019). Cannabis has long been used as a source of fiber, grain, and medicines in cultures around the world, and remains in use in these capacities today (Clark and Merlin, 2013; Rull, 2022). Dispersal and subsequent speciation within the *Cannabaceae* family may have been caused by local adaptation to various geographic and environmental parameters, e.g. colder, more light-exposed climates of the proposed northern Chinese center of origin (Jin et al., 2023; Pate, 1987).

The high capability for migration resulting from its large pollen dispersal radius has allowed for many regions to support historic *Cannabis* populations, particularly throughout Eurasia (Rull, 2022; Clark and Merlin, 2013). Indeed, Canadian regulations for hemp growers require a minimum isolation distance of at least 5km to prevent cross-contamination of fields, and some studies have even proposed contamination of Spanish cannabis fields via pollen traveling across 90km of the Mediterranean Ocean from Morocco (Small and Antle, 2003; Cabezudo et al., 1997). The multipurpose utility of cannabis for early human

civilizations and its ease of cultivation has allowed it to serve as a pioneer crop in the history of human adoption of agriculture (Clark and Merlin, 2013). Its spread across Eurasia prior to the appearance of humans may have led to multiple independent domestication events in China and Europe. Some have suggested that this spread and multilocal model of domestication may correspond to subspecies differentiation between *C. sativa* subsp. *sativa*, *C. sativa* subsp. *indica*, respectively, with *C. sativa* subsp. *ruderalis* remaining as a wild ancestor (Rull, 2022).

However, taxonomic delineation of *Cannabis* is controversial, and debate has occurred regarding proposed classification schemes, though many accept it as a monospecific genus with a single species *Cannabis sativa*, and two subspecies *C. sativa* subsp. *sativa*, and *C. sativa* subsp. *indica* (McPartland et al., 2023; Rull, 2022). More practically, *Cannabis* is often referred to by biotypes based on end-use for fiber, grain, or high-cannabinoid cultivation (Small, 2015; Rull, 2022). Intensive historic artificial selection and the world-wide distribution of *Cannabaceae* have historically given rise to varied product-based, taxonomic, and legal classifications, leading to complex schemes of nomenclature for individual *Cannabis* plants. Various recommendations for classification of *Cannabis* as either hemp or marijuana based off cannabinoid content, with subspecies corresponding to narcotic effect and domestication status (Small, 2015) or as a single species with three subspecies*, C. sativa* subsp. *sativa*, *C. sativa* subsp. *indica*, and *C. sativa* subsp. *ruderalis* (Rull, 2022) have been given.

The differentiation of supposed *Cannabis* subspecies remains under strong debate, with various studies finding substantial genetic variation between geographic populations,

proposed subspecies, and between cultivars based on usage pattern (fiber, grain, high cannabinoid) (Ren et al., 2021, Balant et al., 2022). Most recent studies using molecular markers to infer phylogenetic relationships have suggested somewhat distinct regional groups in Europe, Northern China, Eastern China, and in Iran (Ren et al., 2021). Though true phylogenetic and morphological differences based on region or biotype may exist, there is inconclusive evidence to confidently classify *Cannabis* past the species level.

Believed to be among the first human cultivated crop species, molecular markers and fossil evidence suggests the emergence of fiber and high-cannabinoid cannabis biotypes in the early Neolithic Era (12KYA) (McPartland et al., 2019). Despite the historic interest in classification and recent push to include genomic data in these efforts there is limited support for multiple or single center of origin hypotheses for domestication, leaving the true origin of *Cannabis* domestication unknown (McPartland et al., 2018; Rull, 2022). Though the center of domestication is still unclear, the impact of cannabis products is not. Grain, fiber, and high-cannabinoid biotypes have been used by cultures throughout the world to create sails, clothes, medicines, foods, ropes, and many other essential goods for thousands of years (Clark and Merlin, 2013).

The initial spread of *Cannabis* across Eurasia is believed to have occurred in the paleolithic era by nomadic peoples migrating westward from Northern China, and using it primarily as a drug (Rull, 2021). Indeed, narcotic use of cannabis is ancient and fossil and historical evidence suggests that it may have been used in religious practices as long ago as 4700 YA in Western China (Russo et al., 2008). Usage of cannabis fiber for ropes, clothes, and sails has been the most widespread and perhaps historically impactful of all, possibly

beginning approximately 8500 YA as funerary shrouds and clothes in China (Schultes, 1970; Mukherjee et al., 2008; Booth, 2005). Intensive European cultivation has been proposed to have begun during the copper and bronze ages (McPartland et al., 2018). Continually increasing in popularity well into the establishment and expansion of the roman empire, hemp fibers dating to this era have been found throughout the roman frontier regions such as Israel and the Iberian Peninsula (Murphy et al., 2011; Rull, 2022). Cannabis fiber cultivation reached its height in Europe during the age of exploration and was introduced to the Americas during this era as well, where it was intensively cultivated to support maritime industry (Clark and Merlin, 2013; Husbands, 1909).

Stigma and Regulation

Despite its historic utility, significant stigma surrounding the use and cultivation of cannabis is prevalent across the world, and many countries have banned its use over the last century. In the United States the first governmental rebuke of cannabis began in 1922 when it was labeled as a narcotic or potentially harmful (Mikos and Kam, 2019). This was followed by the Marihuana Tax Act of 1937 which placed taxes on the cultivation, sale, importation, and possession of the crop and led to a significant decrease of cannabis production in all US states. In 1970 the Controlled Substances Act and the Comprehensive Drug Abuse Prevention and Control Act created the modern 5-tiered schedule system for psychoactive compounds, ranging from Schedule I to Schedule V. Schedule I drugs are described as having little or no human benefit as well as high potential for addiction in users, and cannabis was initially placed in this category at the creation of the bill. This was a highly controversial categorization, and much debate has occurred over the safety and utility of

cannabis in the US and abroad. In 1996, California became the first state to legalize cannabis use and cultivation. In 2014, the Agricultural Act was passed, providing the modern legal framework for *Cannabis* classification into hemp (plant material containing <0.3% THC dry w/w) and marijuana (plant material with <0.3% THC dry w/w), and allowing the cultivation of both for either commercial use or for research purposes. This legal definition is bafflingly based on the singular taxonomic delineation of *Cannabis* proposed by Cronquist and Small (1976) for the supposed differentiation of *C. sativa* from *C. indica*. After the passage of the 2018 Farm Bill, hemp was reclassified as an agricultural commodity, and CBD was clarified as separate from THC. This catapulted the hemp-derived CBD industry into the forefront, and hemp-derived CBD and THC products can now be easily found in most major grocery stores across the US.

Overall, *Cannabis* is a useful and versatile emerging crop species with potential applications in medicine, nutrition, and material science. Though it has been cultivated world-wide throughout history, it still faces many challenges to become aligned with modern standards for distinct, stable, and uniform cultivars. Significant hurdles remain in the way of developing high performing cultivars such as a lack of homozygous parental lines, incomplete genomic resources, and vague regulations making legal status at times unclear. Despite these challenges, public interest in cannabis remains strong, and it is unlikely that this will change soon.

Basic Biology of Cannabis

Cannabis sativa is herbaceous, wind pollinated, largely dioecious, and highly genetically variable both within and between varieties (Pisupati et al., 2018; Soler et al.,

2017; Ren et al., 2021). Sex determination in cannabis is heavily influenced by heteromorphic XY sex chromosomes with female individuals being homogametic (XX) and males being heterogametic (XY) (Prentout et al., 2020). Monoecy is present in some varieties, and monoecious individuals carry homomorphic sex chromosomes (XX) and bisexual inflorescences, indicating that autosomes contain important genes sufficient to induce male floral sexual expression (Faux et al., 2014). Many commercial cannabis cultivars are highly variable in several key agronomic traits, namely flowering time, germination rate, secondary metabolite profile, disease resistance, and some domestication traits such as seed shattering (Stack et al., 2020; Suchoff et al., 2024). Such high variability is likely a result of the millennia of strong artificial selection in multiple different environments by different cultures across Eurasia and more recent feralization and losses in genetic variation from prohibitive regulations and clandestine breeding efforts.

Sex expression in *C. sativa* is highly plastic and, despite its determinative sex chromosomes, can be altered by environmental stimuli and exogenous chemical application (Baek and Vergara, 2025). Sexual dimorphism is present in cannabis, with males often being significantly thinner stemmed, flowering earlier, supporting unique floral structures, and having comparatively low levels of phytocannabinoids relative to female plants (Moliterni et al., 2004). By contrast, female plants have much wider stem diameter and often express some or all the morphological traits usually associated with biotype, e.g. high trichome density in high-cannabinoid types and long stalked growth patterns in fiber types. Monoecious plants often contain a blend of agronomic traits characteristic of fiber or

grain biotypes such as an elongated growth pattern and dense grain-filling floral heads (Baldini et al., 2018).

Cannabis is commonly photoperiod sensitive, requiring a 12/12-hour light/dark regime to induce flowering. Though many varieties flower under short day conditions, variation is also present for photoperiod sensitivity or lack thereof, with differences for flowering time both between and within male and female individuals and populations. Male individuals flower approximately two weeks earlier than females, and timing differences remain consistent in monoecious populations, suggesting an autosomal genetic origin of this trait (Van Baek., 2025; Steel et al., 2023). Indeed, a recent transcriptomic analysis implicated *CsFloweringDate (CsFD)* as a potential causative gene underlying male early flowering (Shi et al., 2025). An ortholog of *AtFD*, a well described regulator of flowering time in *Arabidopsis,* it is conceivable that *CsFD* may act in concert with other genes to regulate the sex specific flowering times of *Cannabis* plants. Photoperiod insensitive plants are colloquially termed “autoflowers” and have distinctive morphology, often growing much faster and remaining smaller during growth. Recent advances in understanding the basis of photoperiod insensitivity have been made, with Toth et al. (2021) observing it as a simple recessive trait and discovering a QTL on chromosome 1 containing the putative causal locus *autoflower1*.

Interestingly, molecular markers developed from this QTL were insufficient to correctly predict photoperiod insensitivity across all varieties, suggesting independent mechanisms for the evolution of this trait. Flowering time is also partially genetically controlled, and in the same study it was found that the QTL *early1* conferred approximately

2-4 weeks earlier flowering time. From this QTL, an ortholog of the major Rice flowering gene *Heading date 16* was discovered, suggesting a potentially multilocus control mechanism for flowering date in cannabis. It is also possible that the different alleles uncovered for flowering time may reflect varietal differences present between the different cultivars used in each study.

Cannabis Breeding and Biotechnology

Cannabis has been under extensive artificial selection for thousands of years, but more recent prohibitive regulations have led to both decreases in genetic variance in the case of clandestine breeding and increases in genetic variance following feralization and local adaptation (Aina et al., 2025). In addition, the high levels of heterozygosity, characteristic of a wind-pollinated crop, makes the majority of commercially available Cannabis cultivars highly unstable and non-uniform. The most prominent of recent breeding efforts has been the development and validation of high-quality reference genomes, the advent of hemp varieties containing predominantly non-intoxicating Cannabidiol (CBD) or type III plants, and genetic mapping of the cannabinoid biosynthesis pathway (McPartland et al., 2017; Luo et al., 2019; Grassa et al., 2021). More recent work has focused on utilizing transcriptomic and genetic analysis to begin to tease apart the physiological complexity of secondary metabolite biosynthesis (namely terpenoids and cannabinoids) and to uncover the basis of various agronomic and morphological traits (Fulvio et al., 2025; de Ronne et al., 2024; Trubanov et al., 2025).

Stable transformation methods capable of producing T_1 transgenic events in Cannabis have until now been unavailable due to its uncommon, and varietally dependent

recalcitrance to regeneration (Galan-Avila et al., 2021). Nanoparticle- and agrobacterium-mediated methods for generation of stable T_0 lines exist and have been used in functional characterization of candidate genes affecting inflorescence development and grain production (Ahmed et al., 2021; Galan-Avila et al., 2021; Xu et al., 2024). Recent advances in Cannabis transformation at the Wisconsin Crop Innovation Center (WCIC) have enabled the generation of stable T_1 events and have facilitated the release of the total cannabinoid knockout line BadgerG.

Genomic resources to aid in cannabis research are relatively new, with the first published cannabis assembly 'Purple Kush' being released in 2011 with a reported genome size of 534MB (Van Bakel et al., 2011). Following the release and subsequent improvement of the Purple Kush reference genome in 2018, the cs10 reference genome was released in 2021, reporting a genome size of 736MB, and was widely lauded as the most complete and accurate of available cannabis genomes upon its release (Laverty et al., 2018; Grassa et al., 2021). However, some issues still remained following these releases: it was indicated that the cs10 reference genome may be physically disordered, the limited scope of varieties sequenced may not have reflected the extensive genetic variability thought to be present both within and between subpopulations, and disagreement over the total genome size and repetitive DNA content remained (Kovalchuk et al., 2020; Faux et al., 2014; Soler et al., 2017; Gao et al., 2020). In addition to these larger scale genomic issues, there is disagreement between candidate gene discovery studies on the naming conventions of annotated genes; see Bassolino et al., 2023; Kundan et al., 2022. This was somewhat ameliorated by the release of the 'Pink Pepper' reference genome, a high-quality and haplotype-resolved

assembly, which was unlikely to be physically disordered due to the long read sequencing technology used in its construction (Ryu et al., 2024).

Recently, a cannabis pangenome has been released, comprising 181 new and 12 previously released assemblies and including expression data and sequence for both sex chromosomes (Lynch et al., 2025). Significant variation among lines was observed, and it was found that transposable elements (TEs) and repetitive sequences comprise on average 68% of the genome, contrasting with previous reports of 60% and 74.5%, respectively (Pisupati et al., 2018; Gao et al., 2020). The pangenome will be instrumental for breeding programs to develop cross-variety-validated markers capable of accurately detecting traits in diverse germplasm and will be a huge boon for future genetic mapping studies as a way of circumventing the high reference bias arising from strong intra-varietal variation (Stack et al., 2025).

Cannabis is most famous for the cannabinoids present within the glandular trichome heads of its female flowers (Livingston et al., 2020). Cannabinoid biosynthesis is primarily mediated by two enzymes: Tetrahydrocannabinolic Acid Synthase (THCAS) and Cannabidiolic Acid Synthase (CBDAS), which act in competition with one another to catalyze the formation of Tetrahydrocannabinolic Acid (THCA) or Cannabidiolic Acid (CBDA) from Cannabigerolic Acid (CBGA) (De Meijer et al., 2003). Since the initial domestication event(s) of *Cannabis*, selection for cannabinoid content has likely been ongoing, but in recent decades selection pressure appears to have increased dramatically. From 1995-2014, the average percentage of THC within samples confiscated by the DEA increased from 4% to 12% (El Sohly et al., 2016). On Leafly.com, one of the main websites detailing the strains of

recreational cannabis on the market and in dispensaries, there are now 2427 reported strains containing containing >20% THC, reflecting a continuing trend of increases in THC content within available products. This is coupled with findings that high-CBD varieties are primarily individuals from high-THC backgrounds into whom the B^D/B^D (non-functional THCAS alleles) have been introgressed, possibly explaining the rapid development of high yielding CBD varieties following the 2014 and 2018 farm bills (Grassa et al., 2021).

Notable Secondary Metabolites

Cannabis can present many phytochemicals within their various tissues, producing more than 588 unique secondary metabolites, including 144 cannabinoids (Yadav et al., 2023; Pourseyed Lazarjani et al., 2020). This vast chemical diversity within plant tissues situates cannabis as a crop with strong potential as a biofactory for the rapid production of medicinal compounds (Tarkowski et al., 2025). Accompanying the titular cannabinoids are a vast number of terpenes, alkaloids, hydrocarbons, volatile sulfur compounds, and flavonoids. While cannabinoids are often the most abundant and controversial chemicals, other chemical constituents can help to maintain plant health, primarily through increasing environmental resiliency (Naing et al., 2021; Jin et al., 2020). Other, less abundant phytochemicals can also have strong effects on human sensory perception of cannabis and in particular the terpenes and volatile sulfur compounds are thought to contribute to its distinctive floral scent (Sommanno et al., 2020; Paryani et al., 2024). The array of non-cannabinoid phytochemicals has been suggested to contribute to a phenomenon coined the “entourage effect” in which various chemical constituents are theorized to interact with each other, producing altered experiential and medical outcomes (Andre et al., 2024; Finlay

et al., 2020; Worth, 2019). While evidence is suggestive of a synergistic effect between terpenes and cannabinoids in medical applications, the “entourage effect” remains controversial and unproven, and leveraging this putative effect is still far from reality (Finlay et al., 2020). Cannabis secondary metabolism research has been heavily focused on cannabinoids and occasionally on terpenes, and to date much remains unknown about the medical applications, adaptive value, or degree of variance between genotypes of the secondary metabolites present in leaves, flowers, and stems.

THC and CBD biosynthesis has been well studied and was thought to operate under the control of a single locus (B) mendelian model with two alleles B^T (encoding THCAS) and B^D (encoding CBDAS) responsible for THCA and CBDA production, respectively (De Meijer et al., 2003). These alleles segregate in a classical 1:2:1 mendelian ratio corresponding to high THC in B^T/B^T, an intermediate chemical split in B^T/B^D, and high CBD in B^D/B^D genotypes (De Meijer et al., 2003; Stack et al., 2021). Upstream from this, the O allele (encoding CBGAS) catalyzes the conversion of Geranyl diphosphate (GPP) and Olivetolic acid (OA) into CBGA, a precursor to both THCA and CBDA (De Meijer et al., 2003). More recent genetic mapping efforts have localized the THCA/CBDA synthase region to chromosome 7, where the genes were confirmed to be mutually exclusive but inextricably linked in repulsion (Grassa et al., 2021). The CBDAS enzyme is notoriously promiscuous, producing up to 3-5% THC even in the absence of THCAS (Luo et al., 2021). This can lead to significant risk for growers, who may be forced to destroy their harvest if their plants test higher than the federal limit of 0.3% THC.

Chemistry and Biosynthesis of Anthocyanins

Anthocyanins are a subclass of flavonoids characterized by their versatile roles in homeostatic and adaptive processes in both vegetative and reproductive tissue (LaFountain and Yuan, 2021). They are most known for their roles as the vibrant pink, purple, or red pigments in reproductive tissues of most flowering plants and absorb light at around 480-560nm (Stintzing et al., 2002). Primarily localized to the vacuole, anthocyanins require an acidic environment to emit visible coloration and can support a wide variety of chemical modifications which can alter hue, color intensity, molecular stability, and antioxidant activity (Cai et al., 2022; Saigo et al., 2020). The chemical diversity of anthocyanins supports many tissue pigmentation changes and underlies their different physiological roles in plants across taxa (Saigo et al., 2020). Cannabis in particular displays striking phenotypic variation for anthocyanin pigmentation intensity and patterning both within and between cultivars (Figure 1.1).

Figure 1.1: Phenotypic variation for anthocyanin accumulation in *C. sativa*. Cannabis plants with exclusive floral (left), leaf (center), and stem (right) pigmentation patterns are shown.

One of the most useful properties of anthocyanins is their antioxidant capability, an important physiological role in both plants and humans. Antioxidants can aid in recovery from UVB-radiation damage or reactive oxygen species (ROS) overaccumulation and oxidative toxicity (Naing et al., 2021). In *Arabidopsis* anthocyanin deficient mutants, hyper-sensitivity to ROS over-accumulation was observed along with corresponding breakdowns in photosynthetic capacity (Xu et al., 2017). ROS buildup has multiple sources from both endogenous stress response signal cascade mechanisms and from environmental damage/natural wear on primary metabolism machinery (Cerqueira et al., 2023). Anthocyanin regulatory elements are induced by ROS regardless of its source, and aid in scavenging their free radicals before they damage DNA, proteins, or lipids. The balance of necessary ROS for cell signaling and antioxidant induction for quenching them is a vital part of oxido-reductive homeostasis, and crucial for abiotic stress response in plants (Singh et al., 2019). The role of anthocyanins in this process is somewhat clear though often their induction is redundant for ROS quenching, and they are at times upregulated independently, suggesting that there remains more to uncover about their physiological functions.

Antioxidants are recognized as being widely beneficial for humans and are lauded for their roles in reducing inflammation, cardioprotection, as anti-cancer agents, and as anti-diabetic treatments (Khoo et al., 2017; Kumar and Pandey, 2013). Uses in clinical and sustainability settings have also been investigated, with anthocyanins showing promise as preventative medicines for some chronic diseases (Wang and Stoner, 2008), as well as useful ingredients for fortification of nutritious foods for better management of Type II Diabetes (Neacsu et al., 2022). Due to the ease of identification and potential medical

appeal, anthocyanins have been reconstructed in *E. coli* for rapid synthesis and targeted addition of specific moieties (Zhang et al., 2014). The near ubiquity of anthocyanins across berries, grains, and vegetables makes them common in most human diets, situating them as an important nutrient with potential for biofortification applications in preventative healthcare.

Compounding on the nutritional and clinical benefits, anthocyanins are of particular importance in cannabis as key drivers of consumer preferences and as a culturally impactful trait in the broader public perception of high-cannabinoid cannabis. For example, on leafly.com, one of the largest records of illicit cannabis strains, there are 225 strains with the keyword “Purple” in their name, and hundreds more with alternative spellings. In addition to the popularity of the trait in naming conventions, it has had a significant influence on the symbology of cannabis in pop culture as can be seen through the popularity of media such as Purple Haze by Jimi Hendrix and in the works of many, many modern rappers. Anthocyanin accumulation in cannabis is an important cultural and visual trait for growers and retailers and creating cultivars high in these compounds can be an effective tool for breeders developing and marketing new varieties.

Color differentiation can play a pivotal role in ecological processes, conferring reproductive advantages to plants displaying high pigmentation in their floral tissue and in some cases aiding in speciation and adaptation to novel environments (Streisfeld et al., 2013). Indeed, it has been proposed that *Cannabacae* speciation into *Humulus* and *Cannabis* may have been spurred through high intensity UVB radiation at high altitudes, resulting in rapid diversification of individuals, though this proposal has been disputed (Pate,

1983). In the case of *Mimulus lewisii* and *M. cardinalis* for example, an anthocyanin QTL affecting floral pigmentation traits and pollinator preferences for one species over the other contributes to reproductive barriers between the two closely related species (Yuan et al., 2013).

The anthocyanin biosynthetic pathway is an offshoot of the wider phenylpropanoid metabolism pathway in plants and is based primarily on eight main enzymatic steps (Figure 2). These steps are often subdivided into early biosynthetic genes (EBGs) and late biosynthetic genes (LBGs). EBGs are offshoots of the Shikimate pathway which produces aromatic amino acids and are upstream to many secondary metabolic pathways. By contrast, LBGs are more specifically dedicated to anthocyanin biosynthesis and act to fine-tune the glycosylation decorations (Saigo et al., 2020). Much of the current knowledge regarding anthocyanin biosynthesis comes from analysis of natural color mutants in the seed coats, vegetative, and floral tissues of crop and model species (Tohge et al., 2017; Saito et al., 2013).

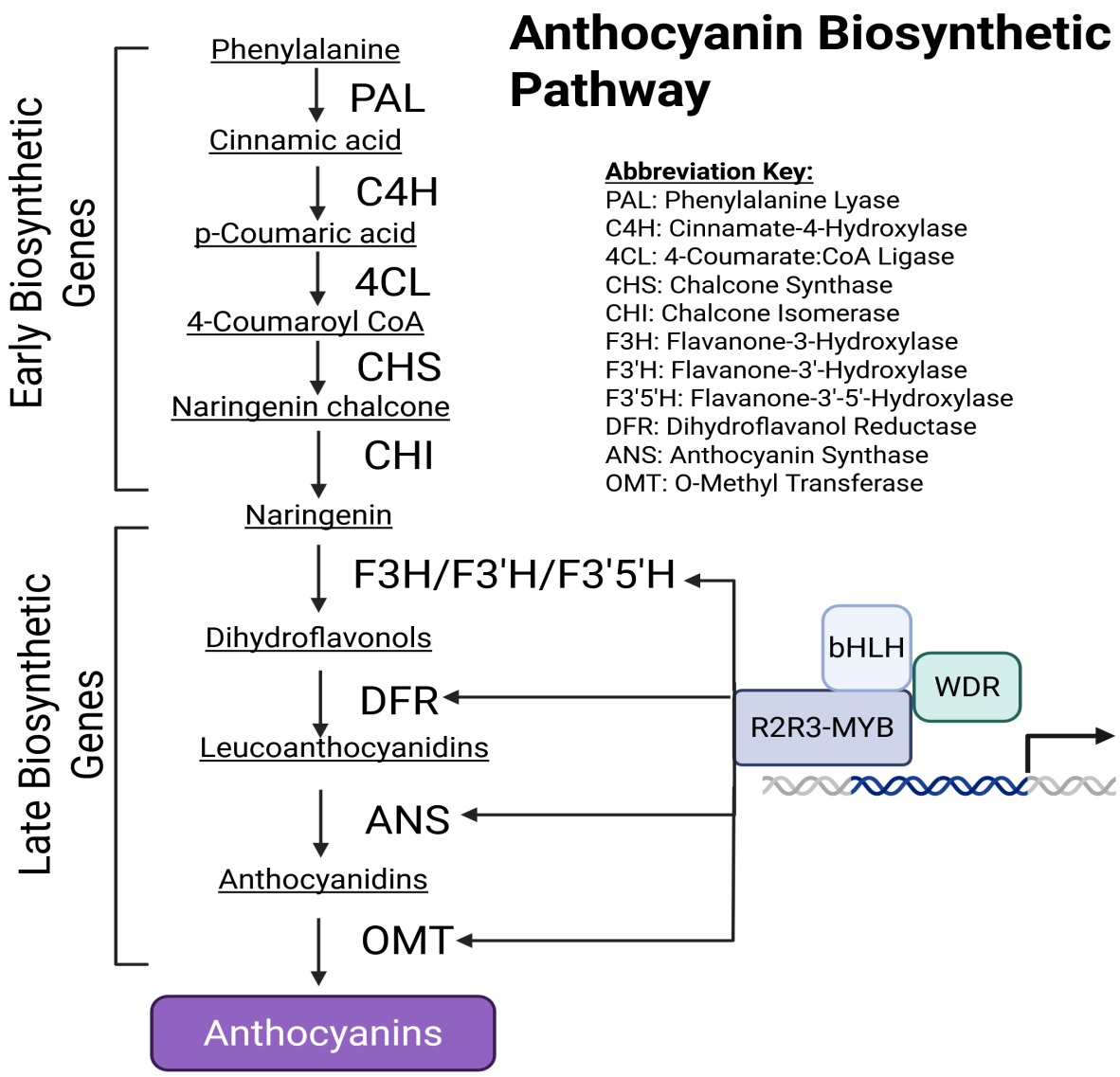

Figure 1.2. *Anthocyanin Biosynthetic Pathway*. The biosynthetic path of anthocyanins begins with the early biosynthetic genes, started by PAL, which catalyzes the transformation of phenylalanine to cinnamic acid, followed by C4H mediated conversion to p-coumaric acid and its subsequent conversion to p-coumaroyl COA by 4CL. Chalcone Synthase (CHS) is the first committed enzyme of the flavonoid pathway and catalyzes the reaction of three molecules of Malonyl COA and p-coumaroyl CoA into the flavanone precursor Naringen Chalcone. This is followed by the isomerization into Naringen via Chalcone Isomerase (CHI). Naringen is a promiscuous precursor molecule for a variety of secondary metabolites. Next is three rounds of hydroxylation from various Flavanone Hydroxylase (F3H/F3'H/F3'5'H) enzymes to form Dihydroflavonol, a direct precursor to anthocyanin. Dihydroflavonol is reduced by Dihydroflavonol Reductase (DFR) then converted to anthocyanin by Anthocyanin Synthase (ANS), the last step common to all anthocyanin chemotypes. Anthocyanins are then glycosylated or otherwise decorated before being transported to the vacuole by a variable suite of methyl or glucosyl transferase enzymes (OMT/UGT) based on the specific identity of the anthocyanin in question.

In cannabis, the EBGs and LBGs have been identified and annotated within many reference genomes (Docimo et al., 2013; Gagalova et al., 2024). Both Gagalova (2024) and Docimo (2013) reported that cannabis biosynthetic enzymes exist in single copy forms, though this may not hold true across extant populations. Keracyanin (Cyanidin-3-rutinoside) has been proposed to be the most abundant anthocyanin (Bassolino et al., 2023), though Gagalova (2024) found that the most common anthocyanins in floral and vegetative tissue were variable by variety, possibly suggesting ongoing intraspecific divergence in anthocyanin glycosylation. Genomic and RNAseq analysis of previously annotated biosynthetic genes has indicated that the varieties Willow-Alpha, CA19210, CK19206, and CaliKush all have single copy versions of the late biosynthetic genes, and that an OMT transferase is the most likely candidate for the final localization of anthocyanins to the acidic vacuole (Gagalova et al., 2024). It is possible that OMT is the most abundant transferase in only these varieties given the chemical diversity of anthocyanins reported. This chemical diversity is perhaps unsurprising in the context of heavy selection pressure by clandestine breeders and phylogenetic disorder of *Cannabis* as a species and may be an indication of rapid divergence in secondary metabolite profiles between varieties and populations.

Genetic Regulation of Anthocyanin Accumulation

The biosynthetic pathway of anthocyanins is a source of rich variation both within and between species and serves as an excellent model system for investigation the interplay of evolution, molecular genetics, and developmental biology. The regulation of anthocyanin biosynthesis is well described in plants, mainly due to the ease of visual identification of alterations to the regulatory network through distinctive phenotypic changes. Furthermore,

anthocyanin biosynthetic processes and many of the same regulatory players are widely conserved across taxa and maintain strong enough sequence similarity such that *Arabidopsis* activators can successfully induce pigmentation in species as far diverged as *Humulus lupus* (Gatica-Arias et al., 2012). MYB proteins are canonical regulators of this pathway and can bind to the promoter regions to up or downregulate transcription of their target gene. They are classified into subgroups by the R1R2R3 repeat moieties and combinations therein, e.g. R3-MYB, R2R3-MYB, or 1R-MYB. In brief, 1R-MYBs (single repeat or MYB-like) mediate protein-protein interactions and R2R3-MYBs are *cis*-regulatory elements. R3-MYBs are more cryptic, though some reports suggest that they act as competitive inhibitors of R2R3-MYBs in developmental patterning processes (Zheng et al., 2021).

The so-called MBW complex, comprised of a MYB transcription factor, a bHLH cofactor, and a WDR scaffold, is well established as the main canonical regulator of anthocyanin accumulation. In *Arabidopsis*, it was found that the LBGs and EBGs (to a lesser extent) were highly upregulated by an MBW complex consisting of *PAP1*/*MYB75*, *TTG1* (*bHLH*), and *WD40*, with overexpression of this suite sufficient to induce anthocyanin pigmentation (Gonzalez et al., 2008). There are, however, MYBs which can repress anthocyanin biosynthesis either through *cis*-regulatory repression, or through competition with capable MYBs for MBW cofactors (LaFountain and Yuan, 2023). This interplay of activation and repression, sometimes by the same protein, can result in intriguing patterns of trait expression across tissues.

MYB transcription factors are numerous and subfamilies of MYBs are involved in cell fate determination, environmental response, secondary metabolite production, and are key in mediated protein-protein interactions (Bhatt et al., 2025). MYBs are often tissue specific, though other MBW constituents can be more general. For example, in *C. sativa*, *CsbHLH112* and *CsbHLH114* are putatively involved in MBW complexes regulating both trichome secondary metabolism and anthocyanin accumulation, though this has yet to be confirmed through functional analysis (Bassolino et al., 2023; Fulvio et al., 2025). MBW complexes can also be controlled through autoregulatory feedback mechanisms which can have stunning effects on pigmentation patterning in plant tissues (Zheng et al., 2021).

Investigation into the molecular mechanisms of MYB-mediated environmental response has revealed a complex web of crosstalk between environmental stimuli and several plant hormones, the chief of which being Jasmonic Acid (JA), Salicylic Acid (SA), and Ethylene (ETH). JA, SA, and ETH are all canonical stress response hormones and are vital parts of signal transduction cascades to coordinate physiological responses. Interestingly, sucrose and phosphorus (or lack thereof) have also been shown to be potent induction agents for anthocyanin accumulation in *Arabidopsis* (Li et al., 2014; Li et al., 2023). Sucrose-mediated anthocyanin induction also requires the presence of *AtMYB75/PAP1* and may not be conserved in species lacking this gene (Teng et al., 2005). MYBs and MBW complexes are crucial for proper stress response in plants, and the activation of anthocyanin biosynthesis may be an artifact arising from MBW complexes playing other roles in environmental response.

In cannabis, MYB family proteins have been catalogued and annotated independently several times, leading to much confusion over correct nomenclature (Bassolino et al., 2021; Kundan et al., 2022; Yin et al., 2022). Despite underrepresentation among cannabis research, MYBs appear to maintain their roles in mitigating environmental stress, inducing secondary metabolite production, and initiating tissue specification. Overexpression of *CsMIXTA*, an R2R3-MYB, in *N. benthamiana* leaves was sufficient to double the observed amount of trichomes (Haiden et al., 2022). Further, cannabis MYBs and proanthocyanidins, putatively MYB-mediated phytochemicals, have been shown to be effective in reducing heavy metal toxicity (Yin et al., 2022). Prior work using transcriptomic analysis also found that higher levels of anthocyanin accumulation in cannabis was accompanied by increased transcript abundance of *CsMYB78* and *CsMYB33*, both putative orthologs of *AtMYB113*, a canonical activator of anthocyanin biosynthesis. Interestingly, *CsMYB78* does not induce heavy anthocyanin pigmentation in *N. benthamiana* leaves while *CsMYB33* does. Overall, there remains significant work to be done regarding the MYB family of transcription factors in cannabis, and future research will help to uncover their roles in tissue specification, pigmentation, and environmental response.

Aims and Rationale

Aesthetic qualities can strongly drive consumer preference for plant varieties, particularly in specialty crops, and the purple serrated leaflet of cannabis is not only an iconic symbol of the culture surrounding the crop, but a valuable aesthetic trait for differentiating and marketing varieties. Previous work has focused on uncovering the chemical composition (Cerrato et al., 2021), biosynthetic genes (Docimo et al., 2013;

Gagalova et al., 2024), and regulatory elements (Bassolino et al., 2020; Kundan et al., 2022) affecting anthocyanin accumulation in cannabis, though few functional analyses have been performed to characterize the many putative candidate genes uncovered. In this study, we first validated and added to previous mapping studies for anthocyanin accumulation in leaves, flowers, and stems using an F_1 mapping population, uncovering large significant marker association clusters on chromosomes 1 and 6. Of these, the marker cluster on chromosome 6 explaining approximately 67% of the phenotypic variation for flower anthocyanin contained a strong candidate gene: *CsMYB78*, an ortholog of *AtMYB1*, an environmentally inducible activator of anthocyanin biosynthesis in several crop species (Schwinn et al., 2016; An et al., 2018). PCR-based fine mapping in concert with visualizations from recent bioinformatic resources revealed significant genotypic variation surrounding this locus. Further RNA sequencing uncovered significant sequence variation between green and purple genotypes allowing for SNP discovery from consensus sequences of either genotype. Prior to obtaining full length sequence data for candidate genes, overexpression experiments using candidate gene sequences of *CsMYB78* and *CsMYB33*, a putative ortholog of *AtMYB113,* from Kundan et al. (2022) were performed. Interestingly, *CsMYB78* was unable to induce visible anthocyanin pigmentation in transformants while *CsMYB33* induced heavy pigmentation. The absence of *CsMYB33* in GWAS results, the sequence variability surrounding *CsMYB78*, and the results of RNA sequencing suggest that the anthocyanin regulatory network in cannabis may be uniquely complex and could hint at genotype specific changes in regulatory players. It is possible that decades of intense clandestine selection for purple leaves and flowers along with the natural variation of a

highly heterozygous crop cultivated in nearly every part of the world has led to intra-varietal differences between regulatory mechanisms of secondary metabolite pathways.

References

Ahmed, S., Gao, X., Jahan, M. A., Adams, M., Wu, N., & Kovinich, N. (2021). Nanoparticle-based genetic transformation of *Cannabis sativa*. *Journal of Biotechnology*, *326*, 48–51. https://doi.org/10.1016/j.jbiotec.2020.12.014

André, R., Gomes, A. P., Pereira-Leite, C., Marques-da-Costa, A., Monteiro Rodrigues, L., Sassano, M., Rijo, P., & Costa, M. do C. (2024). The entourage effect in cannabis medicinal products: A comprehensive review. *Pharmaceuticals*, *17*(11). https://doi.org/10.3390/ph17111543

Baek, Y., & Vergara, D. (2025). A review of sexual strategies in *Cannabis sativa* L. under genomic and environmental controls. *Agrosystems, Geosciences & Environment*, *8*(1), e70050. https://doi.org/10.1002/agg2.70050

Baldini, M., Ferfuia, C., Piani, B., Sepulcri, A., Dorigo, G., Zuliani, F., Danuso, F., & Cattivello, C. (2018). The performance and potentiality of monoecious hemp (*Cannabis sativa* L.) cultivars as a multipurpose crop. *Agronomy*, *8*(9). https://doi.org/10.3390/agronomy8090162

Balant, M., Rodríguez González, R., Garcia, S., Garnatje, T., Pellicer, J., Vallès, J., Vitales, D., & Hidalgo, O. (2022). Novel insights into the nature of intraspecific genome size diversity in *Cannabis sativa* L. *Plants*, *11*(20). https://doi.org/10.3390/plants11202736

Bassolino, L., Fulvio, F., Pastore, C., Pasini, F., Gallina Toschi, T., Filippetti, I., & Paris, R. (2023). When *Cannabis sativa* L. turns purple: Biosynthesis and accumulation of anthocyanins. *Antioxidants*, *12*(7). https://doi.org/10.3390/antiox12071393

Bassolino, L., Buti, M., Fulvio, F., Pennesi, A., Mandolino, G., Milc, J., Francia, E., & Paris, R. (2020). In silico identification of MYB and bHLH families reveals candidate transcription factors for secondary metabolic pathways in *Cannabis sativa* L. *Plants*, *9*(11). https://doi.org/10.3390/plants9111540

Bassolino, L., Fulvio, F., Cerrato, A., Citti, C., Cannazza, G., Capriotti, A. L., Alberti, I., Terracciano, I., Pecchioni, N., & Paris, R. (2024). Metabolic characterization and transcriptional profiling of polyphenols in *Cannabis sativa* L. inflorescences with different chemical phenotypes. *Planta*, *260*(4), 76. https://doi.org/10.1007/s00425-024-04505-z

Bhatt, V., Sharma, A., & Chauhan, R. (2025). MYB transcription factors: Regulation of secondary metabolism and abiotic stress tolerance in plants. *Plant Gene*, *33*, 100457. https://doi.org/10.1016/j.plgene.2023.100457

Booth, M. (2005). *Cannabis: A history*. St. Martin's Press.

Cabezudo, B., Recio, M., Sánchez-Laulhé, J., Trigo, M. M., Toro, F. J., & Polvorinos, F. (1997). Atmospheric transportation of marijuana pollen from North Africa to the southwest of Europe. *Atmospheric Environment*, *31*(20), 3323–3328. https://doi.org/10.1016/S1352-2310(97)00161-1

Cai, D., Li, X., Chen, J., Jiang, X., Ma, X., Sun, J., Tian, L., Vidyarthi, S. K., Xu, J., Pan, Z., & Bai, W. (2022). A comprehensive review on innovative and advanced stabilization

approaches of anthocyanin by modifying structure and controlling environmental factors. *Food Chemistry*, *366*, 130611. https://doi.org/10.1016/j.foodchem.2021.130611

Cerqueira, J. V. A., de Andrade, M. T., Rafael, D. D., Zhu, F., Martins, S. V. C., Nunes-Nesi, A., Benedito, V., Fernie, A. R., & Zsögön, A. (2023). Anthocyanins and reactive oxygen species: A team of rivals regulating plant development? *Plant Molecular Biology*, *112*(4), 213–223. https://doi.org/10.1007/s11103-023-01362-4

Clark, R. C., & Merlin, M. D. (2013). *Cannabis: Evolution and ethnobotany*. University of California Press.

Cronquist, A., & Small, E. (1976). A practical and natural taxonomy for *Cannabis*. *Taxon*, *25*(4), 405–435. https://doi.org/10.2307/1220524

De Meijer, E. P. M., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. M. C., Ranalli, P., & Mandolino, G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*, *163*(1), 335–346. https://doi.org/10.1093/genetics/163.1.335

Docimo, T., Consonni, R., Coraggio, I., & Mattana, M. (2013). Early phenylpropanoid biosynthetic steps in *Cannabis sativa*: Link between genes and metabolites. *International Journal of Molecular Sciences*, *14*(7), 13626–13644. https://doi.org/10.3390/ijms140713626

El Sohly, M. A., Mehmedic, Z., Foster, S., Gon, C., Chandra, S., & Church, J. C. (2016). Changes in cannabis potency over the last 2 decades (1995–2014): Analysis of current data in the United States. *Biological Psychiatry*, *79*(7), 613–619. https://doi.org/10.1016/j.biopsych.2016.01.004

Finlay, D. B., Sircombe, K. J., Nimick, M., Jones, C., & Glass, M. (2020). Terpenoids from cannabis do not mediate an entourage effect by acting at cannabinoid receptors. *Frontiers in Pharmacology*, *11*, 359. https://doi.org/10.3389/fphar.2020.00359

Faux, A.-M., Berhin, A., Dauguet, N., & Bertin, P. (2014). Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). *Euphytica*, *196*(2), 183–197. https://doi.org/10.1007/s10681-013-1023-y

Fulvio, F., Pieracci, Y., Ascrizzi, R., Bassolino, L., Flamini, G., & Paris, R. (2025). Insights into terpenes profiling and transcriptional analyses during flowering of different *Cannabis sativa* L. chemotypes. *Phytochemistry*, *229*, 114294. https://doi.org/10.1016/j.phytochem.2024.114294

Gagalova, K. K., Yan, Y., Wang, S., Matzat, T., Castellarin, S. D., Birol, I., Edwards, D., & Schuetz, M. (2024). Leaf pigmentation in: Characterization of anthocyanin biosynthesis in colorful cannabis varieties. *Plant Direct*, *8*(11), e70016. https://doi.org/10.1002/pld3.70016

Galan-Ávila, A., Gramazio, P., Ron, M., Prohens, J., & Herraiz, F. J. (2021). A novel and rapid method for *Agrobacterium*-mediated production of stably transformed *Cannabis sativa* L. plants. *Industrial Crops and Products*, *170*, 113691. https://doi.org/10.1016/j.indcrop.2021.113691

Gao, S., Wang, B., Xie, S., Xu, X., Zhang, J., Pei, L., Yu, Y., Yang, W., & Zhang, Y. (2020). A high-quality reference genome of wild *Cannabis sativa*. *Horticulture Research*, *7*(1), 73. https://doi.org/10.1038/s41438-020-0295-3

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *The Plant Journal*, *53*(5), 814–827. https://doi.org/10.1111/j.1365-313X.2007.03373.x

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., & Schwartz, C. J. (2021). A new cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist*, *230*(4), 1665–1679. https://doi.org/10.1111/nph.17243

Haiden, S. R., Apicella, P. V., Ma, Y., & Berkowitz, G. A. (2022). Overexpression of CsMIXTA, a transcription factor from *Cannabis sativa*, increases glandular trichome density in tobacco leaves. *Plants*, *11*(11). https://doi.org/10.3390/plants11111519

Husbands, J. D. (1909). *Seeds and plants imported during the period from October 1 to December 31, 1908*. (Inventory No. 17). USDA Bureau of Plant Industry Bulletin.

Jin, D. (2019). *Cannabis sativa*: The plant of a thousand and one molecules. In M. A. ElSohly (Ed.), *Cannabis sativa L. – Botany and biotechnology* (pp. 3–23). Springer.

Jin, J.-J., Yang, M.-Q., Fritsch, P. W., van Velzen, R., Li, D.-Z., & Yi, T.-S. (2020). Born migrators: Historical biogeography of the cosmopolitan family Cannabaceae. *Journal of Systematics and Evolution*, *58*(4), 461–473. https://doi.org/10.1111/jse.12552

Jin, J.-J., Yang, M.-Q., Fritsch, P. W., van Velzen, R., Li, D.-Z., & Yi, T.-S. (2023). Recent divergence of *Cannabis* within Cannabaceae. *Systematic Botany*, *48*(1), 23–35.

Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., Ashnest, J. R., Mau, M., Schröder, J., Rieseberg, L. H., & Mauleon, R. (2020). The genomic view of cannabis

domestication. *Plant Biotechnology Journal*, *18*(5), 1102–1110. https://doi.org/10.1111/pbi.13246

Kundan, M., Gani, U., Fayaz, M., Angmo, T., Kesari, R., Rahul, V. P., Gairola, S., & Misra, P. (2022). Two R2R3-MYB transcription factors, CsMYB33 and CsMYB78, are involved in the regulation of anthocyanin biosynthesis in *Cannabis sativa* L. *Industrial Crops and Products*, *188*, 115546. https://doi.org/10.1016/j.indcrop.2022.115546

Kumar, S., & Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: An overview. *The Scientific World Journal*, 2013, Article ID 162750. https://doi.org/10.1155/2013/162750

LaFountain, A. M., & Yuan, Y.-W. (2021). Repressors of anthocyanin biosynthesis. *New Phytologist*, *231*(3), 933–949. https://doi.org/10.1111/nph.17402

LaFountain, A. M., & Yuan, Y.-W. (2023). Anthocyanin pathway gene regulation in model and crop species. *Plant Science*, *329*, 111571. https://doi.org/10.1016/j.plantsci.2023.111571

Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., Deikus, G., Sebra, R., Hughes, T. R., Page, J. E., & van Bakel, H. (2018). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangement at the THC/CBD acid synthase locus. *Genome Research*, *28*(10), 1462–1477. https://doi.org/10.1101/gr.242594.118

Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D. C. J., Rensing, K. H., Laflamme-Yonkman, J., Castellarin, S. D., Bohlmann, J., Page, J. E., & Samuels, A. L. (2020). Cannabis glandular trichomes alter morphology and metabolite content during

flower maturation. *The Plant Journal*, *101*(1), 37–56. https://doi.org/10.1111/tpj.14516

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., Zhang, Y., Grzybowski, A. T., Harth, S., Lin, W., Lee, H., Yu, C., Shin, J., Deng, K., Benites, V. T., Wang, G., Baidoo, E. E. K., Chen, Y., Dev, I., … Keasling, J. D. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*, *567*(7746), 123–126. https://doi.org/10.1038/s41586-019-0978-9

McPartland, J. M., Guy, G. W., & Hegman, W. (2018). Cannabis is indigenous to Europe and cultivation began during the Copper or Bronze Age: A probabilistic synthesis of fossil pollen studies. *Vegetation History and Archaeobotany*, *27*(4), 635–648. https://doi.org/10.1007/s00334-018-0678-7

McPartland, J. M., Hegman, W., & Long, T. (2019). Cannabis in Asia: Its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Vegetation History and Archaeobotany*, *28*(6), 691–702. https://doi.org/10.1007/s00334-019-00731-8

McPartland, J. M., Small, E., & Hammond, K. M. (2023). A classification of cannabis: Evolutionary and domestication insights. *Journal of Ethnopharmacology*, *306*, 116260. https://doi.org/10.1016/j.jep.2022.116260

Mikos, R. A., & Kam, C. D. (2019). Has the "M" word been framed? Marijuana, cannabis, and public opinion. *PLoS ONE*, *14*(10), e0224289. https://doi.org/10.1371/journal.pone.0224289

Moliterni, V. M. C., Cattivelli, L., Ranalli, P., & Mandolino, G. (2004). The sexual differentiation of *Cannabis sativa* L.: A morphological and molecular study. *Euphytica*, *140*(1), 95–106. https://doi.org/10.1007/s10681-004-4758-7

Mukherjee, A., Roy, S. C., De Bera, S., Jiang, H.-E., Li, X., Li, C.-S., & Bera, S. (2008). Results of molecular analysis of an archaeological hemp (*Cannabis sativa* L.) DNA sample from north-west China. *Genetic Resources and Crop Evolution*, *55*(4), 481–485. https://doi.org/10.1007/s10722-008-9343-9

Murphy, T. M., Ben-Yehuda, N., Taylor, R. E., & Southon, J. R. (2011). Hemp in ancient rope and fabric from the Christmas Cave in Israel: Talmudic background and DNA sequence identification. *Journal of Archaeological Science*, *38*(10), 2579–2588. https://doi.org/10.1016/j.jas.2011.05.004

Naing, A. H., & Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiologia Plantarum*, *172*(3), 1711–1723. https://doi.org/10.1111/ppl.13373

Neacsu, M., Christie, J. S., Duncan, G. J., Vaughan, N. J., & Russell, W. R. (2022). Buckwheat, fava bean and hemp flours fortified with anthocyanins and other bioactive phytochemicals as sustainable ingredients for functional food development. *Nutraceuticals*, *2*(3), 150–161. https://doi.org/10.3390/nutraceuticals2030011

Paryani, T. R., Sosa, M. E., Page, M. F. Z., Martin, T. J., Hearvy, M. V., Ojeda, M. A., Koby, K. A., Grandy, J. J., Melshenker, B. G., Skelly, I., & Oswald, I. W. H. (2024). Nonterpenoid chemical diversity of cannabis phenotypes predicts differentiated aroma

characteristics. *ACS Omega*, *9*(26), 28806–28815. https://doi.org/10.1021/acsomega.4c03225

Pate, D. W. (1987). Chemical ecology of cannabis. *Journal of the International Hemp Association*, *4*(2), 29–35.

Pisupati, R., Vergara, D., & Kane, N. C. (2018). Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. *BMC Genomics*, *19*(1), 156. https://doi.org/10.1186/s12864-018-4494-3

Prentout, D., Razumova, O., Rhoné, B., Badouin, H., Henri, H., Feng, C., Käfer, J., Karlov, G., & Marais, G. A. B. (2020). An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. *Genome Research*, *30*(2), 164–172. https://doi.org/10.1101/gr.251207.119

Pourseyed Lazarjani, M., Torres, S., Hooker, T., Fowlie, C., Young, O., & Seyfoddin, A. (2020). Methods for quantification of cannabinoids: A narrative review. *Journal of Cannabis Research*, *2*(1), 35. https://doi.org/10.1186/s42238-020-00040-2

Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M. L., Yang, Y., Liu, A., Ravikanth, G., Nawaz, M. A., Mumtaz, A. S., Salamin, N., & Fumagalli, L. (2021). Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Science Advances*, *7*(29), eabg2286. https://doi.org/10.1126/sciadv.abg2286

Rull, V. (2021). The origins and domestication of *Cannabis*: New insights from fossil pollen records. *Vegetation History and Archaeobotany*, *30*(6), 845–855. https://doi.org/10.1007/s00334-021-00829-6

Rull, V. (2022). Origin, early expansion, domestication and anthropogenic diffusion of cannabis, with emphasis on Europe and the Iberian Peninsula. *Perspectives in Plant Ecology, Evolution and Systematics*, *55*, 125670. https://doi.org/10.1016/j.ppees.2022.125670

Ryu, B.-R., Gim, G.-J., Shin, Y.-R., Kang, M.-J., Kim, M.-J., Kwon, T.-H., Lim, Y.-S., Park, S.-H., & Lim, J.-D. (2024). Chromosome-level haploid assembly of *Cannabis sativa* L. cv. Pink Pepper. *Scientific Data*, *11*(1), 1442. https://doi.org/10.1038/s41597-024-04288-8

Russo, E. B., Jiang, H.-E., Li, X., Sutton, A., Carboni, A., del Bianco, F., Mandolino, G., Potter, D. J., Zhao, Y.-X., Bera, S., Zhang, Y.-B., Lü, E.-G., Ferguson, D. K., Hueber, F., Zhao, L.-C., Liu, C.-J., Wang, Y.-F., & Li, C.-S. (2008). Phytochemical and genetic analyses of ancient cannabis from Central Asia. *Journal of Experimental Botany*, *59*(15), 4171–4182. https://doi.org/10.1093/jxb/ern260

Saigo, T., Wang, T., Watanabe, M., & Tohge, T. (2020). Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. *Current Opinion in Plant Biology*, *55*, 93–99. https://doi.org/10.1016/j.pbi.2020.04.001

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., & Fernie, A. R. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: Structural and genetic diversity. *Plant Physiology and Biochemistry*, *72*, 21–34. https://doi.org/10.1016/j.plaphy.2013.02.001

Schultes, R. E. (1970). Random thoughts and queries on the botany of cannabis. In S. H. C. C. R. B. Joyce (Ed.), *The botany and chemistry of cannabis* (pp. 11–33). J. & A. Churchill.

Shi, J., Toscani, M., Dowling, C. A., Schilling, S., & Melzer, R. (2025). Identification of genes associated with sex expression and sex determination in hemp (*Cannabis sativa* L.). *Journal of Experimental Botany*, *76*(1), 175–190. https://doi.org/10.1093/jxb/erae429

Singh, A., Kumar, A., Yadav, S., & Singh, I. K. (2019). Reactive oxygen species-mediated signaling during abiotic stress. *Plant Gene*, *18*, 100173. https://doi.org/10.1016/j.plgene.2019.100173

Smart, E. (2015). The cannabis classification conundrum. *Botanical Journal of the Linnean Society*, *179*(2), 161–172. https://doi.org/10.1111/boj.12235

Small, E. (2015). Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. *The Botanical Review*, *81*(3), 189–294. https://doi.org/10.1007/s12229-015-9157-3

Small, E., & Antle, T. (2003). A preliminary study of pollen dispersal in *Cannabis sativa* in relation to wind direction. *Journal of Industrial Hemp*, *8*, 37–50. https://doi.org/10.1300/J237v08n02_06

Soler, S., Gramazio, P., Figàs, M. R., Vilanova, S., Rosa, E., Llosa, E. R., Borràs, D., Plazas, M., & Prohens, J. (2017). Genetic structure of *Cannabis sativa* var. indica cultivars based on genomic SSR (gSSR) markers: Implications for breeding and germplasm management. *Industrial Crops and Products*, *104*, 171–178. https://doi.org/10.1016/j.indcrop.2017.04.043

Sommanno, S. R., Chittasupho, C., Ruksiriwanich, W., & Jantrawut, P. (2020). The cannabis terpenes. *Molecules*, *25*(24), 5792. https://doi.org/10.3390/molecules25245792

Stack, G. M., Quade, M. A., Wilkerson, D. G., Monserrate, L. A., Bentz, P. C., Carey, S. B., Grimwood, J., Toth, J. A., Crawford, S., Harkess, A., & Smart, L. B. (2025). Comparison of recombination rate, reference bias, and unique pangenomic haplotypes in *Cannabis sativa* using seven de novo genome assemblies. *International Journal of Molecular Sciences*, *26*(3), 1165. https://doi.org/10.3390/ijms26031165

Steel, L., Welling, M., Ristevski, N., Johnson, K., & Gendall, A. (2023). Comparative genomics of flowering behavior in *Cannabis sativa*. *Frontiers in Plant Science*, *14*, 1227898. https://doi.org/10.3389/fpls.2023.1227898

Streisfeld, M. A., Liu, D., & Rausher, M. D. (2013). Predictable patterns of constraint among anthocyanin-regulating transcription factors in *Ipomoea*. *New Phytologist*, *200*(1), 49–57. https://doi.org/10.1111/nph.12351

Suchoff, D. H., Davis, J. M., Copeland, J. E., & Creamer, R. (2024). Agronomic performance and cannabinoid profile of industrial hemp cultivars in North Carolina. *Industrial Crops and Products*, *214*, 118141. https://doi.org/10.1016/j.indcrop.2024.118141

Tohge, T., de Souza, L. P., & Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *Journal of Experimental Botany*, *68*(15), 4013–4028. https://doi.org/10.1093/jxb/erx177

Toth, J. A., Smart, L. B., Smart, C. D., Stack, G. M., Carlson, C. H., Philippe, G., & Rose, J. K. C. (2021). Limited effect of environmental stress on cannabinoid profiles in high-

cannabidiol hemp (*Cannabis sativa* L.). *GCB Bioenergy*, *13*(10), 1666–1674. https://doi.org/10.1111/gcbb.12880

Trubanov, N., Isobe, S., Shirasawa, K., Watanabe, A., Kelesidis, G., Melzer, R., & Schilling, S. (2025). Genome-specific association study (GSAS) for exploration of variability in hemp (*Cannabis sativa*). *Scientific Reports*, *15*(1), 8371. https://doi.org/10.1038/s41598-025-92168-5

Van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., & Page, J. E. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology*, *12*(10), R102. https://doi.org/10.1186/gb-2011-12-10-r102

Van Baek, J. (2025). A novel insight into sex determination and flowering in *Cannabis sativa*. *The Plant Journal*, *101*(3), 514–528. https://doi.org/10.1111/tpj.15722

Wang, L.-S., & Stoner, G. D. (2008). Anthocyanins and their role in cancer prevention. *Cancer Letters*, *269*(2), 281–290. https://doi.org/10.1016/j.canlet.2008.05.020

Worth, R. (2019). The science of cannabis: The basics. *Journal of Cannabis Research*, *1*(1), 1–12. https://doi.org/10.1186/s42238-019-0002-1

Xu, Z., Mahmood, K., & Rothstein, S. J. (2017). ROS induces anthocyanin production via late biosynthetic genes and anthocyanin deficiency confers the hypersensitivity to ROS-generating stresses in *Arabidopsis*. *Plant and Cell Physiology*, *58*(8), 1364–1377. https://doi.org/10.1093/pcp/pcx073

Xu, G., Liu, Y., Yu, S., Kong, D., Tang, K., Dai, Z., Sun, J., Cheng, C., Deng, C., Yang, Z., Tang, Q., Li, C., Su, J., & Zhang, X. (2024). CsMIKC1 regulates inflorescence development

and grain production in *Cannabis sativa* plants. *Horticulture Research*, *11*(8), uhae161. https://doi.org/10.1093/hr/uhae161

Yadav, S. P. S., Kafle, M., Ghimire, N. P., Shah, N. K., Dahal, P., & Pokhrel, S. (2023). An overview of phytochemical constituents and pharmacological implications of *Cannabis sativa* L. *Journal of Herbal Medicine*, *42*, 100798. https://doi.org/10.1016/j.hermed.2023.100798

Yin, M., Pan, L., Liu, J., Yang, X., Tang, H., Zhou, Y., Huang, S., & Pan, G. (2022). Proanthocyanidins alleviate cadmium stress in industrial hemp (*Cannabis sativa* L.). *Plants*, *11*(18), 2450. https://doi.org/10.3390/plants11182364

Yuan, Y.-W., Sagawa, J. M., Young, R. C., Christensen, B. J., & Bradshaw, H. D. Jr. (2013). Genetic dissection of a major anthocyanin QTL contributing to pollinator-mediated reproductive isolation between sister species of *Mimulus*. *Genetics*, *194*(1), 255–263. https://doi.org/10.1534/genetics.112.146852

Zhang, Y., Butelli, E., & Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Current Opinion in Plant Biology*, *19*, 81–90. https://doi.org/10.1016/j.pbi.2014.05.011

Zheng, X., Om, K., Stanton, K. A., Thomas, D., Cheng, P. A., Eggert, A., Simmons, E., Yuan, Y.-W., Conradi Smith, G. D., Puzey, J. R., & Cooley, A. M. (2021). The regulatory network for petal anthocyanin pigmentation is shaped by the MYB5a/NEGAN transcription factor in *Mimulus*. *Genetics*, *217*(2), iyaa036. https://doi.org/10.1093/genetics/iyaa036

Chapter 2: Genetic Mapping and SNP Discovery for Anthocyanin Accumulation

Abstract

The molecular basis of anthocyanin accumulation in *Cannabis* is not well understood. Previous mapping studies revealed significant marker associations on chromosomes 6 and 8, and identified a candidate R2R3-MYB transcription factor, *CsMYB78* as a putative regulator of anthocyanin biosynthesis in floral and leaf tissues (S. Kim, Unpublished). To validate previous findings and further investigate genomic regions affecting anthocyanin accumulation across tissue types, we performed a genome wide association study (GWAS) on an F_1 population derived from highly heterogenous parents. In this analysis, significant marker associations were found on chromosome 1 from 85.08Mb-87.92Mb and on chromosome 6 from 66.75Mb-77.75Mb, overlapping with previously identified marker clusters. Within the marker cluster on chromosome 1, there were three candidate genes sharing homology with known regulators of anthocyanin biosynthesis, *CsGLABRA3*, *CsMYB17*, and *CsANTHOCYANINLESS2*. Within the cluster on chromosome 6, two candidate regulatory genes were found, *CsMYB78*, and *CsbHLH143*. PCR-based fine mapping attempts uncovered cultivar-dependent genetic variation surrounding *CsMYB78*, complicating SNP-based molecular marker development. To circumvent this, RNA sequencing was performed, and consensus gene sequences were reconstructed. Marker development for varying consensus SNPs is underway. This study uncovered candidate genes putatively impacting anthocyanin accumulation in cannabis and reports the development of a molecular marker for the rapid integration of alleles causing purple

pigmentation. Taken together, these results suggest tissue-specific regulation mechanisms of anthocyanin biosynthesis in cannabis flowers as opposed to leaves and stems, paving the way for the development of unique, and aesthetic new cultivars.

Introduction

Cannabis sativa L. is a crop with versatile uses for its fibers, grains, and diverse secondary metabolites. The most famous secondary metabolites in cannabis are the titular cannabinoids, well known for their psychoactive and medicinal properties and their economic importance (Livingston et al., 2020). Cannabinoids are among hundreds of secondary metabolites produced in cannabis including terpenes, esters, hydroxyls, volatile organic compounds (VOCs), and flavonoids, many of which confer agronomic or economically important traits. Of these, the flavonoids are a class of phenylpropanoid metabolites with varied functions in plant defense against biotic and abiotic stresses, cellular structure, maintaining homeostasis, and in some cases attracting pollinators (Naing et al., 2021). The anthocyanins are water-soluble glycoside pigments best known for the vibrant pink, purple, and red colors they give to plant tissues when localized to the highly acidic vacuole (Li et al., 2022). Anthocyanins also have an important role in regulating oxido-reductive homeostasis and reactive oxygen species (ROS) scavenging following induction from signal cascades or from primary metabolism machinery breakdown (Naing et al., 2021). In cannabis, anthocyanins are most recognizable in the flowers and remain visible in plant tissue after commercial processing methods, providing a good way for breeders and consumers to visually discriminate between cultivars. Historically purple cannabis has been anecdotally associated with higher quality by consumers, despite a dearth of research to

support this. However, the psychological tie between pigmentation and quality remains an important driver of consumer preference in the floral cannabis market and is an important trait for developing highly appealing cultivars.

The basic anthocyanin biosynthetic pathway and the key regulatory transcription factors (TFs) are well described in many model and crop species and are fairly well conserved across taxa. The *PAP1*, *PAP2*, *AtMYB113*, and *AtMYB114* suite of R2R3-MYB TFs are known to upregulate anthocyanin accumulation in many species and are commonly accepted to act in concert with a bHLH TF and a WD repeat scaffold to control the expression of the late biosynthetic genes (LBGs) and thus the pigmentation of the cell (Gonzalez et al., 2008). This protein trimer is referred to as an MBW complex and is primarily responsible for the induction of pigmentation across most flowering plants. Anthocyanin accumulation is an important stress response mechanism and MBW complexes and other MYB TFs can induce anthocyanins in response to many hormonal or environmental cues including Jasmonic Acid (JA) signaling, Cytokinin (CK) signaling, UVB-radiation, temperature stress, nutrient stress, nutrient toxicity, ROS buildup, and heavy metal toxicity (Li et al., 2022). Significant differences exist across clades with respect to the aglycone decorations of anthocyanin moieties and in the MYB-mediated stress-response TF networks impacting canonical MBW activity under abiotic or biotic stress conditions (Saigo et al., 2020). Loss or changes to the patterning of anthocyanins across tissue types can contribute to pre-zygotic and post-zygotic mating barriers such as pollinator preference and environmental adaptation and changes in environmental response activity within moieties (Zhou et al., 2025). Due to the conservation of core regulators and the comparatively high variation in

auxiliary regulatory networks at both the gene and protein level, anthocyanin accumulation can be an important model trait for investigating the interplay between genetics and development in plants.

There is relatively little research on the regulation of anthocyanin accumulation in cannabis, though there have been some studies identifying and annotating putative MYBs. Bassolino et al. (2020) were the first to comprehensively categorize the MYB and bHLH transcription factors into subfamilies based on the cs10 reference genome, and suggested putatively named *CsbHLH112* (LOC115717140), *CsbHLH113* (LOC115717045), *CsbHLH114* (LOC115703590), *CsMYB82* (LOC115708959, an ortholog of *CsRUBY*) and *CsMYB87* (LOC115695758, an ortholog of *AtMYB1*) as primary candidates for regulation of trichome formation and flavonoid biosynthesis. Indeed, transcriptomic analysis of developing trichomes have recently shown *CsbHLH112* and *CsbHLH114* to be highly upregulated during the early trichome developmental stages, confirming one of their suspected functions and suggesting that they may also be involved in anthocyanin regulation. Concurrently Kundan et al. (2022) performed an *in-silico* analysis of the same cs10 reference genome and discovered congruent MYB candidate genes, which they confusingly named *CsMYB33* (LOC115708959) and *CsMYB78* (LOC115695758). These candidate genes were further analyzed by RNA sequencing in purple and green cannabis plants, and it was found that *CsMYB78* was the highest upregulated transcript in purple vs. green plants with *CsMYB33* following closely behind. Interestingly, when these candidate genes were transformed into *N. benthamiana* transient expression assays, *CsMYB33*

induced heavy pigmentation and *CsMYB78* was unable to induce strong anthocyanin pigmentation despite its high correlation with anthocyanin accumulation.

Previous genetic mapping work in the Ellison lab uncovered a strong overlapping peak on chromosome 6 from 74.54Mb – 75.77Mb associated with flower and leaf purple pigmentation in two independent mapping populations. Within this region was the putative candidate gene *CsMYB78*, previously identified by both Kundan et al. (2022) and Bassolino et al. (2020) as a possible regulator of anthocyanin biosynthesis. A peak on chromosome 8 was discovered as well spanning from 51.90Mb – 54.98Mb associated with stem pigmentation, and within this region was the putative UV-B radiation response gene Ultraviolet-B receptor (UVR8), and the flavonoid biosynthetic pathway gene Coumaroyl-CoA:anthocyanidin 3-O-glucoside-6''-O-coumaroyltransferase 2. Despite the strong significant marker associations uncovered in this study, model statistical power was hampered by a lack of replication of either population.

In this study, we aimed to confirm previous findings and to uncover new regions of the genome associated with anthocyanin accumulation in floral, leaf, and stem tissue types. To do this we performed a Genome Wide Association Study (GWAS) on an F_1 population (Early Harvest118 x Cherry Wine) between highly heterogeneous parents. We found a significant marker association cluster on chromosome 1 for leaf and stem anthocyanin and validated another cluster on chromosome 6 associated with flower anthocyanin. *CsMYB78*, previously indicated as a strong candidate gene, was validated in this replication as well. Fine mapping and bioinformatic analysis revealed significant genetic variation between varieties surrounding *CsMYB78*, hindering SNP marker development. RNA sequencing was

performed, and variants were called from recovered sequence data to discover consensus target SNPs and elucidate potential transcript differences between green and purple genotypes. This work provides strong new candidate genes and validates previous findings, contributing to further understanding of the molecular basis of anthocyanin accumulation in cannabis, and to facilitate rapid development of aesthetic and resilient new cultivars.

Methods

Population Development

Plants previously developed by S. Kim (Unpublished) as part of a BC_2F_2 mapping population derived from Early Harvest 118 (7-Mile Farms, HempLogic) x Cherry Wine (Jerry McCoy) (both day-neutral high cannabidiol (CBD) cultivars) were used in this experiment. Early Harvest 118 was used as a pollen donor, and Cherry Wine as a recurrent parent. Early Harvest 118 was previously phenotyped as a green population and Cherry Wine as purple. Upon genotyping mapping population and parental genotypes, it was discovered that among progeny genotypes, markers segregated in a 1:2:1 ratio, possibly indicating genetic heterogeneity among parental lines. Three Early Harvest 118 and four Cherry Wine individuals were selected to represent parental genotypes in downstream sequencing analyses. Observations of parental genotypes revealed high levels of heterozygosity among samples, as well as high genetic variation within both parental populations, making linkage map construction difficult. For this reason, the GWAS population is heretofore referred to as an F_1 mapping population, to better reflect the true distribution of progeny marker segregation. 355 plants were grown at the Walnut Street Greenhouse at the University of Wisconsin-Madison during the summer of 2024 for approximately 3.5 months before leaf

tissue was collected and stored for DNA extraction on August 8th, 2024. A long-day light schedule of 16h light and 8h dark was used throughout the experiment.

Phenotyping

Anthocyanin phenotyping was performed on floral (F), leaf (L), and stem tissue (S) using a visual scale ranging from zero to three, with zero indicating a lack of visible pigmentation and three indicating deep purple pigmentation (Figure 2.1). All phenotypic measurements were taken under the same conditions, and all plants were phenotyped and harvested on August 23rd, 2024. Raw phenotypic data can be found in Supplementary Table 2.1.

Figure 2.1. Phenotypic Scale for Anthocyanin Accumulation. Pigmentation intensity was visually assessed in three different tissue types, flowers, leaves, and stems on a scale of 0 to 3. A) an example plant with a score of 0 across tissue types. B) A plant with a leaf pigment score of 1. C) A plant with a leaf and flower score of 2. D) A plant with a pigmentation intensity score of 3. Images of all plants were taken in the Walnut Street Greenhouse at the University of Wisconsin-Madison, and bottom right reference images were generated by S. Kim (Unpublished).

Genome wide association analysis

DNA was extracted from lyophilized and pulverized leaf tissue (~0.2g) using the NucleoSpin 96 Plant II Kit (Macherey-Nagel, Durin, Germany) according to manufacturer instructions and sent to the University of Wisconsin Biotechnology Center for Genotyping-By-Sequencing library preparation using the *ApeK*I restriction enzyme. Sequencing was conducted on an Illumina NovaSeq X Plus sequencer with 2X150bp reads. Raw reads were filtered using the Skewer trimming software (Jiang et al., 2014), excluding any reads with a Phred quality score less than 20. Reads were further processed using the Tassel GBS Pipeline Version 2 with the Bowtie2 alignment software to remove wrongly barcoded reads, and to align raw fastq files to the cs10 reference genome (GCA_900626175.2) (Glaubitz et al., 2014; Langmead et al., 2012; Grassa et al., 2021). SNPs were called using Tassel DiscoverySNPCallerPluginV2 with an initial minimum allele frequency of 1 and locus coverage of 10%. VCF files were obtained using ProductionSNPCallerPluginV2 and further trimmed in Tassel5 to keep markers with no missing data or multi-allelic sites, and a minimum and maximum percent heterozygosity of 0.10 and 0.55, respectively.

After filtering, 355 samples and 11,270 markers remained. GAPITv4.3 (Lipka, 2012) was used for association mapping with both a mixed linear model with a kinship matrix (MLM) and a fixed and random model circulating probability unification (FarmCPU) model with a kinship matrix to verify marker associations. QQ plots generated from the MLM revealed a normal distribution of residuals with strong inflation of middling p-values, indicating that some marker peaks may be false positives. To account for this, analysis was

subsequently run using the FarmCPU model, a more conservative model presumably better able to reduce false positive results. A -log(p) significance threshold was determined as the Bonferroni corrected p-value -log10(0.05/n) for the MLM and -log10(0.01/n) for the FarmCPU model where n is the number of SNP markers used for each trait. Significant cluster areas were manually determined using significant markers identified in both FarmCPU and MLM analyses, and cluster areas were delineated as regions of chromosomes with flanking significant markers within the MLM Manhattan plots. Candidate genes from the cs10 reference genome were identified using the NCBI Genome Data Browser (Rangwala et al., 2021) and the NCBI Entrez-Direct tool to retrieve mRNA annotation information.

CsMYB78 Amplification

Indigo, an open-pollinated grain cultivar segregating for anthocyanin accumulation with a well-defined floral phenotype was selected for downstream sequence analysis of the putative candidate gene *CsMYB78*. DNA was extracted from four green and four purple samples. Primers were designed using the NCBI primer blast tool using the region annotated as LOC115695758: NC_044377.1[75409427..75419464] in the cs10 reference genome (Grassa et al., 2021). Primer sequences are listed in Supplementary Table 2.1. Amplification was performed in a 25 ml reaction volume and consisting of 10 µL nuclease free water, 12.5 µL LongAmp HotStart Platinum II TAQ master mix, 0.5 µL forward primer, 0,5 µL reverse primer, and 1 µL template diluted by half in water. Primer sequences are listed in Supplementary Table 2.2. The reaction was amplified in a BioRad T100 thermocycler (BioRad, USA) programmed for 30 seconds at 94°C as initial denaturation, followed by 28 cycles of 30 sec at 94°C for denaturation, 30 sec at 60 °C as annealing, 12 min at 65 °C for

extension, and final extension at 65°C for 10 min. For partial amplification products, an extension step of 5 min at 65°C was used. PCR products were examined by gel electrophoresis at 60 mV for 50 minutes in a 0.75% (w/v) agarose gel in 1 x TAE buffer. The electrophoresis gel was mixed with 0.01% GelRed prior to pouring and sample loading. A NEB DNA ladder 10 kb was used to determine fragment size. Gels were imaged using a BioRad Gel Doc XR+ (BioRad, USA).

***CsMYB78* RNA sequencing**

Apical floral tissue was taken from mature (N = 5) purple and (N = 3) green Indigo plants and placed immediately into liquid nitrogen and stored at -80°C before being extracted using the Qiagen RNeasy Power Plant RNA Extraction kit (Qiagen, USA) with 20uL of phenolic separation solution, or the Qiagen RNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer protocol. RNA samples were DNase treated after extraction using the Turbo DNase kit (Qiagen, USA) according to manufacturer instructions. RNA samples were run on a Biomek Plate Reader for quantification and on an Agilent 4200 Tapestation (High Sensitivity RNA screentape) to determine RNA Integrity Number (RIN) scores. RNA libraries were constructed using the NEB Ultra II Directional mRNA Sample Preparation Guide using 100ng of RNA input. The samples were Poly(A)+ enriched using oligo(dT) bead capture then fragmented for 5 minutes before synthesis of double stranded cDNA and purification with NEB SPRIselect beads. 3’ ends of reads were then adenylated and adapters ligated before further 0.8X purification of double stranded cDNA (NEBNext Sample Purification beads) and 8 cycles of PCR amplification of purified cDNA. Final purification was performed on amplified samples (NEBNext Sample Purification beads), and mRNA libraries were

quantified with PicoGreen Assay (Thermo Fisher Scientific, USA) on an Agilent Biotek Synergy H1 Microplate and assayed on an Agilent 4200Tapestation (High Sensitivity D1000 screentape). RNA yields in ng are listed in Supplementary Table 2.3. In between all steps samples were kept at a maximum of -20°C and were thawed on ice according to protocol directions before library preparation.

RNA sequencing was performed on five purple and three green samples by the University of Wisconsin Biotechnology Center on an Illumina Novaseq X plus sequencer using 0.5 10B 2x150bp lanes. Prior to analysis, raw fastq files were preprocessed using skewer trimming software (Jiang et al., 2014) to remove adapter sequences from generated sequence reads and were filtered to exclude reads with a Phred quality score less than 20. Bioinformatic analysis was performed using the University of Wisconsin Bioinformatics Resource Center RNAseq pipeline, and fastq files were aligned to the cs10 reference genome (Grassa et al., 2021) and mapped using STAR (Spliced Transcripts Alignment to a Reference) to generate final binary alignment map (BAM) files for downstream analysis.

SNP Discovery and Marker Development

To generate consensus FASTA files of putative *CsMYB78* sequence, RNAseq BAM files were first grouped into green and purple sample bins, then subset to the regional bounds of *CsMYB78* within cs10 (NC_044377.1:75409156-75419265), and variants were called using samtools and bcftools (Li, 2011). Following variant calling from subset BAM files, consensus FASTA sequences were constructed using the NCBI cs10 genome assembly

(GCF_900626175.2) and bcftools consensus to create two consensus FASTA files of green and purple samples for downstream analysis.

Alignments of consensus FASTA sequences to cs10 genomic sequence of *CsMYB78* was performed using the ClustalW algorithm in MEGAx11 (Tamura, 2021) and exported to the NCBI multiple sequence alignment viewer for large scale visualization (https://www.ncbi.nlm.nih.gov/projects/msaviewer/) and the ggmsa package in Rv4.3 for fine scale visualization and SNP marker visual identification (Zhou et al., 2022) (Figure 2.5c). To determine local sequence homology of *CsMYB78* across cannabis pangenomic accessions and to assess SNP marker conservation, instantBLASTN alignment was performed on the putative regions surrounding *CsMYB78* as retrieved using BLAST with the cannabis pangenome as a database and *CsMYB78* sequence retrieved from cs10 as a query (Lynch et al., 2025) (Figure 2.5b). Visualization of instantBLASTN alignments of the FCS1a, GERv1b, COSVa, NLv1b, PPFBa, BCMa, SN1v3b, and TKFBa cannabis genome assemblies was performed in Persephone genome viewer (https:persephonesoft.com/). Sequences were retrieved from the cannabis pangenome assemblies (Lynch et al., 2025) available using BLAST with the full-length sequence of Indigo purple samples. Sequences were aligned using instant BLASTN.

Results

Genotyping-By-Sequencing

After variant and genotype cleaning, a total of 11,270 markers and 355 individuals were left for GWAS analysis. Markers were fairly evenly distributed across chromosomes,

ranging from 697 markers on chromosome 7 to 1386 markers on chromosome 3. On average the marker count per chromosome was 1122, indicating a relatively equal spread of markers across the 10 chromosomes. Among flower, leaf, and stem anthocyanin association tests, 165 significant SNPs were found across chromosomes 1, 2, 3, 5, 6, and 7 using the MLM model parameters. The percentage of phenotypic variation explained (PVE) by significant markers ranged from 0 to 74.58%. Marker significance and PVE for each trait and both models can be found in Supplementary Tables 2.3-2.8.

Leaf Pigmentation (Anthocyanins)

Six major clusters comprised of 46 significant SNPs were identified for purple leaf pigmentation (Figure 2.2, Supplementary Tables 2.4 and 2.5) Cross referencing MLM cluster regions with FarmCPU results confirmed the association of leaf anthocyanin pigmentation with a cluster on chromosome 1 from [85266373..85417293], a cluster on chromosome 6 from [7141542..7141563], two significant markers on chromosome 8 at positions 43821206 and 51441251, and one significant marker on chromosome 9 at position 56362569.

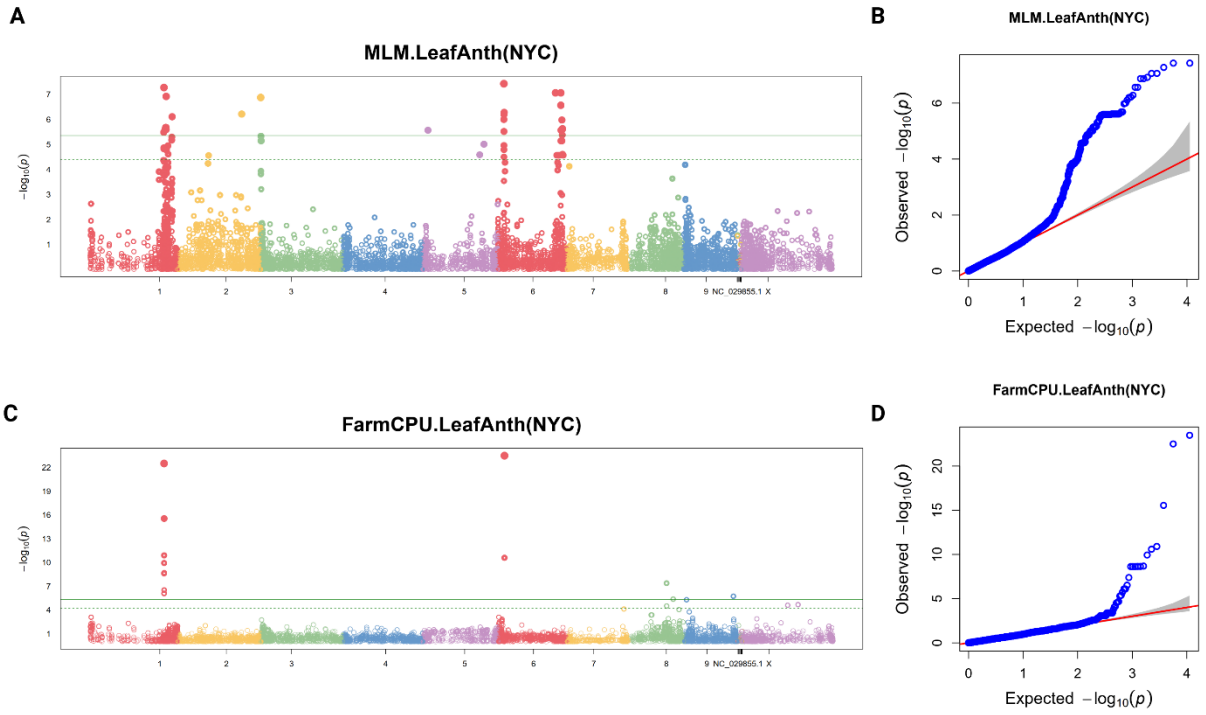

Figure 2.2. Manhattan plots and Quantile-Quantile (QQ) plots for leaf anthocyanin accumulation. A-B) Manhattan plot and associated QQ plot using MLM parameters. C-D) Manhattan plot and associated QQ plot using FarmCPU model parameters. The solid green line signifies a Bonferroni corrected 0.01 alpha threshold and the dashed green line a Bonferroni corrected 0.05 alpha significance threshold.

Flower Pigmentation

Nine clusters comprised of 64 significant SNPs were identified for flower pigmentation (Figure 2.3, Supplementary Tables 2.6 and 2.7). There were two significant markers under the FarmCPU model on chromosome 4 at 77029671 and on chromosome 6 at 66752055, providing support for the cluster area on chromosome 6 at [66752055..75779274] identified with the MLM model and in previous work (S. Kim unpublished).

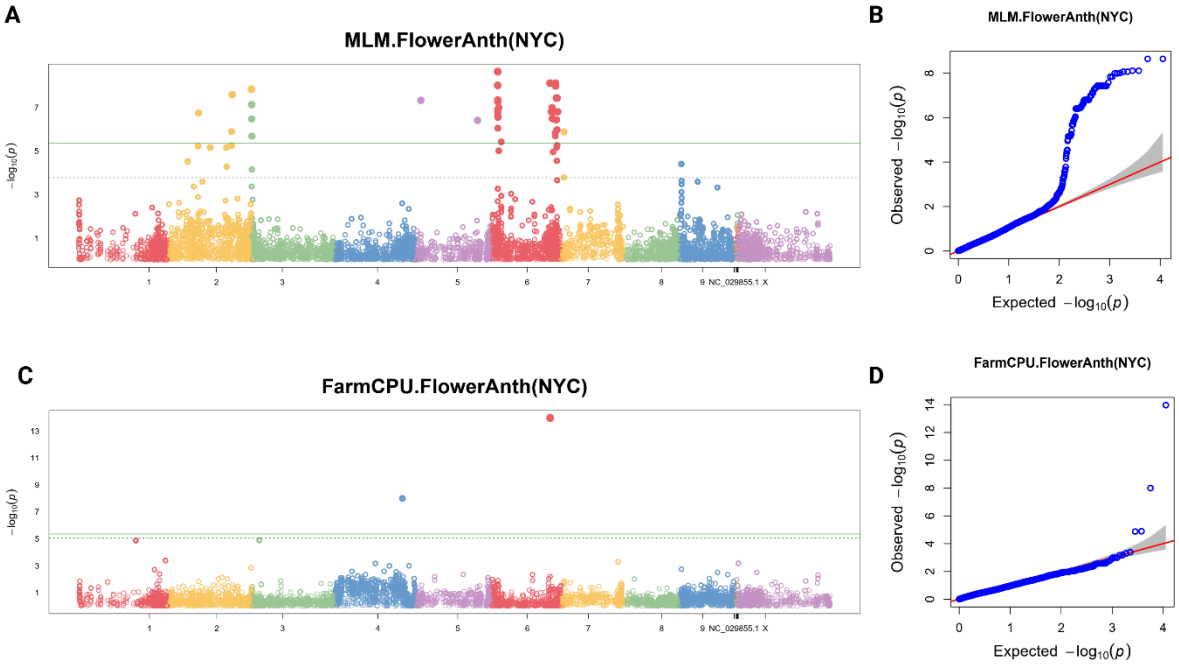

Figure 2.3. Manhattan Plots and Quantile-Quantile (QQ) Plots for flower anthocyanin accumulation. A-B) Manhattan plot and associated QQ plot using MLM parameters. C-D) Manhattan plot and associated QQ plot using FarmCPU model parameters. The solid green line signifies a Bonferroni corrected 0.01 alpha threshold and the dashed green line a Bonferroni corrected 0.05 alpha significance threshold.

Stem Pigmentation

Ten clusters comprised of 55 significant SNPs were identified for stem anthocyanin in the MLM analysis (Figure 2.4, Supplementary Tables 2.8 and 2.9). Within the FarmCPU model there were two significant markers, one on chromosome 1 at position 94777564 and another on chromosome 5 at position 71299408, confirming these regions for candidate gene identification.

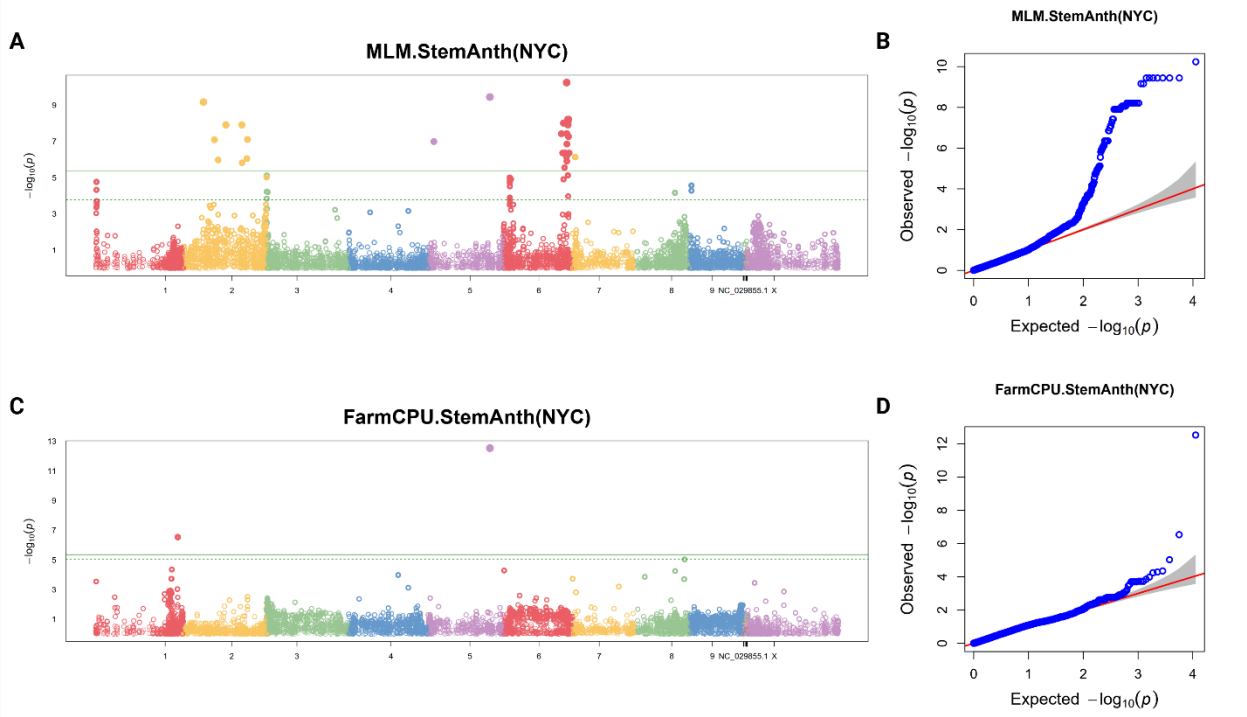

Figure 2.4. Manhattan Plots and Quantile-Quantile (QQ) Plots for stem anthocyanin accumulation. A-B) Manhattan plot and associated QQ plot using MLM parameters. C-D) Manhattan plot and associated QQ plot using FarmCPU model parameters. The solid green line signifies a Bonferroni corrected 0.01 alpha threshold and the dashed green line a Bonferroni corrected 0.05 alpha significance threshold.

Identification of Candidate Genes

Candidate genes were mined from the cs10 reference genome using the bounds of peak regions discovered in the MLM analysis and confirmed using the FarmCPU analysis. Putative candidate gene annotations were identified using XM transcript annotation records for the cs10 reference genome assembly (Grassa et al., 2021). Within the marker association regions, there were a total of 279 candidate genes for flower anthocyanin within the cluster boundaries on chromosome 6 and 228 candidate genes for leaf and stem anthocyanin within the marker cluster boundaries on chromosome 1. Gene description information from the

NCBI Entrez-Direct database is included in supplementary tables 2.10 and 2.11, with candidate genes sharing homology with known anthocyanin regulators in bold. Notable genes annotated as orthologs of known regulators of anthocyanin biosynthesis within the region on chromosome 6 were *CsMYB78* (LOC115695758), a previously identified candidate transcription factor affecting anthocyanin biosynthesis (S. Kim, Unpublished), and *CsbHLH143* (LOC115725351) on chromosome 6, a putative MBW complex constituent possibly associated with anthocyanin accumulation (Wang et al., 2022). Notable candidate genes sharing homology with known anthocyanin regulators within the region on chromosome 1 were *CsMYB17* (LOC115706528), *CsGLABRA3* (LOC115706692), and *CsANTHOCYANINLESS2* (LOC115706571).

Fine Mapping and Marker Development

Marker development for *CsMYB78* has been underway since its candidacy (S. Kim, Unpublished) but has been hindered by high genetic variation at this locus both within and between cultivars. PCR primers based on the cs10 version of *CsMYB78* were unable to retrieve full length DNA amplicon sequences from both green and purple samples but did reveal the presence of a large ~1.6Kb insertion in the 5' UTR of purple samples relative to green (Fig 2.5a). With the release of the cannabis pangenome and the development of the persephone software for visualizing genomic data, large scale insertions surrounding *CsMYB78* can be seen across pangenome accessions (Fig 2.5b). Further RNA sequencing of green and purple samples revealed significant overall variation relative to cs10 within the transcriptome of green and purple samples, suggesting that a *de novo* transcriptomic assembly will be required to fully elucidate differential transcript accumulation. However,

alignments of consensus sequences of green and purple samples revealed strong sequence variation between genotypes (Fig 2.5c). Development of a florescent-based co-dominant marker assay for this region is underway.

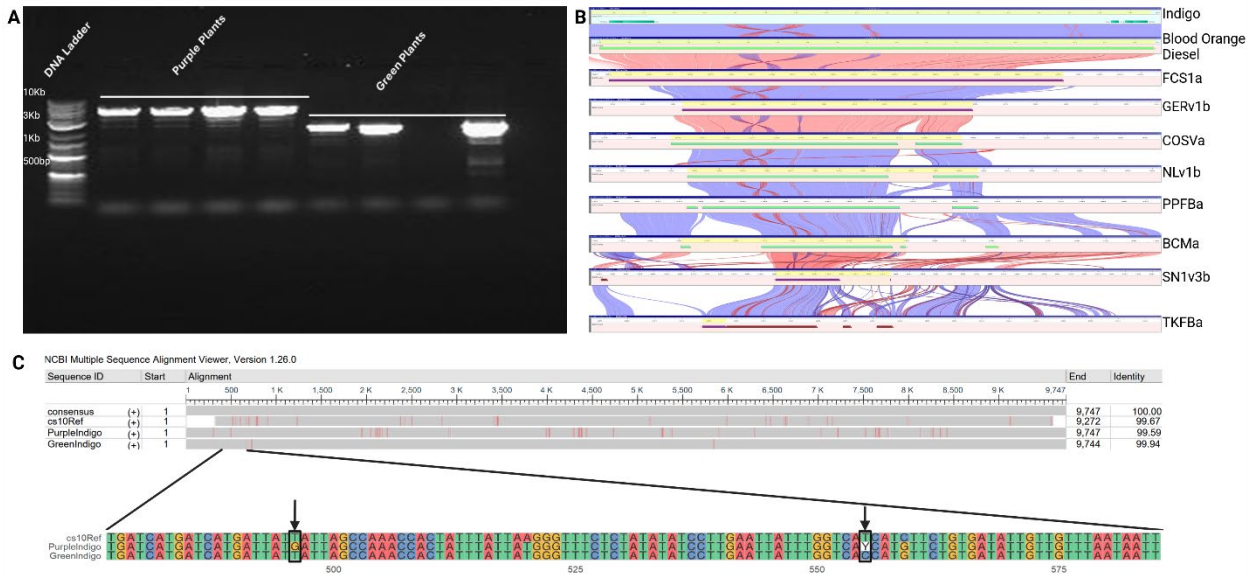

Figure 2.5. Genetic Variability Surrounding *CsMYB78*. A) 0.7% agarose gel containing PCR fragments amplified from Indigo genomic DNA sequence is shown, with a ladder to the left for reference. Purple samples show a 1.6Kb insertion relative to green individuals. B) Genotypic variability around the region of *CsMYB78*. (a) *CsMYB78* genomic regions of pangenome accessions are aligned with instantBLASTN using (in order) Indigo, Blood Orange Diesel, FCS1a, GERv1b, COSVa, NLv1b, PPFBa, BCMa, SN1v3b, and TKFBa. The ribbons represent regions of sequence homology. Genomes constructed antiparallel to others have ribbons colored red. Blue ribbons represent sequence in the same phase. Sequences were visualized in persephonesoft.com. C) Multiple sequence alignment of cs10 *CsMYB78* sequence relative to consensus sequences of purple and green Indigo samples. Putative SNP markers are outlined and indicated with arrows.

Discussion

Known for its secondary metabolites, grains, and fibers, *Cannabis sativa* is a versatile crop with an abundance of public interest. However, the development of cannabis cultivars has been slowed by a lack of knowledge of the genetic basis underpinning aesthetic and

stress tolerance traits. Anthocyanin pigmentation is an important trait strongly influencing consumer preference in the floral cannabis market and potentially conferring some abiotic stress tolerance and thus is a strong target for breeding marketable and environmentally resilient cultivars. In this study we validated significant marker associations with floral pigmentation on chromosome 6 from 74.54Mb – 75.77Mb using GWAS on a population of 355 green and purple individuals. We also uncovered a significant marker cluster associated with leaf and stem pigmentation on chromosome 1 from 85.26Mb – 85.41Mb, allowing for investigation of candidate genes imposing tissue-specific influence on the diverse pigmentation phenotypes visible across different cannabis cultivars. Several significant candidate genes were discovered in these regions including *CsMYB78* (an ortholog of *AtMYB1*) and *CsbHLH143* in the region on chromosome 6, and *CsMYB17*, *CsGLABRA3* (*CsGL3*), and *CsANTHOCYANINLESS2* (*CsANL2*) in the region on chromosome 1. Molecular marker development for *CsMYB78* is currently underway, facilitated by alignments of green and purple consensus sequences.

Candidate Genes

Within the bounds of the marker cluster on chromosome 6 we uncovered *CsMYB78* (LOC115695758), a putative ortholog of the R2R3-MYB TF *AtMYB1*, an environmentally induced positive regulator of anthocyanin biosynthesis in Onion (*Allium cepa* L.), Apple (*Malus domestica*), and Eggplant (*Solanum melongena* L.) (Schwinn et al., 2016; An et al., 2018; Yang et al., 2021; Filyushin et al., 2023). This gene resides on chromosome 6 from 75,409,856..75,419,127 in the cs10 reference genome (Grassa et al., 2021) and is proposed to influence floral anthocyanin accumulation. In Eggplant, *SmMYB1* expression was

correlated with *SmbHLH143* in fruit peel tissue, indicating that these genes may interact to form an MBW complex for the regulation of anthocyanin accumulation in different tissue types. Previous association studies from S. Kim (Unpublished) yielded *CsMYB78* as a strong candidate regulator of anthocyanin accumulation within the region on chromosome 6, as well as Ultraviolet-B receptor (*UVR8*) (LOC115700277) and Coumaroyl-CoA:anthocyanidin 3-O-glucoside-6''-O-coumaroyltransferase 2 (LOC115700963), which were not reproduced in this analysis.

Many cs10 gene annotations were recently censored or altered, suggesting that *CsMYB78* may not be a true ortholog of *AtMYB1* and casting some doubt on the annotations of other cs10 genes. Despite this, several studies have annotated and phylogenetically clustered MYB and bHLH gene families in cannabis and confirmed *CsMYB78* as an R2R3-MYB TF related to *AtMYB1* (Bassolino et al., 2020). Kundan et al. (2022) tested the proposed activation of anthocyanin accumulation via *CsMYB78* using transient expression assays in *N. benthamiana* and found that *CsMYB78* was only able to induce low levels of anthocyanins. This was in sharp contrast to *CsMYB33*, an R2R3-MYB TF identified as a strong candidate gene for anthocyanin accumulation by both Kundan et al. (2022) and Bassolino et al. (2020), which was able to induce high anthocyanin accumulation in transient expression assays. Interestingly, no markers physically close to *CsMYB33* (LOC115708959) (Chr 3: 16,488,389..16,492,725) were found to be significant in this analysis or in previous mapping studies performed by the Ellison lab, hinting at differential regulation mechanisms between disparate cannabis cultivars. This idea is further supported by the existence of distinct,

tissue specific marker clusters in association mapping (Figures 2.2, 2.3, and 2.4), and the diverse pigmentation patterns or breakdown thereof between cannabis genotypes.

The bHLH TF *bHLH143* (LOC115725351), a poorly characterized gene in *Arabdopsis* was discovered on chromosome 6 from 70,709,522..70,713,911 and is associated with flower anthocyanin pigmentation. In *Capsicum anuum* L., an ortholog of *bHLH143* increased in expression under continuous blue light radiation, coinciding with upregulation in anthocyanin biosynthetic enzymes and known regulators (Zhou et al., 2024). bHLH TFs are crucial for the proper formation of MBW complexes and subsequent upregulation of anthocyanin biosynthetic enzymes, though this function is only one of the many roles within the protein family. Further bioinformatic and molecular analysis will be required to determine the function of *bHLH143* in cannabis, but it is possible that this gene may contribute to flower tissue specific anthocyanin induction or to stress-responsive induction.

Three potential candidate genes were discovered within the marker cluster on chromosome 1 for leaf anthocyanin. The regulatory enzyme homeobox-leucine zipper protein *ANTHOCYANINLESS 2* (LOC115706571), a possible activator of anthocyanin biosynthesis, was discovered on chromosome 1 from 87067285 – 87073409. Kubo et al. (2008) performed expression analysis on *ANL2* using *in-situ* hybridization and β-Glucuronidase (GUS) staining to determine tissue localization of mRNA accumulation and expression and found that *ANL2* specifically accumulates in shoot apex, young leaves, root tips, anthers and ovules. In rosette tissue, specific subepidermal *ANL2* expression was observed corresponding to anthocyanin accumulation, while neither strong expression nor anthocyanin accumulation was observed in rosette epidermal cells. In *anl2* knockout

mutants, root hair organization was disturbed, and cell wall extensibility was lessened, suggesting that *ANL2* is also involved in root architecture organization or in regulating cell wall polysaccharide composition. Thus, *ANL2* may be an upstream regulator of different developmental or biochemical pathways dependent on tissue type (Kubo et al., 1999; Mabuchi et al., 2016).

A putative ortholog of *GLABRA3* (*GL3*) (LOC115706692), a strong canonical bHLH TF regulating anthocyanin accumulation in model and crop species was discovered on chromosome 1 from 86615985 - 86618426 (Gonzalez et al., 2008; Tominaga-Wada et al., 2013). *GL3* is a well-known and characterized constituent of the canonical MBW complex and regulator of anthocyanin biosynthesis, positioning this as a strong candidate gene for future analysis. Interestingly, *CsGL3* did not appear to be annotated in previous studies investigating the diversity of bHLH TFs in cannabis. In *Arabidopsis, GL3* displays tissue specific expression and has been shown to affect both anthocyanin accumulation in apical organs and root hair epidermal patterning in primary and lateral roots (Bernhardt et al., 2005).

The final candidate gene found on chromosome 1 for leaf anthocyanin was an orthologous R2R3-MYB TF to *AtMYB17*, a gene putatively involved in floral initiation and cell fate determination in *Arabidopsis* and in activating flavonoid biosynthesis in Pear fruit (Zhang et al., 2009; Premathilake et al., 2020). The functionality of *MYB17* may be diverse, as in *Cichorium intybus* it was found to upregulate fructan biosynthesis and degradation enzymes in a hairy root culture transient expression system (Wei et al., 2017).

The identification of five novel candidate genes on chromosomes 1 and 6 provide strong targets for future investigation into the transcriptional regulation of anthocyanin accumulation in cannabis. Furthermore, given the homology of candidates to known genes across other species, the roles of orthologs of *ANL2, GL3, MYB17, CsMYB78,* and *bHLH143* could suggest different MBW complex constituents based on tissue type and/or environmental stimuli.

Marker Development for *CsMYB78*

A main goal of this study was the development and validation of molecular markers capable of differentiating green and purple genotypes across cannabis cultivars to aid in rapid development of aesthetically pleasing and environmentally resilient cultivars. This goal was motivated by the strong tendency of cannabis plants to display delayed pigmentation accumulation, often remaining green until the onset of senescence, and hindering efforts to effectively select purple plants during reproductive life stages. Moreover, PCR amplification and RNA sequencing revealed significant genetic variation around *CsMYB78* both within and between varieties, stymying efforts to develop broadly compatible molecular markers. PCR amplification of full-length sequences of *CsMYB78* was successful for genomic DNA of purple Indigo samples and initially revealed a ~1.6Kb insertion in the 5'UTR of *CsMYB78* in purple samples relative to green (Fig. 2.5a). In contrast, only sequence covering the 5'UTR and first two exons of *CsMYB78* was amplified from green samples, indicating some genetic variability surrounding this region. Further, the presence of the aforementioned insertion was not broadly indicative of a purple phenotype in disparate varieties, disqualifying it from use in a gel-based genotyping scheme. Pangenomic alignment of full-length Indigo

sequences also showed differential insertions in the promoter, terminator, and exons of *CsMYB78* across accessions (Figure 2.5b), complicating attempts at developing markers on conserved SNPs.

To circumvent challenges in constructing primers for a highly variable region the RNA of purple and green Indigo samples was sequenced, and consensus gene sequences were constructed and aligned. This revealed several putative SNP and indel markers, which are being cross validated using the cannabis pangenome and available genomic DNA of several cultivars.

Limitations

In this study, genome wide association mapping was used to identify significant clusters of markers associated with flower, leaf, or stem pigmentation using a population originally constructed as a BC_2F_2 mapping population for use in linkage mapping by S. Kim (unpublished). Upon sequencing four Cherry Wine (P1) and three Early Harvest 118 (P2) parents were genotyped along with 348 F_1 progenies it was revealed that not only were the parental genotypes highly heterozygous but that the parental genotype and phenotype was variable between individuals. This, combined with the high heterozygosity of both parents and offspring, made effective construction of a genetic map unfeasible due to the inability to accurately assign marker dosage, and uncertain identification of recombination events. Due to these limitations, the population was referred to as an F_1 population throughout. Thus, a Genome Wide Association Study (GWAS) was pursued.

Future Directions

Fluorescent genotyping markers are an excellent tool for rapid genotype determination and can ensure rapid integration of a target allele into breeding populations. In this study, marker assays were developed based on SNPs present between purple and green consensus FASTA sequences from Indigo and cs10 DNA and RNA. SNP sequences have been sent to 3CR (3CR, Essex, United Kingdom) for fluorescent allelic discrimination assay preparation and design according to in-house specifications. The development of these tools will aid in creating vigorous purple cannabis breeding populations.

Cannabis is a highly heterozygous, wind-pollinated, outcrossing crop. This is reflected in the vast majority of commercially available varieties, which lack uniformity and stability for many agronomic traits. This makes linkage mapping challenging, as parental marker dosage is difficult to determine without genotype data of both parent plants independently from their populations. The development and validation of cannabis inbred parental lines would aid in genetic mapping efforts and allow for linkage map construction to better determine the boundaries and exact locations of quantitative trait loci (QTL).

The presence of many different insertions in the promoters, terminators, and coding regions of *CsMYB78* is an unusually high level of genetic variation and may suggest that *CsMYB78* is a site of high transposon activity. Future bioinformatic analysis could investigate the presence of palindromic repeat sequences across pangenomic accessions, indicative of transposable element insertion to better understand the factors affecting *CsMYB78* expression, and to potentially expose alterations to core gene components affecting its function.

Overall, these studies have revealed a framework of genetic architecture suggesting differential mechanisms of tissue specific regulation of anthocyanin accumulation in cannabis. This knowledge will enable the future development of uniformly and attractively patterned purple cannabis cultivars, facilitate molecular investigation into the transcriptional regulation of secondary metabolite accumulation, and provide growers and breeders with a tool to facilitate environmental resilience in their crops.

References

An, J.-P., Wang, X.-F., Li, Y.-Y., Song, L.-Q., Zhao, L.-L., You, C.-X., & Hao, Y.-J. (2018). EIN3-LIKE1, MYB1, and ETHYLENE RESPONSE FACTOR3 Act in a Regulatory Loop That Synergistically Modulates Ethylene Biosynthesis and Anthocyanin Accumulation. *Plant Physiology*, *178*(2), 808–823. https://doi.org/10.1104/pp.18.00068

Anisimova, A. V., Filyushin, M. A., Shchennikova, A. V., & Kochieva, E. Z. (2023). Antocyanidin-3-O-glucosyltransferase genes in pepper (Capsicum spp.) and their role in anthocyanin biosynthesis. *Plants, 12*(4), 812. https://doi.org/10.3390/plants12040812

Bassolino, L., Wizenberg, A., Hart, J., Garfinkel, A. R., Pugh, C., & Ellison, S. (2020). Characterization of MYB and bHLH gene families in Cannabis sativa reveals candidates regulating trichome development and flavonoid biosynthesis. *Frontiers in Plant Science, 11*, 1347. https://doi.org/10.3389/fpls.2020.01347

Bernhardt, C., Zhao, M., Gonzalez, A., Lloyd, A., & Schiefelbein, J. (2005). The bHLH genes GL3 and EGL3 participate in an intercellular regulatory circuit that controls cell patterning in the Arabidopsis root epidermis. *Development, 132*(2), 291–298. https://doi.org/10.1242/dev.01549

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLOS ONE, 9*(2), e90346. https://doi.org/10.1371/journal.pone.0090346

Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/MYB transcriptional complex in Arabidopsis seedlings. *The Plant Journal, 53*(5), 814–827. https://doi.org/10.1111/j.1365-313X.2007.03373.x

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. G., ... & Schwartz, C. J.(2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist, 230*(4), 1665–1679. https://doi.org/10.1111/nph.17243

Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics, 15*, 182. https://doi.org/10.1186/1471-2105-15-182

Kubo, H., Peeters, A. J., Aarts, M. G., Pereira, A., & Koornneef, M. (1999). ANTHOCYANINLESS2, a homeobox gene affecting anthocyanin distribution and root development in Arabidopsis. *The Plant Cell, 11*(7), 1217–1226. https://doi.org/10.1105/tpc.11.7.1217

Kundan, M., Gani, U., Fayaz, M., Angmo, T., Kesari, R., Rahul, V. P., ... & Misra, P. (2022). Two R2R3-MYB transcription factors, CsMYB33 and CsMYB78 are involved in the regulation of anthocyanin biosynthesis in Cannabis sativa L. *Industrial Crops and Products, 188*, 115546. https://doi.org/10.1016/j.indcrop.2022.115546

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, C., Yu, W., Xu, J., Lu, X., & Liu, Y. (2022). Anthocyanin biosynthesis induced by MYB transcription factors in plants. *International Journal of Molecular Sciences, 23*(19), 11701. https://doi.org/10.3390/ijms231911701

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., ... & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics, 28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D. C. J., Rensing, K. H., Laflamme-Yonkman, J., ... & Samuels, A. L. (2020). Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *The Plant Journal, 101*(1), 37–56. https://doi.org/10.1111/tpj.14516

Lynch, R. C., Padgitt-Cobb, L. K., Garfinkel, A. R., Knaus, B. J., Hartwick, N. T., Allsing, N., ... & Michael, T. P.(2025). Domesticated cannabinoid synthases amid a wild mosaic cannabis pangenome. *Nature*. https://doi.org/10.1038/s41586-025-09065-0

Mabuchi, A., Soga, K., Wakabayashi, K., & Hoson, T. (2016). Phenotypic screening of Arabidopsis T-DNA insertion lines for cell wall mechanical properties revealed ANTHOCYANINLESS2, a cell wall-related gene. *Journal of Plant Physiology*, *191*, 29–35. https://doi.org/10.1016/j.jplph.2015.11.011

Naing, A. H., & Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiologia Plantarum, 172*(3), 1711–1723. https://doi.org/10.1111/ppl.13373

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., ... & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidence. *Nucleic Acids Research, 50*(D1), D543–D552. https://doi.org/10.1093/nar/gkab1038

Premathilake, A. T., Ni, J., & Chen, X. (2020). Transcriptome analysis provides new insights into the transcriptional regulation of methyl jasmonate-induced flavonoid biosynthesis in pear calli. *Plant Physiology and Biochemistry, 156*, 569–582. https://doi.org/10.1016/j.plaphy.2020.09.030

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., Rudnev, D., Shkeda, A., Weitz, E. M., & Schneider, V. A. (2021). Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome Research*, *31*(1), 159–169. https://doi.org/10.1101/gr.266932.120

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., … & Martin, C. (2016). The Onion (Allium cepa L.) R2R3-MYB Gene MYB1 regulates anthocyanin biosynthesis. *Frontiers in Plant Science, 7*, 1865. https://doi.org/10.3389/fpls.2016.01865

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution, 38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Tominaga-Wada, R., Nukumizu, Y., Sato, S., & Wada, T. (2013). Control of plant trichome and root-hair development by a tomato (Solanum lycopersicum) R3 MYB transcription factor. *PLoS ONE, 8*(7), e67688. https://doi.org/10.1371/journal.pone.0067688

Wang, Q., Ma, C., Xu, Y., Liu, L., Zhang, D., & Wang, Y. (2022). CiMYB17, a stress-induced chicory R2R3-MYB transcription factor, activates promoters of genes involved in fructan synthesis and degradation. *Plant Cell, Tissue and Organ Culture, 131*(3), 509–519. https://doi.org/10.1007/s11240-017-1302-y

Wei, L., Zhang, Y., Zhuang, Z., Li, Y., Zhang, X., & Xie, Y. (2017). CiMYB17, a stress-induced chicory R2R3-MYB transcription factor, activates promoters of genes involved in fructan synthesis and degradation. *Journal of Plant Physiology, 219*, 39–49. https://doi.org/10.1016/j.jplph.2017.09.005

Yang, L., Zhang, Y., Dong, L., Yang, J., & Ren, J. (2021). Microplastics in soil: A review on methods, occurrence, sources, and potential risk. *Science of The Total Environment, 780*, 146546. https://doi.org/10.1016/j.scitotenv.2021.146546

Zhang, X., Zhang, M., Takano, T., & Liu, S. (2009). Sugars induce anthocyanin accumulation and flavanone 3-hydroxylase expression in grape berries. *Plant Growth Regulation, 58*(3), 251–260. https://doi.org/10.1007/s10725-009-9381-8

Zhou, Y., Gao, L., Chen, J., Li, Y., Wang, Y., Li, Y., … & He, C. (2024). Integrated transcriptome and metabolome analysis reveals anthocyanin biosynthesis mechanisms in pepper (Capsicum annuum L.) leaves under continuous blue light. *BMC Plant Biology, 24*(1), 58. https://doi.org/10.1186/s12870-023-04491-5

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., … & Chanda, S. K. (2022). Therapeutic target database update 2022: Facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Research, 50*(D1), D1398–D1407. https://doi.org/10.1093/nar/gkab953

Zhou, Z.-L., Ling, T. C., Zhao, J.-L., Wang, X.-Z., Wang, L.-L., Li, L., … & Duan, Y.-W. (2025). Parallel loss of anthocyanins triggers the incipient sympatric speciation in an alpine ginger. *Plant Diversity, 47*(3), 429–439. https://doi.org/10.1016/j.pld.2025.03.004

Supplementary Tables

Supplementary Table 2.1. Raw Phenotypic Data for Association Mapping

Sample	Leaf Anthocyanin	Flower Anthocyanin	Stem Anthocyanin
EH-32-001	0	2	0
EH-32-002	0	0	0
EH-32-003	0	0	0
EH-32-004	0	0	0
EH-32-005	0	0	0
EH-32-006	0	1	0
EH-32-007	0	0	0
EH-32-008	0	0	0
EH-32-009	0	0	0
EH-32-010	0	0	0
EH-32-011	0	1	0
EH-32-012	0	0	0
EH-32-013	0	0	0
EH-32-014	0	0	0
EH-32-015	0	1	0
EH-32-016	0	0	0
EH-32-017	0	0	1
EH-32-018	0	0	0
EH-33-001	0	3	3
EH-33-002	3	2	3
EH-33-003	0	2	2
EH-33-004	1	3	2
EH-33-005	0	0	0
EH-33-006	2	2	2
EH-33-007	1	3	0
EH-33-008	1	2	2
EH-33-009	0	0	0
EH-33-010	1	3	0
EH-33-011	3	0	0
EH-33-012	1	3	2
EH-33-013	0	2	3
EH-33-014	0	0	2
EH-33-015	2	3	2
EH-33-016	0	3	3
EH-33-017	3	3	2
EH-33-018	0	3	3

EH-32-019	0	2	0
EH-32-020	0	0	0
EH-32-021	0	0	0
EH-32-022	0	0	1
EH-32-023	0	0	0
EH-32-024	0	0	1
EH-32-025	0	0	0
EH-32-026	0	0	0
EH-32-027	0	0	0
EH-32-028	0	0	0
EH-32-029	0	0	0
EH-32-030	0	0	0
EH-32-031	0	2	0
EH-32-032	0	0	0
EH-32-033	0	0	0
EH-32-034	0	1	0
EH-32-035	0	0	0
EH-32-036	0	1	0
EH-33-020	2	3	3
EH-33-021	0	0	0
EH-33-022	2	3	2
EH-33-023	1	2	2
EH-33-024	1	2	2
EH-33-025	2	3	3
EH-33-027	1	3	1
EH-33-028	0	3	3
EH-33-029	3	3	3
EH-33-030	0	3	3
EH-33-031	3	3	2
EH-33-033	2	2	3
EH-33-034	0	3	2
EH-33-035	3	3	2
EH-33-036	2	2	3
EH-32-037	0	0	0
EH-32-038	0	0	0
EH-32-039	0	0	0
EH-32-040	0	0	0
EH-32-041	0	0	0
EH-32-042	0	0	0
EH-32-043	0	0	0
EH-32-044	0	0	0
EH-32-045	0	0	2
EH-32-046	0	3	0
EH-32-047	0	0	0
EH-32-048	1	0	0

EH-32-049	0	0	0
EH-32-050	0	0	1
EH-32-051	0	0	0
EH-32-052	0	0	2
EH-32-053	0	0	1
EH-33-037	0	2	2
EH-33-038	3	3	2
EH-33-039	2	3	1
EH-33-041	1	3	1
EH-33-042	0	3	2
EH-33-043	2	3	1
EH-33-044	1	3	3
EH-33-045	2	2	0
EH-33-046	3	3	3
EH-33-048	1	2	2
EH-33-049	1	2	0
EH-33-050	2	3	3
EH-33-053	1	2	3
EH-32-054	0	0	0
EH-32-055	0	0	1
EH-32-056	0	2	2
EH-32-057	0	0	0
EH-32-058	0	0	0
EH-32-059	0	0	0
EH-32-060	0	0	0
EH-32-061	0	0	0
EH-32-062	0	0	0
EH-32-063	0	2	0
EH-32-064	0	0	0
EH-32-065	0	0	0
EH-32-066	0	0	0
EH-32-067	0	0	2
EH-32-068	0	0	2
EH-32-069	0	0	0
EH-32-070	0	0	1
EH-33-054	1	3	1
EH-33-055	2	2	2
EH-33-056	3	3	2
EH-33-057	2	2	3
EH-33-058	2	2	2
EH-33-059	1	2	1
EH-33-062	1	2	1
EH-33-063	0	0	1
EH-33-064	2	3	3
EH-33-066	1	1	3

EH-33-068	3	2	2
EH-33-069	0	3	3
EH-33-070	2	3	2
EH-32-071	0	0	0
EH-32-072	0	0	0
EH-32-073	0	0	0
EH-32-074	0	0	0
EH-32-075	0	0	0
EH-32-076	0	0	0
EH-32-077	0	0	0
EH-32-078	0	0	0
EH-32-079	0	0	1
EH-32-080	0	0	0
EH-32-081	0	0	1
EH-32-082	0	0	0
EH-32-083	0	0	0
EH-32-084	0	0	0
EH-32-085	0	0	0
EH-32-086	0	0	0
EH-32-087	0	0	0
EH-32-088	0	0	0
EH-33-071	2	2	2
EH-33-073	0	3	0
EH-33-074	2	3	0
EH-33-077	2	3	3
EH-33-078	3	3	3
EH-33-079	2	2	1
EH-33-080	2	2	2
EH-33-081	3	3	2
EH-33-083	1	2	2
EH-33-087	2	0	2
EH-33-088	2	2	3
EH-32-089	0	0	0
EH-32-090	0	1	0
EH-32-091	0	0	0
EH-32-092	0	0	0
EH-32-093	0	2	0
EH-32-094	0	0	0
EH-32-095	0	0	0
EH-32-096	0	2	0
EH-32-097	0	0	0
EH-32-098	0	0	0
EH-32-099	0	0	0
EH-32-100	0	0	0
EH-32-101	0	0	0

EH-32-102	0	0	0
EH-32-103	0	2	0
EH-32-104	0	0	0
EH-32-105	0	0	0
EH-32-106	0	0	0
EH-33-089	2	3	1
EH-33-091	2	2	1
EH-33-092	2	3	0
EH-33-093	2	2	2
EH-33-095	1	3	2
EH-33-096	2	2	2
EH-33-097	0	3	1
EH-33-098	1	2	2
EH-33-099	1	2	2
EH-33-100	1	3	1
EH-33-101	2	3	1
EH-33-102	2	2	3
EH-33-103	2	3	0
EH-33-104	1	2	2
EH-32-107	0	0	0
EH-32-108	0	0	0
EH-32-109	0	0	0
EH-32-110	0	0	0
EH-32-111	0	0	0
EH-32-112	0	0	0
EH-32-113	0	0	0
EH-32-114	0	0	0
EH-32-115	0	0	0
EH-32-116	0	2	0
EH-32-117	0	1	0
EH-32-118	0	0	0
EH-32-119	0	0	0
EH-32-120	0	0	0
EH-32-121	0	3	0
EH-32-122	0	2	0
EH-32-123	0	2	0
EH-32-124	0	0	0
EH-33-107	1	3	2
EH-33-108	2	3	2
EH-33-109	3	3	3
EH-33-110	2	2	1
EH-33-111	3	3	3
EH-33-112	0	3	2
EH-33-113	1	2	2
EH-33-114	1	3	1

EH-33-115	1	2	2
EH-33-116	NA	NA	NA
EH-33-117	1	2	2
EH-33-118	3	3	0
EH-33-119	3	3	1
EH-33-120	2	3	2
EH-33-121	3	2	3
EH-33-122	2	3	1
EH-33-123	1	2	1
EH-33-124	2	1	2
EH-32-125	0	0	0
EH-32-126	0	2	0
EH-32-127	0	0	0
EH-32-128	0	0	0
EH-32-129	0	0	0
EH-32-130	0	1	0
EH-32-131	0	0	0
EH-32-132	0	1	0
EH-32-133	0	1	0
EH-32-134	0	0	0
EH-32-135	0	0	0
EH-32-136	0	2	0
EH-32-137	0	3	0
EH-32-138	0	0	0
EH-32-139	0	1	0
EH-32-140	0	2	0
EH-32-141	0	0	0
EH-32-142	0	0	0
EH-33-125	3	3	2
EH-33-126	1	2	2
EH-33-127	1	2	2
EH-33-128	3	2	2
EH-33-129	2	2	3
EH-33-130	1	2	2
EH-33-131	2	3	2
EH-33-132	1	2	2
EH-33-133	2	2	2
EH-33-134	1	3	2
EH-33-138	1	3	2
EH-33-139	1	2	2
EH-33-140	2	3	1
EH-33-141	1	3	3
EH-32-143	0	0	0
EH-32-144	0	0	0
EH-32-145	0	1	0

EH-32-146	0	1	0
EH-32-147	0	0	0
EH-32-148	0	0	0
EH-32-149	0	0	0
EH-32-150	0	0	0
EH-32-151	0	0	0
EH-32-152	0	1	0
EH-32-153	0	0	0
EH-32-154	0	0	0
EH-32-155	0	0	0
EH-32-156	0	0	0
EH-32-157	0	0	0
EH-32-158	0	0	0
EH-32-159	0	1	0
EH-32-160	0	1	0
EH-33-143	1	2	1
EH-33-144	1	3	1
EH-33-145	0	2	2
EH-33-146	2	3	2
EH-33-147	1	3	2
EH-33-148	1	2	2
EH-33-150	1	3	3
EH-33-151	1	2	2
EH-33-152	2	1	2
EH-33-154	3	3	3
EH-33-155	0	3	3
EH-33-156	2	3	2
EH-33-157	2	3	2
EH-33-159	3	3	2
EH-32-161	0	0	0
EH-32-162	0	0	0
EH-32-163	0	2	0
EH-32-164	0	0	0
EH-32-165	0	2	0
EH-32-166	0	0	0
EH-32-167	0	1	0
EH-32-168	0	0	0
EH-32-169	0	0	0
EH-32-170	0	2	0
EH-32-171	0	0	0
EH-32-172	0	0	0
EH-32-173	0	0	0
EH-32-174	0	0	0
EH-32-175	0	0	0
EH-32-176	0	0	0

EH-32-177	0	1	0
EH-33-161	2	1	3
EH-33-163	2	2	3
EH-33-166	2	2	3
EH-33-167	1	1	2
EH-33-171	2	2	1
EH-33-173	3	2	3
EH-33-175	2	2	3
EH-33-176	1	3	0
EH-33-178	2	3	1
EH-32-179	0	3	0
EH-32-180	0	2	0
EH-32-181	0	0	2
EH-32-182	0	0	0
EH-32-183	0	0	0
EH-32-184	0	1	0
EH-32-185	0	1	0
EH-32-186	0	0	0
EH-32-187	0	0	0
EH-32-188	0	0	0
EH-32-189	0	2	0
EH-32-190	0	2	0
EH-32-191	0	0	0
EH-32-192	0	2	0
EH-32-193	0	2	0
EH-32-194	0	0	0
EH-32-195	0	0	0
EH-33-180	1	2	2
EH-33-181	2	3	3
EH-33-182	3	3	2
EH-33-185	2	2	3
EH-33-186	1	3	3
EH-33-187	0	2	2
EH-33-188	0	3	1
EH-33-189	2	3	2
EH-33-190	1	3	2
EH-33-191	1	2	1
EH-33-193	2	3	3
EH-33-194	1	3	2
EH-33-195	1	3	2
EH-32-196	0	0	0
EH-32-197	0	0	0
EH-32-198	0	0	0
EH-32-199	0	2	2
EH-33-197	1	2	2

EH-33-199	1	3	2
CW1	1	1	1
CW2	0	0	1
CW3	0	0	1
EHPar1	3	3	3
EHPar2	0	0	0
EHPar3	3	3	0
EHPar5	3	3	3
EH-33-047	1	3	1
EH-33-106	0	3	2
EH-33-153	3	3	2

Supplementary Table 2.2. Primer Sequences for *CsMYB78* (LOC115695758) Amplification

Primer Name	**Primer Sequence**
Cs10AnthfullF	CGATGAGGTTCGGGAGGATG
Cs10AnthfullR	TCTTCTTCGACCAAACCGGG

Supplementary Table 2.3. RNA Yields for Indigo Samples

Sample	**Picogreen ng/uL**
Indigo-1-Purple	9.7
Indigo-2-Purple	14.4
Indigo-3-Purple	2.4
Indigo-4-Purple	4.8
Indigo-5-Purple	10.1
Indigo-9-Green	9.4
Indigo-11-Green	1.6

Indigo-12-Green	2.4

Supplementary Table 2.4. Significant MLM SNPs for Flower Anthocyanin

SNP	Chr	Position	P.value	PhenoVarExplained(%)
SNC_044375.1_35831916	2	35831916	1.79E-07	0.00
SNC_044375.1_73349016	2	73349016	1.28E-06	0.00
SNC_044375.1_74052320	2	74052320	2.62E-08	0.00
SNC_044375.1_96056289	2	96056289	1.45E-08	0.00
SNC_044375.1_96056315	2	96056315	1.45E-08	0.00
SNC_044372.1_23623	3	23623	3.37E-07	0.00
SNC_044372.1_35269	3	35269	7.45E-08	0.00
SNC_044372.1_35305	3	35305	7.45E-08	0.00
SNC_044372.1_35472	3	35472	2.08E-06	0.00
SNC_044372.1_35507	3	35507	2.08E-06	0.00
SNC_044372.1_476699	3	476699	2.08E-06	0.00
SNC_044374.1_6600329	5	6600329	4.74E-08	0.00
SNC_044374.1_71299408	5	71299408	3.93E-07	0.00
SNC_044374.1_71299409	5	71299409	3.93E-07	0.00
SNC_044374.1_71299412	5	71299412	3.93E-07	0.00
SNC_044374.1_71299413	5	71299413	3.93E-07	0.00
SNC_044374.1_71299417	5	71299417	3.93E-07	0.00
SNC_044374.1_71299454	5	71299454	3.93E-07	0.00
SNC_044374.1_71299466	5	71299466	3.93E-07	0.00
SNC_044377.1_7127426	6	7127426	9.69E-09	0.00
SNC_044377.1_7141542	6	7141542	2.25E-09	0.00
SNC_044377.1_7141563	6	7141563	2.25E-09	0.00
SNC_044377.1_7240309	6	7240309	2.46E-07	0.00
SNC_044377.1_7240320	6	7240320	1.23E-07	0.00
SNC_044377.1_7240324	6	7240324	2.46E-07	0.00
SNC_044377.1_7240325	6	7240325	4.51E-08	0.00
SNC_044377.1_7240343	6	7240343	1.60E-07	0.00
SNC_044377.1_7240410	6	7240410	5.43E-08	0.00
SNC_044377.1_7243758	6	7243758	9.00E-07	0.00
SNC_044377.1_7513278	6	7513278	2.85E-07	0.00
SNC_044377.1_8426020	6	8426020	1.02E-07	0.00
SNC_044377.1_11173368	6	11173368	3.82E-06	0.00
SNC_044377.1_66752055	6	66752055	7.66E-09	0.00
SNC_044377.1_68129206	6	68129206	1.57E-07	0.00
SNC_044377.1_69139015	6	69139015	1.02E-07	0.00
SNC_044377.1_69180792	6	69180792	3.26E-07	0.00

SNC_044377.1_70265802	6	70265802	1.57E-07	0.00
SNC_044377.1_70265860	6	70265860	1.57E-07	0.00
SNC_044377.1_70265862	6	70265862	1.57E-07	0.00
SNC_044377.1_70265903	6	70265903	1.57E-07	0.00
SNC_044377.1_71519544	6	71519544	1.57E-07	0.00
SNC_044377.1_72643737	6	72643737	1.50E-06	0.00
SNC_044377.1_72786284	6	72786284	2.03E-06	0.00
SNC_044377.1_72950220	6	72950220	7.66E-09	0.00
SNC_044377.1_72950244	6	72950244	8.48E-09	6.76
SNC_044377.1_72950247	6	72950247	8.48E-09	6.70
SNC_044377.1_72985627	6	72985627	3.65E-07	0.00
SNC_044377.1_73091192	6	73091192	1.02E-08	12.69
SNC_044377.1_73091193	6	73091193	1.02E-08	32.82
SNC_044377.1_74169162	6	74169162	1.11E-06	0.00
SNC_044377.1_74169190	6	74169190	3.69E-08	0.00
SNC_044377.1_74180252	6	74180252	3.69E-08	0.00
SNC_044377.1_74180880	6	74180880	3.69E-08	0.00
SNC_044377.1_74180893	6	74180893	3.69E-08	0.00
SNC_044377.1_74180976	6	74180976	3.65E-07	0.00
SNC_044377.1_74180981	6	74180981	3.65E-07	0.00
SNC_044377.1_74181172	6	74181172	3.65E-07	0.00
SNC_044377.1_74539222	6	74539222	3.69E-08	0.00
SNC_044377.1_74539291	6	74539291	3.69E-08	0.00
SNC_044377.1_74893099	6	74893099	1.06E-06	0.00
SNC_044377.1_74893243	6	74893243	3.69E-08	0.00
SNC_044377.1_74893275	6	74893275	3.69E-08	0.00
SNC_044377.1_75779274	6	75779274	1.57E-07	0.00
SNC_044378.1_3405033	7	3405033	1.33E-06	8.34

Supplementary Table 2.5 Significant FarmCPU SNPs for Flower Anthocyanin

SNP	CHR	POS	P.VALUE	PHENOVAREXPLAINED(%)
SNC_044373.1_77029671	4	77029671	1.01E-08	6.51
SNC_044377.1_66752055	6	66752055	1.08E-14	72.19

Supplementary Table 2.6 Significant MLM SNPs for Leaf Anthocyanin

SNP	Chr	Position	P.value	PhenoVarExplained(%)
SNC_044371.1_85083607	1	85083607	3.28E-06	16.92
SNC_044371.1_85083608	1	85083608	3.28E-06	5.74
SNC_044371.1_85327650	1	85327650	5.30E-08	10.91

SNC_044371.1_86474655	1	86474655	2.54E-06	0.51
SNC_044371.1_86968836	1	86968836	2.54E-06	0.05
SNC_044371.1_86999970	1	86999970	2.54E-06	0.00
SNC_044371.1_86999979	1	86999979	2.54E-06	0.07
SNC_044371.1_87000019	1	87000019	2.54E-06	0.17
SNC_044371.1_87241272	1	87241272	2.54E-06	0.14
SNC_044371.1_87241511	1	87241511	2.07E-06	0.00
SNC_044371.1_87241605	1	87241605	2.54E-06	0.24
SNC_044371.1_87555061	1	87555061	2.54E-06	0.50
SNC_044371.1_87750430	1	87750430	2.07E-06	0.00
SNC_044371.1_87750522	1	87750522	2.54E-06	0.02
SNC_044371.1_87750547	1	87750547	2.54E-06	0.11
SNC_044371.1_87853555	1	87853555	2.54E-06	2.26
SNC_044371.1_87911523	1	87911523	1.20E-07	26.31
SNC_044371.1_87918190	1	87918190	2.54E-06	0.60
SNC_044371.1_94777564	1	94777564	7.81E-07	0.82
SNC_044375.1_74052320	2	74052320	6.10E-07	0.00
SNC_044375.1_96056289	2	96056289	1.33E-07	0.00
SNC_044375.1_96056315	2	96056315	1.33E-07	0.00
SNC_044374.1_6600329	5	6600329	2.71E-06	0.00
SNC_044377.1_7127426	6	7127426	1.00E-06	0.00
SNC_044377.1_7141542	6	7141542	3.74E-08	0.00
SNC_044377.1_7141563	6	7141563	3.74E-08	0.00
SNC_044377.1_7240320	6	7240320	3.03E-06	0.00
SNC_044377.1_7240325	6	7240325	6.33E-07	0.00
SNC_044377.1_7513278	6	7513278	5.24E-07	0.00
SNC_044377.1_66752055	6	66752055	8.66E-08	0.83
SNC_044377.1_72950220	6	72950220	8.66E-08	0.00
SNC_044377.1_72950244	6	72950244	2.72E-07	0.00
SNC_044377.1_72950247	6	72950247	2.72E-07	0.97
SNC_044377.1_73091192	6	73091192	2.70E-06	1.95
SNC_044377.1_73091193	6	73091193	2.70E-06	0.39
SNC_044377.1_74169162	6	74169162	1.05E-06	15.10
SNC_044377.1_74169190	6	74169190	2.46E-06	0.00
SNC_044377.1_74180252	6	74180252	2.46E-06	0.00
SNC_044377.1_74180880	6	74180880	2.46E-06	0.00
SNC_044377.1_74180893	6	74180893	2.46E-06	0.00
SNC_044377.1_74539222	6	74539222	2.46E-06	0.00
SNC_044377.1_74539291	6	74539291	2.46E-06	0.00
SNC_044377.1_74571488	6	74571488	4.17E-06	4.73
SNC_044377.1_74893099	6	74893099	2.32E-06	0.00
SNC_044377.1_74893243	6	74893243	2.46E-06	0.00
SNC_044377.1_74893275	6	74893275	2.46E-06	0.00

Supplementary Table 2.7 Significant FarmCPU SNPs for Leaf Anthocyanin

SNP	CHR	POS	P.VALUE	PHENOTYPE VARIANCE EXPLAINED(%)
SNC_044371.1_85266373	1	85266373	3.15E-23	0.18
SNC_044371.1_85282085	1	85282085	2.38E-09	0.00
SNC_044371.1_85282094	1	85282094	2.38E-09	0.00
SNC_044371.1_85282102	1	85282102	2.38E-09	0.00
SNC_044371.1_85282141	1	85282141	8.10E-07	1.54
SNC_044371.1_85282181	1	85282181	8.10E-07	0.00
SNC_044371.1_85327650	1	85327650	2.81E-16	20.83
SNC_044371.1_85327651	1	85327651	1.17E-10	0.00
SNC_044371.1_85350805	1	85350805	1.25E-11	0.00
SNC_044371.1_85361946	1	85361946	2.31E-09	0.00
SNC_044371.1_85361955	1	85361955	2.31E-09	0.00
SNC_044371.1_85362561	1	85362561	3.03E-07	0.00
SNC_044371.1_85417293	1	85417293	2.09E-09	0.00
SNC_044377.1_7141542	6	7141542	2.57E-11	43.87
SNC_044377.1_7141563	6	7141563	3.28E-24	2.04
SNC_044379.1_43821206	8	43821206	4.13E-08	5.88
SNC_044379.1_51441251	8	51441251	4.24E-06	8.17
SNC_044376.1_56362569	9	56362569	1.83E-06	0.62

Supplementary Table 2.8. Significant MLM SNPs for Stem Anthocyanin.

SNP	Chr	Pos	P.value	PhenoVarExplained(%)
SNC_044375.1_23166472	2	23166472	6.87E-10	0.00
SNC_044375.1_23166478	2	23166478	6.87E-10	0.00
SNC_044375.1_35831916	2	35831916	8.17E-08	0.00
SNC_044375.1_40139426	2	40139426	1.07E-06	4.23
SNC_044375.1_49088533	2	49088533	1.25E-08	0.00
SNC_044375.1_49088536	2	49088536	1.25E-08	0.00
SNC_044375.1_49088538	2	49088538	1.25E-08	0.00
SNC_044375.1_49088539	2	49088539	1.25E-08	0.00
SNC_044375.1_49088540	2	49088540	1.25E-08	0.00
SNC_044375.1_67571037	2	67571037	1.25E-08	0.00
SNC_044375.1_67571039	2	67571039	1.25E-08	0.00
SNC_044375.1_67756983	2	67756983	1.53E-06	2.42

SNC_044375.1_73349016	2	73349016	9.02E-07	0.00
SNC_044375.1_73362548	2	73362548	8.91E-07	3.06
SNC_044375.1_74052320	2	74052320	7.91E-08	0.00
SNC_044374.1_6600329	5	6600329	1.03E-07	0.00
SNC_044374.1_71299408	5	71299408	3.63E-10	3.16
SNC_044374.1_71299409	5	71299409	3.63E-10	18.70
SNC_044374.1_71299412	5	71299412	3.63E-10	13.09
SNC_044374.1_71299413	5	71299413	3.63E-10	13.44
SNC_044374.1_71299417	5	71299417	3.63E-10	3.26
SNC_044374.1_71299454	5	71299454	3.63E-10	2.41
SNC_044374.1_71299466	5	71299466	3.63E-10	2.78
SNC_044377.1_66752055	6	66752055	3.82E-08	0.00
SNC_044377.1_68129206	6	68129206	4.41E-07	0.00
SNC_044377.1_69180792	6	69180792	9.90E-09	0.00
SNC_044377.1_70265802	6	70265802	4.41E-07	0.00
SNC_044377.1_70265860	6	70265860	4.41E-07	0.00
SNC_044377.1_70265862	6	70265862	4.41E-07	0.00
SNC_044377.1_70265903	6	70265903	4.41E-07	0.00
SNC_044377.1_70341718	6	70341718	2.81E-06	8.14
SNC_044377.1_71519544	6	71519544	4.41E-07	0.00
SNC_044377.1_72643737	6	72643737	7.40E-07	0.00
SNC_044377.1_72786284	6	72786284	5.81E-11	6.59
SNC_044377.1_72950220	6	72950220	3.82E-08	0.00
SNC_044377.1_72950244	6	72950244	1.25E-06	0.00
SNC_044377.1_72950247	6	72950247	1.25E-06	0.00
SNC_044377.1_72985627	6	72985627	8.65E-09	0.00
SNC_044377.1_73091192	6	73091192	1.41E-07	0.00
SNC_044377.1_73091193	6	73091193	1.41E-07	0.00
SNC_044377.1_74169190	6	74169190	6.18E-09	0.00
SNC_044377.1_74180252	6	74180252	6.18E-09	0.00
SNC_044377.1_74180880	6	74180880	6.18E-09	0.00
SNC_044377.1_74180893	6	74180893	6.18E-09	0.00
SNC_044377.1_74180976	6	74180976	8.65E-09	0.03
SNC_044377.1_74180981	6	74180981	8.65E-09	0.00
SNC_044377.1_74181172	6	74181172	8.65E-09	4.42
SNC_044377.1_74319724	6	74319724	1.25E-08	0.00
SNC_044377.1_74539222	6	74539222	6.18E-09	0.00
SNC_044377.1_74539291	6	74539291	6.18E-09	0.00
SNC_044377.1_74893099	6	74893099	5.44E-08	0.00
SNC_044377.1_74893243	6	74893243	6.18E-09	0.00
SNC_044377.1_74893275	6	74893275	6.18E-09	0.00
SNC_044377.1_75779274	6	75779274	4.41E-07	0.00
SNC_044378.1_3405033	7	3405033	7.42E-07	1.05

Supplementary Table 2.9. Significant FarmCPU SNPs for Stem Anthocyanin

SNP	CHR	POS	P.VALUE	PHENOVAREXPLAINED(%)
SNC_044371.1_94777564	1	94777564	2.92E-07	1.59
SNC_044374.1_71299408	5	71299408	2.94E-13	77.92

Supplementary Table 2.10. Candidate Genes for Chr1 [85083607-87918190]. Candidate genes sharing homology with known regulators of anthocyanin biosynthesis are in bold.

LOC#	Start	Stop	Transcript description
115708119	85274322	85279257	probable sugar phosphate/phosphate translocator At3g14410
115706313	85280703	85284688	tricin synthase 1
115706315	85284784	85287490	uncharacterized LOC115706315
115706317	85289557	85295843	uncharacterized LOC115706317
115706319	85296784	85297840	GCN5-related N-acetyltransferase 8
115706318	85301483	85305851	NAC domain-containing protein 21/22
115706320	85325465	85333632	uncharacterized LOC115706320
115706321	85330587	85337075	protein At-4/1
115706323	85337349	85339136	large ribosomal subunit protein eL36y
115706326	85349225	85356634	translation initiation factor IF-2, chloroplastic
115706328	85361080	85363460	UDP-glucuronate 4-epimerase 3
115706329	85367854	85371434	protein NSP-INTERACTING KINASE 2
115706951	85382295	85383474	uncharacterized LOC115706951
115706807	85388939	85390181	uncharacterized LOC115706807
115706805	85396654	85400336	uncharacterized LOC115706805
115706335	85408485	85412145	probable protein S-acyltransferase 14
115706336	85416358	85418122	vestitone reductase
115706337	85419304	85421998	translation machinery-associated protein 22
115704883	85425415	85427163	protein BRI1-5 ENHANCED 1
115705009	85433655	85439105	protein transport Sec1a
115706338	85440634	85445485	uncharacterized LOC115706338
115706339	85448032	85452303	peptidyl-prolyl cis-trans isomerase FKBP43
115706342	85453391	85455406	uncharacterized LOC115706342
115706340	85455869	85458055	proteasome subunit beta type-1
115706509	85461277	85463263	uncharacterized LOC115706509
115708197	85468768	85471310	protein cornichon homolog 4
115706343	85477989	85482393	AT-hook motif nuclear-localized protein 1
115708218	85485343	85490797	65-kDa microtubule-associated protein 3
115706344	85494438	85496950	uncharacterized LOC115706344

115706345	85498787	85501718	LOB domain-containing protein 19
115705047	85542498	85545371	LOB domain-containing protein 30
115708190	85562161	85564739	mediator of RNA polymerase II transcription subunit 18
115703748	85576405	85578333	uncharacterized LOC115703748
115706353	85578553	85579636	dirigent protein 11
115706352	85579789	85582394	probable hexosyltransferase MUCI70
115703864	85584794	85589123	probable E3 ubiquitin-protein ligase ARI2
115706357	85598526	85606019	brefeldin A-inhibited guanine nucleotide-exchange protein 2
115706358	85606319	85607720	probable calcium-binding protein CML23
115704801	85618102	85619908	AT-hook motif nuclear-localized protein 22
115706359	85637562	85644674	kinesin-like protein KIN-UB
115706363	85656923	85660042	4-hydroxy-tetrahydrodipicolinate synthase, chloroplastic
115704474	85676682	85679040	inositol transporter 4-like
115706365	85682039	85687639	uncharacterized LOC115706365
115706367	85708962	85711565	peptidyl-prolyl cis-trans isomerase CYP18-1
115706366	85713470	85717647	uncharacterized LOC115706366
115706370	85725031	85726656	fasciclin-like arabinogalactan protein 10
115706371	85751891	85754670	growth-regulating factor 9
115706373	85755071	85762835	serine/threonine-protein kinase Aurora-3
115707213	85765138	85766998	uncharacterized LOC115707213
115706374	85768477	85773523	probable hexosyltransferase MUCI70
115706379	85806780	85812951	probable serine/threonine-protein kinase At1g54610
115706381	85826664	85832332	protein SPA1-RELATED 3
115706382	85833176	85836454	beta-glucuronosyltransferase GlcAT14A
115706384	85847907	85849097	desiccation protectant protein Lea14 homolog
115706613	85851548	85854538	probable WRKY transcription factor 56
115708268	85855692	85858825	large ribosomal subunit protein mL46
115707045	85860039	85863466	uncharacterized LOC115707045
115707044	85863832	85868396	uncharacterized LOC115707044
115707233	85883363	85885902	3-methyl-2-oxobutanoate hydroxymethyltransferase 1, mitochondrial
115706386	85889850	85893155	pentatricopeptide repeat-containing protein At3g61520, mitochondrial
115706387	85894074	85895969	WRKY transcription factor 22
115706389	85906691	85917113	exocyst complex component SEC10b
115707614	85925108	85928975	uncharacterized LOC115707614
115704849	85952788	85956005	1-aminocyclopropane-1-carboxylate synthase
115706403	85972864	85975442	heavy metal-associated isoprenylated plant protein 39
115706406	86005411	86009723	C-terminal binding protein AN
115706412	86027063	86030534	uncharacterized LOC115706412
115706411	86041164	86045236	uncharacterized LOC115706411
115706410	86047300	86054807	uncharacterized LOC115706410
115706414	86058617	86061046	large ribosomal subunit protein uL5c

115708206	86061576	86064980	sphingoid long-chain bases kinase 2, mitochondrial
115708281	86066298	86068737	uncharacterized LOC115708281
115706415	86069745	86075917	CAAX prenyl protease 1 homolog
115706416	86079405	86082414	probable serine/threonine-protein kinase At1g01540
115706417	86094607	86097168	protein BPS1, chloroplastic
115706422	86112670	86117672	mitogen-activated protein kinase homolog MMK2
115706421	86118004	86121398	probable polygalacturonase
115706423	86124522	86128923	mechanosensitive ion channel protein 1, mitochondrial
115706425	86141795	86146438	uncharacterized LOC115706425
115706426	86147068	86148952	probable receptor-like protein kinase At5g24010
115707468	86153517	86159403	ferric reduction oxidase 2
115706430	86160024	86169739	uncharacterized LOC115706430
115708164	86180797	86181713	brassinosteroid-responsive RING protein 1
115704688	86201472	86203529	uncharacterized LOC115704688
115706451	86224338	86227693	syntaxin-71
115706449	86230696	86238015	uncharacterized LOC115706449
115706450	86237977	86240962	calmodulin-binding receptor-like cytoplasmic kinase 2
115704478	86244653	86246431	pentatricopeptide repeat-containing protein At2g46050, mitochondrial
115706456	86265773	86268570	probable E3 ubiquitin-protein ligase RHB1A
115706457	86273279	86276333	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2
115706459	86284958	86288314	eukaryotic translation initiation factor NCBP
115704479	86288314	86294569	protein DETOXIFICATION 34
115706466	86332352	86339719	AT-rich interactive domain-containing protein 1
115706472	86366105	86369067	ubiquitin-conjugating enzyme E2-23 kDa
115706471	86370780	86382516	ATP-dependent helicase BRM
115707895	86386084	86389732	bifunctional L-3-cyanoalanine synthase/cysteine synthase 1, mitochondrial
115706476	86419422	86420465	uncharacterized LOC115706476
115706475	86420505	86423221	uncharacterized LOC115706475
115706477	86430821	86433476	uncharacterized LOC115706477
115708201	86433653	86437539	origin of replication complex subunit 5
115706480	86473049	86475812	cytochrome P450 86A8
115706481	86476840	86481993	ascorbate transporter, chloroplastic
115706483	86492368	86496100	probable glycerol-3-phosphate acyltransferase 8
115706485	86505820	86508159	aquaporin PIP1-3
115706486	86513103	86520634	signal peptide peptidase-like 4
115706490	86528407	86535214	U-box domain-containing protein 33
115706489	86539074	86542278	pentatricopeptide repeat-containing protein At1g71210, mitochondrial
115703644	86548552	86553706	putative recombination initiation defects 3
115706505	86568529	86574173	uncharacterized LOC115706505
115707936	86575860	86579677	uncharacterized LOC115707936
115706507	86583450	86596088	mediator of RNA polymerase II transcription subunit 12

115706508	86599473	86601197	chlorophyll a-b binding protein CP26, chloroplastic
115708155	86602266	86605296	rop guanine nucleotide exchange factor 3
115704721	86606782	86608044	elicitor-responsive protein 3
115704916	86610811	86612185	gibberellin 20-oxidase-like protein
115706692	**86615985**	**86618426**	**transcription factor GLABRA 3**
115706512	86629385	86633020	beta-amylase 2, chloroplastic
115706513	86633239	86633937	acetyltransferase At1g77540
115706510	86634034	86640638	beta-amylase 7
115706514	86640804	86645143	uncharacterized LOC115706514
115706516	86653068	86658994	acyl-CoA hydrolase 2
115706517	86659701	86662034	F-box/kelch-repeat protein SKIP4
115706518	86662273	86664072	uncharacterized LOC115706518
115708173	86665903	86666469	kiwellin-1-like
115704482	86667649	86668290	kiwellin-1-like
115704483	86668878	86672135	uncharacterized LOC115704483
115704485	86676693	86677268	kiwellin-1-like
115706526	86685446	86688025	voltage-dependent chloride channel 1, chloroplastic
115706536	86690027	86691498	uncharacterized LOC115706536
115706533	86692562	86694532	uncharacterized LOC115706533
115706523	86703050	86708902	serine/threonine-protein kinase BSK3
115706535	86711207	86712188	uncharacterized LOC115706535
115706527	86734540	86739619	AT-hook motif nuclear-localized protein 11
115706529	86753151	86755243	uncharacterized LOC115706529
115706532	86756040	86756606	kiwellin-1-like
115706534	86765581	86769238	uncharacterized LOC115706534
115706520	86770178	86775061	FT-interacting protein 3
115706537	86790845	86793888	uncharacterized LOC115706537
115706525	86796020	86799559	tRNA-specific adenosine deaminase TAD1
115706531	86814644	86817043	remorin
115706528	**86818386**	**86821154**	**transcription factor MYB17**
115706521	86833961	86839477	eukaryotic translation initiation factor 3 subunit A
115706522	86844072	86849042	DEAD-box ATP-dependent RNA helicase 8
115706539	86856400	86866459	uncharacterized LOC115706539
115706541	86867604	86869416	LIM domain-containing protein PLIM2b
115706542	86870098	86881597	K(+) efflux antiporter 2, chloroplastic
115706545	86882270	86884899	bifunctional protein FolD 4, chloroplastic
115706544	86885180	86888781	putative threonine aspartase
115706546	86888892	86893606	phosphomannomutase
115704489	86894412	86902834	(+)-neomenthol dehydrogenase
115706549	86903355	86905764	(+)-neomenthol dehydrogenase-like
115704490	86909820	86911453	(+)-neomenthol dehydrogenase-like
115704491	86912399	86918357	uncharacterized LOC115704491
115706548	86919974	86924618	replication factor C subunit 2
115704492	86924835	86930167	short-chain dehydrogenase/reductase 2b

115706551	86935257	86942567	NAD-dependent malic enzyme 59 kDa isoform, mitochondrial
115707120	86944571	86947468	uncharacterized LOC115707120
115703592	86948923	86952846	uncharacterized LOC115703592
115707599	86962839	86963589	disease resistance protein RGA5-like
115706832	86968964	86970034	enoyl-CoA delta isomerase 2, peroxisomal
115707649	86976846	86981064	uncharacterized LOC115707649
115706653	86981838	86986186	digalactosyldiacylglycerol synthase 2, chloroplastic
115707423	86992611	86993868	tubulin-folding cofactor A
115706558	86996224	86997693	uncharacterized LOC115706558
115706557	86997813	87002408	cell division protein FtsY homolog, chloroplastic
115705018	87005316	87006131	BON1-associated protein 2
115706662	87019119	87019963	BON1-associated protein 2
115705073	87029560	87030367	BON1-associated protein 2-like
115707351	87035816	87038419	E3 ubiquitin-protein ligase At4g11680
115706569	87042167	87046879	shaggy-related protein kinase theta
115706571	**87067285**	**87073409**	**homeobox-leucine zipper protein ANTHOCYANINLESS 2**
115706573	87100308	87104616	probable pectin methyltransferase QUA3
115706574	87105459	87109635	COP9 signalosome complex subunit 1
115706575	87115787	87120321	plant UBX domain-containing protein 9
115706590	87125235	87129722	probable methyltransferase PMT15
115707609	87166867	87168999	F-box protein PP2-A13
115707311	87172743	87177272	F-box protein At4g00755
115706592	87182197	87185454	peroxisomal membrane protein 11C
115704495	87215167	87218541	cytochrome P450 87A3-like
115703624	87216679	87236968	cytochrome P450 87A3
115706599	87240531	87244290	uncharacterized LOC115706599
115706600	87249127	87253258	serine/arginine-rich splicing factor RS41
115707190	87270987	87272165	ethylene-responsive transcription factor ERF012-like
115706610	87282933	87285528	RING-H2 finger protein ATL5
115707376	87286398	87288141	uncharacterized LOC115707376
115703728	87288426	87291624	protein-ribulosamine 3-kinase, chloroplastic
115706615	87292428	87293252	uncharacterized LOC115706615
115706614	87293274	87297912	ribosomal RNA small subunit methyltransferase, chloroplastic
115707878	87300719	87305232	flap endonuclease GEN-like 1
115706616	87311601	87314573	uncharacterized LOC115706616
115706619	87334574	87338390	UDP-galactose/UDP-glucose transporter 2
115706719	87340434	87342594	uncharacterized LOC115706719
115706718	87342726	87348065	protein-lysine N-methyltransferase EFM1
115708335	87352577	87358033	SPX domain-containing membrane protein At4g22990
115706621	87367062	87370756	ATPase GET3A
115708129	87371156	87373998	ubiquitin-related modifier 1 homolog 2
115706922	87392301	87395665	peroxisome biogenesis protein 16

115706624	87405986	87408041	transcription factor TCP9
115707336	87411046	87414453	uncharacterized LOC115707336
115708338	87452171	87458521	subtilisin-like protease SBT4.14
115706957	87460921	87464315	CRS2-associated factor 1, mitochondrial
115706628	87468740	87476045	lysophospholipid acyltransferase LPEAT2
115706939	87492519	87505383	MADS-box protein SOC1
115706706	87506102	87508488	probable fructokinase-7
115704294	87513236	87519439	agamous-like MADS-box protein MADS3
115706634	87544635	87550555	polygalacturonate 4-alpha-galacturonosyltransferase
115704439	87554396	87556217	gibberellin 20 oxidase 1-D
115707107	87569159	87571231	glyoxylate/hydroxypyruvate reductase HPR3
115704860	87574656	87577625	glyoxylate/hydroxypyruvate reductase HPR3
115706636	87579241	87583409	UTP:RNA uridylyltransferase 1
115706864	87588879	87589983	probable carboxylesterase 9
115704944	87592391	87593636	probable carboxylesterase 8
115708266	87596071	87597394	probable carboxylesterase 8
115706638	87613236	87615771	receptor-like serine/threonine-protein kinase At2g45590
115706639	87623227	87630922	calcium-dependent lipid-binding protein
115706645	87639443	87641441	early nodulin-like protein 18
115706644	87644909	87666615	BEACH domain-containing protein C2
115706646	87678649	87681102	9-cis-epoxycarotenoid dioxygenase NCED1, chloroplastic
115706661	87748592	87754817	polyadenylate-binding protein-interacting protein 9
115703622	87765644	87766109	VQ motif-containing protein 10
115706670	87776866	87778696	gibberellin 2-beta-dioxygenase 1
115706677	87827073	87829431	cationic amino acid transporter 8, vacuolar
115704839	87831209	87833331	cationic amino acid transporter 8, vacuolar-like
115706687	87852168	87859216	ubiquitin carboxyl-terminal hydrolase 15
115706688	87859573	87861189	uncharacterized LOC115706688
115706696	87900220	87912806	putative callose synthase 8
115706698	87912946	87914635	pentatricopeptide repeat-containing protein At3g14580, mitochondrial

Supplementary Table 2.11. Candidate Genes for Chr6 [66752055..75779274]. Candidate genes sharing homology with known regulators of anthocyanin biosynthesis are in bold.

LOC#	StartPosition	StopPosition	XMtranscript description
115694829	66761268	66765931	glycosyltransferase BC10-like
115725319	66811027	66814051	cationic peroxidase 2
115695397	66836448	66839484	probable inactive receptor kinase At1g48480

115724638	66882141	66885437	receptor-like protein EIX2
115695702	66935267	66939400	uncharacterized LOC115695702
115725673	66951243	66959313	uncharacterized LOC115725673
115695398	66968688	66969176	uncharacterized LOC115695398
115725034	67009596	67012292	uncharacterized LOC115725034
115725033	67012661	67025930	protein LNK1
115695401	67196647	67197335	uncharacterized LOC115695401
115694842	67397597	67398465	uncharacterized LOC115694842
115695403	67989713	68002079	2-oxoglutarate-dependent dioxygenase 19
115724899	68004699	68007026	uncharacterized LOC115724899
115724900	68007523	68011651	L-2-hydroxyglutarate dehydrogenase, mitochondrial
115695404	68029499	68057915	protein ZINC INDUCED FACILITATOR 1-like
115695738	68055326	68059845	lysine-specific demethylase JMJ32
115725455	68069786	68072803	rop guanine nucleotide exchange factor 1
115725126	68080240	68084074	serine/arginine-rich splicing factor SC35
115725229	68117610	68119730	protein ROOT HAIR DEFECTIVE 3-like
115725228	68121256	68128125	potassium transporter 7
115725227	68128240	68137507	copper-transporting ATPase PAA1, chloroplastic
115695405	68151043	68163905	uncharacterized LOC115695405
115695406	68189903	68191240	uncharacterized LOC115695406
115695407	68216746	68232939	uncharacterized LOC115695407
115695987	68233354	68235588	ubiquinol oxidase, mitochondrial
115724757	68241450	68245112	glucose-6-phosphate 1-dehydrogenase, chloroplastic
115725356	68316593	68329188	calmodulin-binding transcription activator 1
115724962	68388407	68393068	probable protein phosphatase 2C 9
115725191	68396816	68407737	outer envelope protein 64, mitochondrial
115724487	68410350	68412867	uncharacterized LOC115724487
115725647	68454990	68458248	uncharacterized LOC115725647
115695990	68471012	68476576	metacaspase-3
115724527	68480608	68486543	metacaspase-1
115695408	68548338	68551295	uncharacterized LOC115695408
115725390	68712697	68720198	ABC transporter G family member 1
115695410	68734288	68737280	uncharacterized LOC115695410
115725708	68738136	68739926	uncharacterized LOC115725708
115725347	68761753	68766663	polyadenylate-binding protein 2
115724690	68779048	68783548	phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplastic
115724901	68786583	68788122	protein EXORDIUM-like 2
115724528	68800393	68800601	uncharacterized LOC115724528
115724485	68920128	68922341	uncharacterized LOC115724485
115724467	68957205	68961623	7-deoxyloganetin glucosyltransferase
115695961	68991241	69005559	7-deoxyloganetin glucosyltransferase
115725269	69096388	69098643	14-3-3-like protein 16R
115695414	69108676	69110675	uncharacterized LOC115695414
115695415	69110913	69126663	probable DNA helicase MCM9

115695787	69138134	69140835	AT-hook motif nuclear-localized protein 15
115725723	69160188	69163615	xanthotoxin 5-hydroxylase CYP82C4
115695418	69176915	69178849	phytosulfokine receptor 1
115725188	69180200	69183405	dicarboxylate transporter 2.1, chloroplastic
115694681	69326880	69329341	receptor-like protein EIX2
115695423	69456820	69462097	rust resistance kinase Lr10-like
115695424	69479552	69480406	uncharacterized LOC115695424
115695425	69527796	69529052	uncharacterized LOC115695425
115695427	69587092	69588048	uncharacterized LOC115695427
115695429	69596891	69599143	uncharacterized LOC115695429
115695991	69654135	69657940	carbon catabolite repressor protein 4 homolog 4
115724886	69660437	69663308	uncharacterized LOC115724886
115695934	69663904	69667062	mitochondrial arginine transporter BAC1
115725035	69675931	69681191	bifunctional riboflavin biosynthesis protein RIBA 1, chloroplastic
115695433	69748821	69749240	classical arabinogalactan protein 1-like
115695648	69754187	69754880	classical arabinogalactan protein 1-like
115695650	69756607	69757207	classical arabinogalactan protein 1-like
115695434	69760703	69763693	classical arabinogalactan protein 1
115725598	69776530	69779803	uncharacterized LOC115725598
115695770	69785833	69786424	uncharacterized LOC115695770
115725493	69819479	69821254	RING-H2 finger protein ATL54
115725492	69820782	69827158	pentatricopeptide repeat-containing protein At5g62370
115725494	69830496	69894171	geranylgeranyl transferase type-2 subunit beta 1
115725495	69837063	69858834	uncharacterized LOC115725495
115724591	69896444	69901848	uncharacterized LOC115724591
115695435	69902233	69903136	uncharacterized mitochondrial protein AtMg00810-like
115725216	69906383	69916064	uncharacterized LOC115725216
115725217	69916115	69919020	pentatricopeptide repeat-containing protein At5g09450, mitochondrial
115725550	69946139	69951044	cytoplasmic tRNA 2-thiolation protein 1
115724791	69952427	69959726	uncharacterized LOC115724791
115695436	69977213	69982174	2-oxoglutarate-dependent dioxygenase 19
115725610	69991419	70006152	2-oxoglutarate-dependent dioxygenase 19
115724633	70128428	70129537	fatty acid desaturase 4-like 1, chloroplastic
115725303	70219519	70222610	root phototropism protein 3
115725621	70261236	70267944	F-box protein 7
115725562	70274036	70277686	uncharacterized LOC115725562
115725372	70329309	70345593	transcription initiation factor TFIID subunit 1
115724532	70349430	70352878	pentatricopeptide repeat-containing protein At3g18020
115695438	70385219	70390891	uncharacterized LOC115695438
115695624	70408917	70410816	peptidyl-prolyl cis-trans isomerase FKBP12

115724807	70513999	70514408	uncharacterized LOC115724807
115724798	70587569	70596547	uncharacterized LOC115724798
115724775	70600945	70606962	uncharacterized LOC115724775
115724711	70611101	70614466	uncharacterized LOC115724711
115724715	70618164	70620394	uncharacterized LOC115724715
115724662	70700933	70702769	uncharacterized LOC115724662
115725351	**70709522**	**70713911**	**transcription factor bHLH143**
115724728	70823856	70824953	mitochondrial uncoupling protein 4-like
115725729	70899243	70901224	probable CoA ligase CCL11
115724889	70904699	70910147	protein STRICTOSIDINE SYNTHASE-LIKE 13
115725659	70933687	70935667	probable CoA ligase CCL11
115695444	70967044	70970560	uncharacterized LOC115695444
115695955	71020209	71022341	probable CoA ligase CCL11
115695445	71025270	71028741	uncharacterized LOC115695445
115695447	71075872	71083460	disease resistance-like protein DSC1
115724552	71104303	71105762	protein PELPK1
115725343	71107800	71114828	glycerol-3-phosphate acyltransferase 9
115695989	71130670	71131975	protein PELPK1-like
115694782	71335111	71338476	chorismate mutase 2
115724788	71396020	71396534	auxin-induced protein 15A-like
115724784	71397555	71398151	auxin-induced protein X15-like
115724789	71397561	71404633	auxin-induced protein X15-like
115724783	71400100	71400839	uncharacterized LOC115724783
115695450	71405678	71406694	auxin-induced protein 15A-like
115724786	71409642	71410225	auxin-induced protein X15
115725753	71412537	71415446	cytoplasmic 60S subunit biogenesis factor REI1 homolog 2
115695451	71416802	71421564	uncharacterized LOC115695451
115695453	71491558	71492796	F-box protein At5g42460-like
115695620	71502001	71503755	uncharacterized LOC115695620
115725161	71512916	71514787	pentatricopeptide repeat-containing protein At1g07740, mitochondrial
115725167	71515159	71516895	auxin-responsive protein SAUR21
115725162	71520411	71523130	thioredoxin-like fold domain-containing protein MRL7L, chloroplastic
115725163	71530990	71558251	truncated transcription factor CAULIFLOWER A
115695457	71588840	71590019	uncharacterized LOC115695457
115725165	71611500	71612002	auxin-responsive protein SAUR21-like
115725166	71613348	71613917	auxin-responsive protein SAUR21
115725551	71614664	71615164	auxin-responsive protein SAUR24
115725552	71630611	71631154	auxin-induced protein X10A
115724699	71631711	71632258	auxin-responsive protein SAUR21-like
115724696	71651139	71655445	pentatricopeptide repeat-containing protein At1g03100, mitochondrial
115724700	71656444	71657009	auxin-induced protein 15A-like

115724698	71664999	71665698	auxin-responsive protein SAUR21-like
115724697	71667718	71668926	uncharacterized LOC115724697
115695980	71675846	71677809	pentatricopeptide repeat-containing protein At1g80550, mitochondrial
115724597	71717512	71720986	uncharacterized LOC115724597
115725707	71737131	71741152	uncharacterized LOC115725707
115695839	71773765	71775992	protein SICKLE
115695777	71779870	71784295	uncharacterized LOC115695777
115695778	71784313	71785202	uncharacterized LOC115695778
115695783	71785760	71795604	auxin-induced protein 15A-like
115695780	71791321	71791934	auxin-induced protein X15-like
115695779	71801505	71802118	auxin-induced protein 15A-like
115695781	71803039	71803733	auxin-responsive protein SAUR24-like
115695459	71804416	71804724	auxin-induced protein 15A-like
115725651	71805363	71805882	auxin-responsive protein SAUR24-like
115695460	71806541	71807830	F-box protein At3g07870-like
115694746	71813151	71816660	exosome complex component RRP45A
115724709	71828703	71829249	auxin-responsive protein SAUR21
115725226	71859666	71860304	auxin-induced protein 15A-like
115725225	71862688	71863519	auxin-responsive protein SAUR21
115725223	71878944	71883674	transcription initiation factor TFIID subunit 14b
115725224	71903731	71904337	protein SMALL AUXIN UP-REGULATED RNA 51
115724468	71983641	71984557	auxin-induced protein X10A
115694783	72018995	72024257	pentatricopeptide repeat-containing protein At2g22410, mitochondrial
115724499	72048052	72049355	probable ribosome-binding factor A, chloroplastic
115725056	72062074	72067755	CBL-interacting serine/threonine-protein kinase 24
115725181	72106595	72108179	uncharacterized LOC115725181
115725180	72108422	72110765	equilibrative nucleotide transporter 8
115724958	72194067	72202314	nuclear pore complex protein NUP1
115724567	72212734	72217384	axial regulator YABBY 5
115695462	72260987	72261466	uncharacterized LOC115695462
115725110	72444702	72447378	methyl-CpG-binding domain-containing protein 11
115695793	72448132	72450112	thylakoid lumenal 16.5 kDa protein, chloroplastic
115725360	72545638	72549377	glucan endo-1,3-beta-glucosidase 14
115694748	72584134	72585291	cucumber peeling cupredoxin
115725070	72642546	72645201	probable galacturonosyltransferase 9
115695925	72662454	72664093	uncharacterized LOC115695925
115725389	72697797	72699905	amidophosphoribosyltransferase, chloroplastic-like
115725437	72703074	72707704	tRNA-dihydrouridine(47) synthase [NAD(P)(+)]-like
115724995	72721237	72724982	bZIP transcription factor 29
115725653	72775739	72779775	protein BASIC PENTACYSTEINE4
115725076	72785278	72787757	uncharacterized LOC115725076
115725532	72789321	72795164	polyprenol reductase 2

115725577	72803979	72807087	NADP-dependent D-sorbitol-6-phosphate dehydrogenase
115725784	72809170	72816073	putative disease resistance protein RGA4
115725026	72833341	72835651	V-type proton ATPase 16 kDa proteolipid subunit
115725315	72837129	72842036	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic-like
115725364	72947694	72950724	arginine decarboxylase
115725516	72984250	72988289	uncharacterized LOC115725516
115695828	72989609	72993205	epimerase family protein SDR39U1 homolog, chloroplastic
115695821	72996104	72998519	uncharacterized LOC115695821
115724553	73002419	73004761	BURP domain protein USPL1
115724427	73006141	73008524	pentatricopeptide repeat-containing protein At4g21065
115695469	73012637	73013467	uncharacterized LOC115695469
115725150	73029451	73035487	kinesin-like protein KIN-7E
115725151	73044825	73048604	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9
115695470	73051719	73053679	uncharacterized LOC115695470
115724441	73053737	73054390	small ribosomal subunit protein bTHXm
115725553	73084475	73087604	B-box zinc finger protein 19
115724972	73089053	73091247	fructose-bisphosphate aldolase 1, chloroplastic
115694781	73111843	73116908	endoglucanase 24
115725324	73136890	73140901	zinc finger protein GIS2
115695471	73151308	73159701	zinc finger CCCH domain-containing protein 19
115695472	73169371	73171231	protein neprosin-like
115695473	73245797	73253515	protein neprosin-like
115695942	73265447	73267869	uncharacterized LOC115695942
115695236	73277136	73280314	uncharacterized LOC115695236
115725670	73290838	73292543	GATA transcription factor 5
115725686	73351466	73354764	SH3 domain-containing protein 2
115725458	73381692	73389082	squalene synthase 2
115724663	73551707	73554626	myosin-15
115725777	73773209	73776244	stellacyanin-like
115724870	73826394	73827550	cucumber peeling cupredoxin
115695476	73842267	73843742	putative pectinesterase 63
115695479	73904317	73908121	uncharacterized LOC115695479
115695480	73908224	73911506	uncharacterized LOC115695480
115694822	73954484	73956077	putative pectinesterase 63
115725094	73977700	73979619	cucumber peeling cupredoxin
115725502	74011708	74020632	uncharacterized LOC115725502
115694684	74038489	74050187	uncharacterized LOC115694684
115694685	74050049	74052667	actin-related protein 6
115725643	74122963	74124133	uncharacterized LOC115725643
115725642	74124217	74129644	uncharacterized LOC115725642

115695481	74154579	74155972	protein terminal ear1 homolog
115695482	74162590	74164680	protein MEI2-like 4
115725381	74179782	74182540	THO complex subunit 5B
115724961	74183062	74187019	THO complex subunit 5A
115695662	74190957	74191656	uncharacterized LOC115695662
115724609	74203142	74206397	uncharacterized LOC115724609
115725317	74210461	74212531	uncharacterized LOC115725317
115725316	74214327	74219901	uncharacterized LOC115725316
115724892	74243428	74245270	B-box zinc finger protein 20
115724477	74269267	74279067	kinesin-like protein KIN-7D, mitochondrial
115695959	74290929	74294021	D-ribulose kinase
115694841	74295905	74300957	probable inactive receptor kinase At3g08680
115724493	74305083	74309618	uncharacterized LOC115724493
115724492	74310641	74316062	protein CHROMATIN REMODELING 35
115724490	74317874	74323439	coatomer subunit alpha-1
115695728	74327376	74328293	uncharacterized LOC115695728
115725780	74382077	74387262	DNA replication licensing factor MCM4
115725475	74394285	74402671	V-type proton ATPase subunit a3
115724523	74405882	74407838	cysteine protease RD19A
115725121	74444592	74447803	purple acid phosphatase 2
115725120	74451694	74456755	uncharacterized LOC115725120
115695773	74465403	74467832	chromatin remodeling protein SHL
115695772	74473368	74479122	beta-galactosidase 5
115724939	74485701	74494683	uncharacterized LOC115724939
115724909	74533765	74539727	BEL1-like homeodomain protein 7
115724790	74542227	74547473	protein CHROMATIN REMODELING 35
115695484	74554261	74564751	putative B3 domain-containing protein At5g35780
115725299	74561972	74564120	probable rRNA-processing protein EBP2 homolog
115725108	74568690	74573951	SUMO-activating enzyme subunit 2
115725109	74574030	74575263	uncharacterized LOC115725109
115694758	74593941	74597045	probable receptor-like protein kinase At4g39110
115725232	74602074	74605498	26S proteasome regulatory subunit S10B homolog B
115695844	74607242	74611239	bifunctional phosphatase IMPL2, chloroplastic
115725633	74631494	74633064	uncharacterized LOC115725633
115724871	74637014	74638496	embryogenic cell protein 40-like
115724948	74644788	74648437	uncharacterized LOC115724948
115694742	74662399	74663843	protein CASPARIAN STRIP INTEGRITY FACTOR 1-like
115725699	74681563	74686496	chaperone protein dnaJ 10
115695824	74688289	74689967	glutathione transferase GST 23
115725298	74718980	74720457	bZIP transcription factor 11
115694831	74761481	74772156	transcription factor TFIIIB component B''
115694738	74795530	74801355	phosphatidylinositol/phosphatidylcholine transfer protein SFH8
115694739	74801453	74806868	phosphatidylinositol/phosphatidylcholine transfer protein SFH3

115724524	74887182	74889666	AAA-ATPase At3g50940
115724959	74891674	74893618	H/ACA ribonucleoprotein complex subunit 4-like
115725687	74896177	74901364	probable caffeine synthase MTL2
115724674	74942897	74964633	probable caffeine synthase MTL2
115724676	75011830	75027198	S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase 1-like
115724677	75051169	75054374	S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase 3-like
115724675	75059036	75073167	S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase 1-like
115724673	75105266	75138300	probable caffeine synthase MTL2
115695490	75187922	75189943	uncharacterized LOC115695490
115695491	75218088	75226356	uncharacterized LOC115695491
115695494	75261048	75261728	uncharacterized LOC115695494
115695497	75366562	75368457	uncharacterized LOC115695497
115695758	**75409856**	**75419127**	**transcription factor MYB1**
115695751	75482934	75486178	7-methylxanthosine synthase 1-like
115695753	75496650	75499393	7-methylxanthosine synthase 1
115695757	75531975	75541238	uncharacterized LOC115695757
115695762	75584352	75586439	uncharacterized LOC115695762
115695501	75616748	75618861	uncharacterized LOC115695501
115695755	75668949	75670955	G2/mitotic-specific cyclin C13-1
115695754	75672199	75674316	putative cyclin-A3-1
115695752	75674424	75677170	probable caffeine synthase MTL2
115695506	75682148	75683814	uncharacterized LOC115695506
115695759	75706808	75708110	uncharacterized LOC115695759
115695756	75738803	75766534	peroxidase 15
115695924	75772819	75776213	peroxidase 15

Chapter 3: Functional Characterization by Overexpression of *MYB* Candidate Genes in *C. sativa*

Abstract

Cannabis sativa is a historic crop with versatile uses for its grains, fibers, and secondary metabolites. Within the plethora of secondary metabolites contained in cannabis tissues are anthocyanins, polyaromatic flavonoid glycosylates which confer vibrant pink, purple, or red coloration, and can increase some abiotic stress tolerances. Despite strong interest in the genetics of secondary metabolite accumulation in cannabis, the genetic basis governing anthocyanin accumulation in this species remains unknown. In this study, we generated *Cannabis* over-expression mutants of two strong candidate genes putatively governing anthocyanin accumulation, *CsMYB33* and *CsMYB78*. We used the GA*A*NTRY agrobacterium vector system and the Wisconsin Crop Innovation Center Hemp gene-editing protocol and successfully recovered T_0 transgenic events of both insertional mutants. Interestingly, *CsMYB78* was unable to induce strong anthocyanin accumulation despite successful transgene integration. In contrast, *CsMYB33* induced heavy anthocyanin accumulation in all recovered explants. This work lays the foundation for future molecular investigations into the transcriptional regulatory network of anthocyanin accumulation in cannabis and provides novel insights into the molecular basis of secondary metabolite accumulation in this versatile and controversial crop.

Introduction

Research into the major secondary metabolites of cannabis has focused primarily on the psychoactive secondary metabolites Cannabidiol (CBD) and Tetrahydrocannabinol (THC) and on the main physical structures housing cannabinoids in plant tissue, glandular trichomes (Livingston et al., 2020; Xu et al., 2024; Fulvio et al., 2025). This research has been aided by the recent release of molecular and bioinformatic tools such as the cannabis pangenome and tissue specific expression atlas releases (Lynch et al., 2025; Barbosa-Xavier et al., 2024). While much is known of the genetic and physiological basis of cannabinoids, other secondary metabolites such as flavanols, flavonoids, terpenes, esters, volatile organic compounds, and hydrocarbons remain relatively understudied. In other crop and model species these compounds often play diverse physiological roles, can provide biotic or abiotic stress protection and, in some cases, have potential for development of new therapeutic drugs (Singh et al., 2015; Naing et al., 2020; Yadav et al., 2023). The development of bioinformatic resources and new molecular genetic methods such as pangenome assemblies and stable insertional transformation methods have been leveraged to investigate secondary metabolite biosynthesis and regulation, and to create new supply chains for rapid production of plant compounds with medicinal properties (Zhang et al., 2022). And indeed, cannabinoids and some unnatural analogues have been fully synthesized using yeast systems (Luo et al., 2019). However, research characterizing the biosynthesis and regulation of less prominent compounds within cannabis tissues is still hampered by the lack of widely available stable genetic transformation methods for

cannabis, and thus the full medicinal potential of its therapeutic compounds has not yet been leveraged (Tarkowski et al., 2025).

Cannabis is unusually recalcitrant to stable agrobacterium-mediated transformation methodology. The reason for this is likely its poor plant regeneration capability, with the first reported explant regeneration of cannabis in 2021 (Feeney and Punja, 2017; Zhang et al., 2021). Many attempts have been made over the last few decades to establish successful and replicable cannabis transformation methods, and indeed transient expression using hairy root cultures, protoplasts, and seedlings have been developed (Wahby et al., 2017; Beard et al., 2021; Li et al., 2024). Cannabis has also been successfully transformed and regenerated using agrobacterium infiltration and nanoparticle-based delivery systems, though these methods were not sufficient to generate stable T_1 transgenic events (Galan-Avila et al., 2021; Ahmed et al., 2021). The first stable transgenic T_1 event in cannabis is believed to have been developed at the Wisconsin Crop Innovation Center (WCIC) in 2020 with the insertion of the *Td-Tomato* fluorescent reporter gene cassette (https://cropinnovation.cals.wisc.edu/). This resource, unique to the University of Wisconsin, allows for characterization of the functionality of candidate genes regulating secondary metabolite accumulation or identity using overexpression or knockout mutants.

Anthocyanins are a class of glycosylated flavonoids common to most flowering plant taxa. They can provide a myriad of abiotic stress protections, and cause plant tissue to accumulate pink, purple, or red pigmentation when localized to the vacuole (Saigo et al., 2020). Of particular importance for cannabis, anthocyanins are strong visual attractants for consumers and are tied to the historic counterculture movement associated with the

species, making them an ideal target for development of attractive and economically viable cannabis cultivars. Current knowledge on the genetic basis of anthocyanin regulation in cannabis is limited to *in silico* gene identification and some transcriptomic analysis of putative MYB candidate genes (Bassolino et al., 2020; Kundan et al., 2022). Taken together, these studies suggested *CsMYB33* (LOC115708959) and *CsMYB78* (LOC115695758) as strong candidate genes which may upregulate anthocyanin accumulation in cannabis, presumably within an MBW complex. Gagalova et al. (2024) investigated the diversity of flavonoid biosynthetic enzymes and anthocyanin moieties within different cannabis varieties and found that while cyanidin-3-O-rutinoside was the most common anthocyanin, there were distinct differences between varieties in the composition of their overall flavonoid profiles. This study also provided comprehensive analysis of biosynthetic gene sequences, the promoters of which are potentially useful for validating candidate transcription factor (TF) promoter binding capability in *in-situ* hybridization assays. Furthermore, the chemical diversity of anthocyanin moieties, pigmentation patterns, and intensities both within and between varieties could suggest that some regulatory elements involved in anthocyanin biosynthesis or aglycone decoration may vary across cannabis genotypes.

In this study, we aimed to functionally characterize two strong candidate genes, *CsMYB33* and *CsMYB78*, that putatively upregulate anthocyanin accumulation in cannabis by overexpressing them. We report the development of T_0 events in cannabis and discuss results for both *CsMYB33* and *CsMYB78*. This study provides valuable insight into the nature of transcriptional regulation of secondary metabolic pathways in cannabis and is a

steppingstone towards the improvement of existing germplasm in the face of increasing environmental stressors and consumer demand.

Methods

GA*A*NTRY, Golden Gate Cloning, and Vector Assembly

The *Arabidopsis* Ubiquitin promoter AtUBQ10, the terminator AtuNOSt1, and the coding sequence (CDS) for *CsMYB33* and *CsMYB78* (Adapted from Kundan et al., 2022) (Supplementary Table 3.1) were assembled into a pL1F-1 vector (Weber et al., 2011) using a modified Golden Gate reaction protocol with 6uL of 5X NEB CutSmart buffer (New England Biolabs, USA) supplemented with 1M Dithiothreitol (DTT) and Adenosine Triphosphate (ATP), 100ng of pLF-1 vector, 100ng of the promoter, terminator, and CDS, 1uL of BsaIHF®v2, and 1uL of T_4 DNA ligase (New England Biolabs, USA), before being diluted to a final volume of 30uL with nuclease-free H_2O (Engler and Marillonet, 2014). A full breakdown of vector constituents can be found in Supplementary Tables 3.1 and 3.2. Vectors were then transformed into *E. coli* using heat shock at 42°C for 60s. *E. coli* was then plated on Luria broth (LB) selection media containing 100 mg L^{-1} carbenicillin and the scorable marker b-galactosidase (IPTG+X-Gal) and placed in an incubator set to 37°C overnight. At this stage, the bacterially expressed lacZ promoter driven b-galactosidase should be removed by the integration of insertions and carbenicillin resistance should be integrated, and thus surviving white colonies were picked. Plasmid DNA was extracted using the ZYMO Miniprep Plasmid DNA Extraction Kit (Zymo Research, USA). DNA concentration was verified using a nanodrop. DNA cargo was excised via *BbsI* high fidelity (New England Biolabs, USA) restriction enzyme digest from the pL1F-1 vector and integrated using the previously described modified Golden

Gate reaction scheme into the P donor vector (Collier et al., 2018) before being plated on LB selection media containing 50mg L^{-1} kanamycin, 5% (w/v) sucrose, and the scorable marker b-galactosidase. Colonies were picked again by visual identification of surviving white colonies, indicating the excision of the lacZ promoter driven b-galactosidase, the successful integration of kanamycin resistance, and the excision of the sacB sucrose sensitivity gene. Plasmid DNA was extracted for cotransformation along with the helper P vector into *Agrobacterium rhizogenes* ArPORT1(1411) via electroporation with a BIO-RAD GenePulser XCell (BIORad, USA) at 25 μF capacitance, 400 Ω resistance, 1800 volts, and using a 1 mm gap electroporation cuvette. Once transformed, the P donor vector and helper P vector integrated tDNA cargo into the virulence plasmid (Figure 3.1). Successful vector assembly was assessed after each ligation using AseI, PvuI, and SspI (New England Biolabs, USA) restriction enzyme digests to confirm fragment integration.

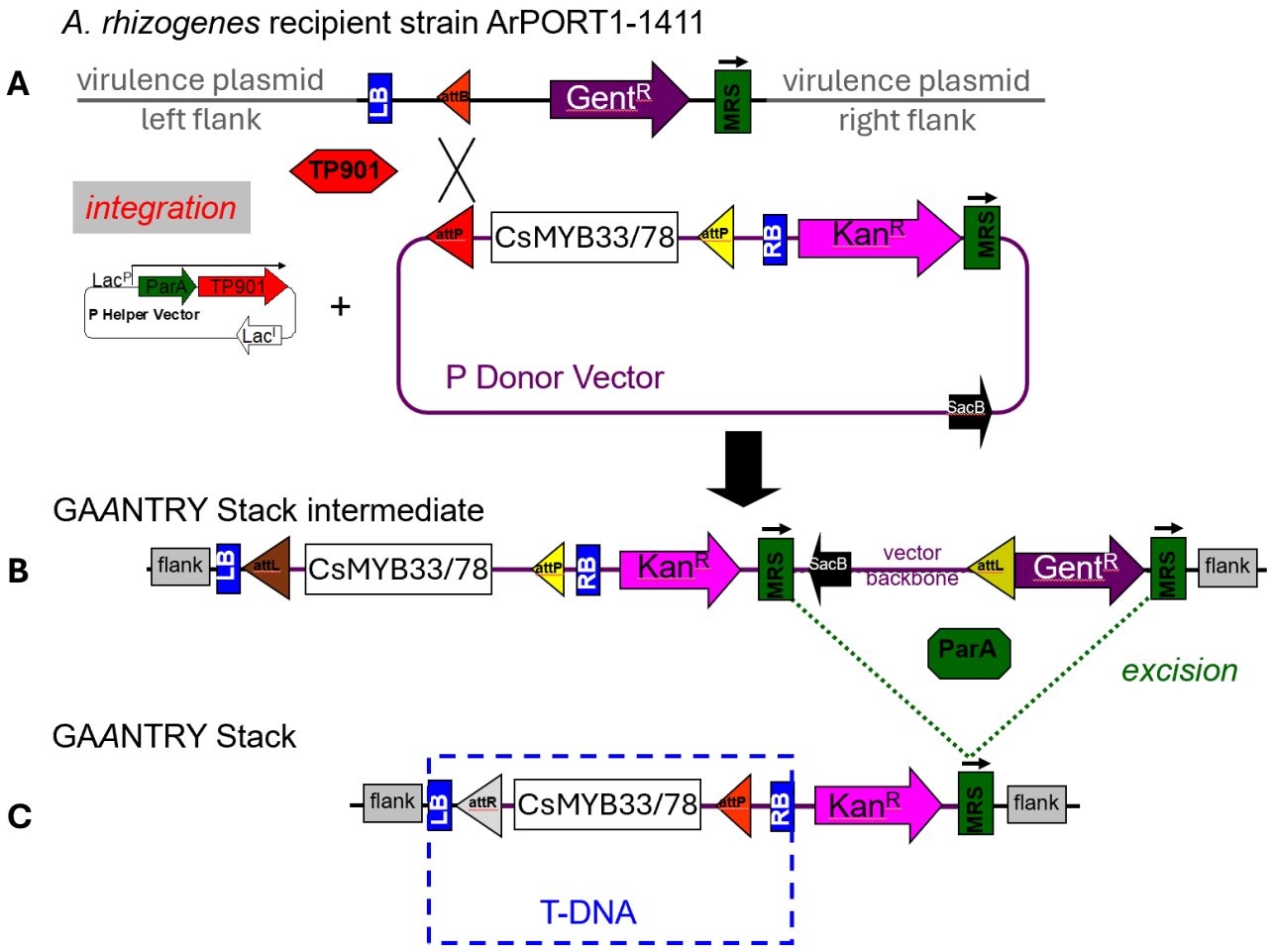

Figure 3.1. GAANTRY Vector Design for Agrobacterium Infiltration. Adapted from Collier et al., 2018. A) The integration of either *CsMYB78* or *CsMYB33* tDNA cargo from the Donor P vector into the ARPORT1-1411 virulence plasmid is facilitated by the TP901 restriction enzyme, encoded by the Helper P vector. Both attB and attP (red) sites are excised, and ligation is facilitated by the T_4 ligase enzyme (not pictured). B) The Donor P vector backbone is then excised from the virulence plasmid by the restriction enzyme ParA, which binds to the MRS recognition site, conferring sucrose insensitivity. C) The final virulence plasmid contains the tDNA cargo between the left and right borders (blue), and a Kanamycin resistance gene (pink) as a selectable marker.

In Planta Transformation

Seeds of the transformable line Badger 5B were sanitized via submersion in a 20% bleach solution for 5 minutes and then rinsed with sterile distilled water (SDW) at the

Wisconsin Crop Innovation Center (WCIC) Cannabis Plant Transformation Lab. Next, seeds were submerged in an imbibing solution in a sterile container and placed in a dark incubator at 37˚C for 24 hours. The same day, agrobacterium was prepared by adding 15uL of previously described vector-containing agrobacterium into 50ml of LB with kanamycin at a concentration of 50mg L^{-1} and was grown overnight in an incu-shaker set at 28°C and 200RPM. The following day, agrobacterium was tested for appropriate optical density (OD) levels at 660nm, and OD levels were confirmed to be between 0.5 and 1.0 for all samples. Agrobacterium was then centrifuged for 20 minutes at 3000RPM. Supernatant was decanted and agrobacterium was resuspended using WCIC inoculation (INO) media and acetosyringone. Agrobacterium samples were then placed on a shaker set to 175RPM at room temperature for one hour. During this time, imbibed seeds were rinsed with SDW and explants were excised to leave only the hypocotyls and first true leaf primordia then rinsed again with SDW. Excised hypocotyls were then placed into a sterile container and covered with 25mL of agrobacterium inoculum before sonication for 30s at 45-55kHz. Inoculated embryos were then shaken at room temperature for 30 minutes at 75RPM. After incubation the same inoculated hypocotyls were placed in 2.25mL of WCIC liquid co-culture media in a sterile container with filter paper to retain moisture and then placed into a Percival growth chamber at 23°C for 3 days. After co-culture, 15 explants per plate were transferred to deep petri dishes (25mm x 100mm) with filter paper and covered with 15mL WCIC liquid regeneration media. Plates were then kept in a controlled growth chamber at 27°C under a 16/8 photoperiod at a light intensity of 80-90um/s.m^2. Greening shoots from explants were harvested via dissection from hypocotyls once they reached 0.5cm and embedded into

WCIC Hemp Rooting media. Putative transgenic events were confirmed via digital droplet PCR using aadA spectinomycin resistance as a marker for successful transgene integration and *CsUBQ* as an internal standard (Protocol described in Collier et al., 2017). Primer sequences are listed in Supplementary Table 3.3. Greening explants were then replaced in a Percival growth chamber kept at 27°C under 16/8 photoperiod at a light intensity of 80-90um/s.m^2. Shoots successfully rooted within a range of 6 weeks to several months. Rooted samples were transferred to soil and kept in greenhouse conditions under a long-day light cycle to stimulate growth before being transferred to a short-day cycle to induce female flower formation. Following this, plants were then treated with Silver Thiosulfate (STS) to induce pollen production and bagged with 5um pore mesh bags to prevent pollen escape. T_1 seeds are currently being generated.

Anthocyanin Extraction and Quantification

Six inches of leaf and floral tissue was collected from lateral inflorescences of successfully rooted explants prior to senescence. The tissue was frozen at -20°C and then lyophilized on a bulk tray freeze dryer (Labconco, USA) for 48 hours prior to extraction and quantification of monomeric anthocyanins. Once lyophilized, tissue was finely ground and placed into amber vials with 1mL of 2% formic acid in methanol and kept at 4°C overnight. Samples were then diluted four-fold in methanol before being placed in plates with 150uL of either pH1.0 or pH4.5 buffer solutions. Samples sat for 15 minutes before being read on an 800TS Absorbance Reader (Biotek Instruments Inc., USA) at 515nm and 690nm. Anthocyanin concentrations were calculated in cyanidin-3-O-rutinoside equivalents, the

most common glycosylated flavonoid across cannabis cultivars (Gagalova et al., 2024) using the following formula adapted from Lee et al. (2001):

$$(A * DF * MW * 10^3)/(l * \mathcal{E})$$

Where A is the differential absorbance between samples at pH1.0 – pH4.5, DF is the dilution factor of 4, MW is the molecular weight of cyanidin-3-O-rutinoside: 595.52 g mol^{-1}, 10^3 is a scaling value to convert final units to mg L^{-1}, l is the path length of 1cm, and $\mathcal{E}$ is the molar extinction coefficient of 34300 L mol^{-1} cm^{-1}. Samples were run in triplicate, and A values were calculated as the average of each sample. Statistical significance was calculated using a Kruskal-Wallis rank sum test, with a p-value significance threshold of 0.05. Raw data can be found in Supplementary Table 3.4.

Results

CsMYB78 Overexpression

16 putatively successful T_0 *CsMYB78* transgenic explants displayed the selective spectinomycin resistance phenotype and of these, 11 were confirmed to contain transgene fragments via ddPCR (Figure 3.2). Of these 11, four survived the transfer to soil and were subsequently selfed to generate T_1 seeds. Two samples were harvested for anthocyanin extraction and quantification, with an average concentration of 0.81mg L^{-1} in cyanidin-3-O-rutinoside equivalents. Anthocyanin pigmentation was not visible in *CsMYB78* transformed plants throughout all stages.

CsMYB33 Overexpression

Seven putatively successful T_0 *CsMYB33* explants displayed spectinomycin resistance phenotypes, and of these, four were confirmed via ddPCR (Figure 3.2). Only two explants survived the transfer to soil and were subsequently grown out and selfed to generate T_1 seeds. Both plants were harvested for anthocyanin extraction and quantification, with an average concentration of 3.18mg L^{-1} in cyanidin-3-O-rutinoside equivalents. Though anthocyanin concentration was not significantly different between the two mutants ($p>0.05$), there were only two samples per group, casting some doubt on statistical power to detect true differences. However, *CsMYB33* samples appear to accumulate overall higher levels of anthocyanins over *CsMYB78* (Figure 3.2).

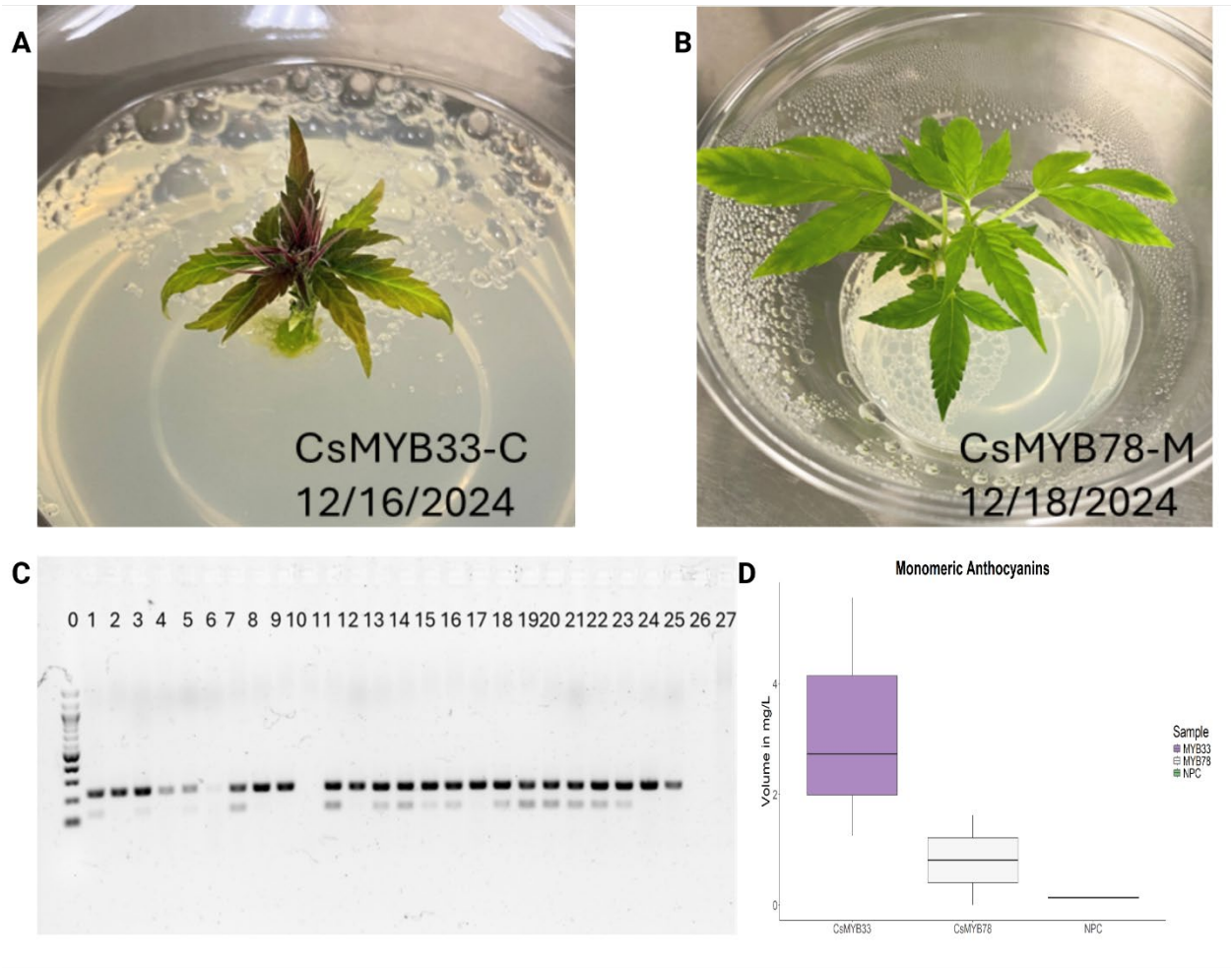

Figure 3.2. *CsMYB33* induces more pigmentation than *CsMYB78*. A) A *CsMYB33* explant plated on rooting media expresses visible pigmentation in leaves and floral structures. B) A *CsMYB78* explant plated on rooting media does not express visible pigmentation. C) Transgene integration is confirmed via ddPCR. Top bands correspond to an internal control CsUBQ (233bp) and bottom bands correspond to a spectinomycin resistance marker aadA (129bp). D) Anthocyanin concentration of mature inflorescences from mutants of both constructs and a non-pigmented control (NPC) are shown. Significance was calculated using a Kruskal-Wallis Rank Sum Test ($p>0.05$).

Discussion

Anthocyanins are a class of glycosylated flavonoids which confer pink, purple, or red pigments to plant tissues. They are aesthetically appealing, provide abiotic stress tolerance, and in the floral cannabis market, can drive consumer preference. Across plant taxa the

accumulation of anthocyanins is primarily spurred by transcriptional activation of key biosynthetic enzymes by R2R3-MYB transcription factors. In this study, we aimed to functionally characterize two R2R3-MYB candidate genes, *CsMYB33* and *CsMYB78*, previously discovered by Kundan et al. (2022), Bassolino et al. (2020), with the latter co-localized with marker clusters associated with flower pigment found in mapping studies by the Ellison lab. tDNA cargo was constructed using the *AtUBQ10* constitutive promoter and inserted via agrobacterium-mediated transformation into the WCIC transformable line Badger-5B. Interestingly, despite its phylogenetic clustering with R2R3-MYB activators of anthocyanin biosynthesis (Bassolino et al., 2020; Kundan et al., 2022), *CsMYB78* was unable to induce visible anthocyanin pigmentation while *CsMYB33* induced strong visible pigmentation.

Throughout the duration of this work, bioinformatic resources for analysis of cannabis genes evolved significantly, with a notable highlight of the recent release of the cannabis pangenome (Lynch et al., 2025). However, there remain significant barriers to overcome the broad integration of available genomic data such as the generation of accurate and consistent genome annotations across and within assemblies. For example, at the time of the genetic mapping in 2023 which led to the discovery of *CsMYB78* as a candidate gene, it was annotated in the cs10 reference genome as an ortholog of *AtMYB113*, though there was some disagreement over this annotation (Bassolino et al., 2020; Grassa et al., 2021). Currently, cs10 genome annotations have been censored by NCBI, and other annotation models support the identity of *CsMYB78* as an *AtMYB1* ortholog (Ryu et al., 2024). However, there are differences within the cannabis pangenome over the identity and locations of many

genes between accessions, leaving the true annotation of any given gene dubious without further analysis confirming function or genomic localization. In this study, *CsMYB33* and *CsMYB78* were chosen as candidate genes due to their believed orthology to known activators of anthocyanin biosynthesis, supported by phylogenetic clustering with subgroup 6 R2R3-MYB genes *PAP1, PAP2, AtMYB113,* and *AtMYB114* (Bassolino et al., 2020; Kundan et al., 2022).

The low levels of anthocyanins induced by *CsMYB78* is perhaps surprising given its high homology with the known activator of anthocyanin biosynthesis *MYB1*, an important anthocyanin activator in *Ipomoea*, onion, and apple (Morita et al., 2006; Schwinn et al., 2016; An et al., 2018; Yang et al., 2021) along with its sequence homology to other subgroup 6 R2R3-MYBs. However, in apple the functionality of *MYB1* is heavily influenced by environmentally inducible cofactors both pre- and post-transcriptionally. For instance, Yang et al. (2021) found that *MdMYB1,* the putative ortholog of *CsMYB78*, was phosphorylated and subsequently stabilized by the light induced kinase *MPK4*. Additionally, An et al. (2018) found that *MdMYB1* transcription is upregulated by *MdEIN-Like 1 (EIL1)* which in turn upregulates *ETHYLENE RESPONSE FACTOR 3,* a key activator of *EIL1,* in a positive feedback loop. *MYB1* orthologs are not exclusively involved in anthocyanin biosynthesis and have been implicated in coordinating phytohormone-mediated physiological responses to salt stress (Wang et al., 2015), in stimulating proper tapetum development by activating *A9* in *Arabidopsis* (Verma and Burma, 2017), and in maintenance of phosphate homeostasis during root development in rice (Gu et al., 2017). Taken together, the low level of anthocyanin induction by *CsMYB78* and the diverse roles of its orthologs indicate that *CsMYB78* may have multiple physiological

functions, or that its function in anthocyanin induction is dependent on specific environmental stimuli and the presence of mediating cofactors.

CsMYB33, by contrast, was able to induce strong visible anthocyanin pigmentation in all putative T_0 events and even appeared to be detrimental to explant health. Indeed, heavy pigmentation led to a higher mortality rate of *CsMYB33* explants both in controlled environment chambers and when transplanted into soil resulting in recovery of less than 50% as compared to *CsMYB78*. Currently annotated in cs10 as an ortholog of *AtMYB113*, Bassolino et al. (2020) suggested that *CsMYB33* may be an ortholog of the citrus R2R3-MYB *CsRUBY*, a well characterized activator of anthocyanin biosynthesis (Wang et al., 2024) and a common reporter gene in plant transformation (Yu et al., 2023). In citrus, *CsRUBY* is inducible by the binding of ethylene responsive cofactors to a retrotransposon in its promoter, which in turn leads to significant upregulation of anthocyanin biosynthesis in fruits. Notably, *CsMYB33,* localized to the distal arm of chromosome 3, was not physically close to any significant SNP markers found during mapping studies by the Ellison lab, possibly implying that the gene may be fixed in the explored populations. Further investigation into the region surrounding *CsMYB33* will be important in determining its potential impact on the transcriptional regulation of anthocyanin induction in cannabis.

Very little investigation has been done into the transcriptional regulation of anthocyanins in cannabis. However, a similar pattern of tissue pigmentation was found by Kundan et al. (2022), who observed that, when investigating *CsMYB33* and *CsMYB78* in transient expression assays in *N. benthamiana*, *CsMYB33* induced approximately three times as much anthocyanin as did *CsMYB78*. The similar pattern of induction exhibited by

Badger-5B explants may suggest that the *CsMYB78* transcript isolated from the accessions studied by Kundan et al. may not encode a functional variant, or at least a variant whose primary function is not activating anthocyanin biosynthesis. Additionally, tDNA cargo was not transcriptionally fused to spectinomycin resistance genes during vector construction (Figure 3.1), which may have led to partial transcription of transgenes containing *CsMYB78,* possibly explaining the low level of anthocyanins observed in explants. However, the pattern similarity of *CsMYB* mutant anthocyanin accumulation with Kundan et al. (2022) makes this somewhat unlikely (Figure 3.3).

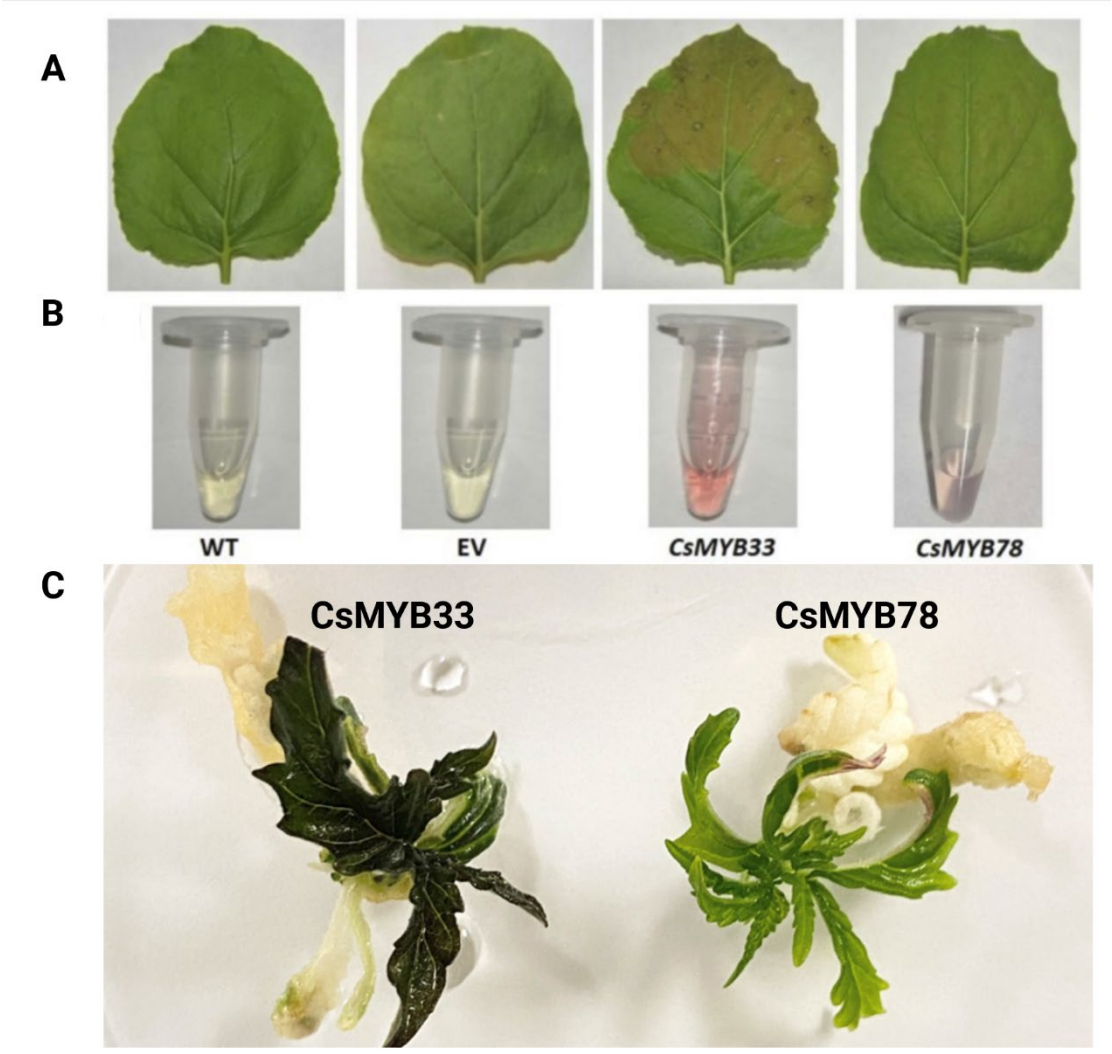

Figure 3.3. Comparison of *CsMYB33* and *CsMYB78* insertional mutants. A-B are adapted from Kundan et al., 2022. A) *N. benthamiana* transient expression assays of wild type (WT), empty vector (EV), *CsMYB33*, and *CsMYB78*. B) Anthocyanin extract supernatants corresponding to *N. benthamiana* samples. C) *CsMYB33* and *CsMYB78* putative T_0 event explants imaged prior to plating on rooting media.

The development of stable, environmentally resilient cannabis lines could help growers and breeders in producing visually attractive and economically viable varieties. Critical to this goal is understanding the regulatory network controlling anthocyanin accumulation. Presumably, anthocyanin biosynthesis in cannabis is regulated by a canonical MBW complex, though results of the current transgenic experiments suggest that

the MYB constituent of this complex may vary between varieties or subpopulations. Recent evidence suggests that gene duplication in cannabis is common, and that transposons are very active throughout its genome (Lynch et al., 2025). It is possible that *CsMYB78* or *CsMYB33* may be a duplicated or neofunctionalized version of a functional MYB gene upregulating anthocyanin biosynthesis. Overall, while this work contributes significantly to our knowledge of the transcriptional regulation of secondary metabolite biosynthesis in cannabis, there remain many unanswered questions and avenues for future study include investigating the transcriptomics of green and purple lines, developing stable mapping populations, and improving bioinformatic resources and analysis tools.

Future Directions

The primary activator of anthocyanin biosynthesis in cannabis remains elusive, as *CsMYB78* appears non-functional or absent in some varieties, and *CsMYB33* is not physically linked to significant SNP markers in both mapping populations studied. Future work in the investigation of cannabis anthocyanin biosynthesis could focus on developing larger and more diverse populations for delineating the regions affecting anthocyanin accumulation more broadly across cannabis germplasm.

The generation of stable transgenic lines with homozygous transgenic genotypes would be an invaluable resource for the rapid development of cannabis cultivars displaying high levels of anthocyanins by providing a true homozygous parent for breeding population development. Developing these stable lines will require putative T_1 events to undergo spectinomycin selective pressure to ensure transgene inheritance, and monomeric

anthocyanin concentration will be measured to determine the effect of transgene on pigmentation accumulation after inheritance.

MYB transcription factors bind to the promoter regions of biosynthetic genes via their R2 and R3 repeat domains and subsequently upregulate the transcription of these genes. Thus, determining the binding affinity of candidate genes to the promoter regions of known cannabis flavonoid biosynthetic enzymes using yeast two-hybrid assays could provide stronger evidence of putative candidate gene potential prior to insertional mutations. This screening method could prove especially useful in the eventuality that many candidate genes are discovered.

Transcriptional regulatory networks are complex and double or triple mutants of candidate genes representing identified cannabis MBW constituents could lead to altered transgenic pigment patterning or intensity phenotypes. Indeed, the true function of any one gene insert may only become clear in the absence or presence of another which stabilizes or aids its activity. Thus, it will be important to create double or triple mutants of desired genes to elucidate the downstream effects of regulatory candidate genes.

Sequencing the RNA of a broad collection of germplasm could highlight transcriptional differences between varieties with purple and green flowers, leaves, or stems and would enable utilizing weighted gene correlation network analyses, a powerful tool to tease out TFs truly affecting differences in anthocyanin accumulation in specific tissue types. This tool is of particular use in determining minor effect candidate genes which may impact only other regulators, allowing for a much fuller picture of the regulatory landscape

within cannabis tissues. Further, utilizing available transcriptomic data from the cannabis pangenome could allow for much broader sample sizes, and potentially lend statistical power to correlation analyses.

Utilizing quantitative reverse-transcriptase PCR or RNA sequencing on transgenic events would determine true expression levels of inserted transgenes and provide knowledge on the insertion efficiency of *CsMYB78* and *CsMYB33* and could help elucidate the basis of the pigmentation differences between the two.

The strong constitutive promoter *AtUBQ10* is a powerful tool for expressing tDNA cargo across all tissue types and was effective in inducing visible purple pigmentation in *CsMYB33* transgenic events. However, the intensity of pigmentation was very strong in purple events and may have been deleterious to plant health in some cases. Thus, future transgenic experiments could use a weaker promoter, allowing for less expression of tDNA cargo and potentially allowing for a lower explant mortality rate. Another potential avenue for alternative promoters could be to utilize the native promoter sequence of *CsMYB78* and *CsMYB33*, both of which may contain TF binding motifs important to their function of inducing anthocyanin pigmentation.

Finally, subjecting putative transgenic events to abiotic stress conditions could facilitate investigation into the full functionality of *CsMYB78* and *CsMYB33*. These genes may require environmentally induced cofactors to stabilize or alter their function, and the application of different stressors such as UV-B radiation or cold temperatures could lead to altered phenotypes in transgenic plants.

References

Ahmed, S., Gao, X., Jahan, M. A., Adams, M., Wu, N., & Kovinich, N. (2021). Nanoparticle-based genetic transformation of Cannabis sativa. *Journal of Biotechnology*, *326*, 48–51. https://doi.org/10.1016/j.jbiotec.2020.12.014

An, J.-P., Wang, X.-F., Li, Y.-Y., Song, L.-Q., Zhao, L.-L., You, C.-X., & Hao, Y.-J. (2018). EIN3-LIKE1, MYB1, and ETHYLENE RESPONSE FACTOR3 Act in a Regulatory Loop That Synergistically Modulates Ethylene Biosynthesis and Anthocyanin Accumulation. *Plant Physiology*, *178*(2), 808–823. https://doi.org/10.1104/pp.18.00068

Barbosa-Xavier, K., Pedrosa-Silva, F., Almeida-Silva, F., & Venancio, T. M. (2024). Cannabis Expression Atlas: A comprehensive resource for integrative analysis of Cannabis sativa L. gene expression. *Physiologia Plantarum*, *176*(6), e70010. https://doi.org/10.1111/ppl.70010

Bassolino, L., Buti, M., Fulvio, F., Pennesi, A., Mandolino, G., Milc, J., Francia, E., & Paris, R. (2020). In Silico Identification of MYB and bHLH Families Reveals Candidate Transcription Factors for Secondary Metabolic Pathways in Cannabis sativa L. *Plants*, *9*(11). https://doi.org/10.3390/plants9111540

Bassolino, L., Fulvio, F., Cerrato, A., Citti, C., Cannazza, G., Capriotti, A. L., Alberti, I., Terracciano, I., Pecchioni, N., & Paris, R. (2024). Metabolic characterization and transcriptional profiling of polyphenols in Cannabis sativa L. inflorescences with different chemical phenotypes. *Planta*, *260*(4), 76. https://doi.org/10.1007/s00425-024-04505-z

Bevan MW, Barnes WM, Chilton MD (1983a) Structure and transcription of the nopaline synthase gene region of T-DNA. Nucleic Acids Res 11: 369–385

Collier, R., Dasgupta, K., Xing, Y.-P., Hernandez, B. T., Shao, M., Rohozinski, D., Kovak, E., Lin, J., de Oliveira, M. L. P., Stover, E., McCue, K. F., Harmon, F. G., Blechl, A., Thomson, J. G., & Thilmony, R. (2017). Accurate measurement of transgene copy number in crop plants using droplet digital PCR. *The Plant Journal*, *90*(5), 1014–1025. https://doi.org/10.1111/tpj.13517

Collier, R., Thomson, J. G., & Thilmony, R. (2018). A versatile and robust Agrobacterium-based gene stacking system generates high-quality transgenic Arabidopsis plants. *The Plant Journal*, *95*(4), 573–583. https://doi.org/10.1111/tpj.13992

Engler, C., & Marillonnet, S. (2014). Golden Gate Cloning. In S. Valla & R. Lale (Eds.), *DNA Cloning and Assembly Methods* (pp. 119–131). Humana Press. https://doi.org/10.1007/978-1-62703-764-8_9

Feeney, M., & Punja, Z. K. (2017). The Role of Agrobacterium-Mediated and Other Gene-Transfer Technologies in Cannabis Research and Product Development. In S. Chandra, H. Lata, & M. A. ElSohly (Eds.), *Cannabis sativa L. - Botany and Biotechnology* (pp. 343–363). Springer International Publishing. https://doi.org/10.1007/978-3-319-54564-6_16

Fulvio, F., Pieracci, Y., Ascrizzi, R., Bassolino, L., Flamini, G., & Paris, R. (2025). Insights into terpenes profiling and transcriptional analyses during flowering of different Cannabis

sativa L. chemotypes. *Phytochemistry*, *229*, 114294. https://doi.org/10.1016/j.phytochem.2024.114294

Gagalova, K. K., Yan, Y., Wang, S., Matzat, T., Castellarin, S. D., Birol, I., Edwards, D., & Schuetz, M. (2024). Leaf pigmentation in: Characterization of anthocyanin biosynthesis in colorful Cannabis varieties. *Plant Direct*, *8*(11), e70016. https://doi.org/10.1002/pld3.70016

Galán-Ávila, A., Gramazio, P., Ron, M., Prohens, J., & Herraiz, F. J. (2021). A novel and rapid method for Agrobacterium-mediated production of stably transformed Cannabis sativa L. plants. *Industrial Crops and Products*, *170*, 113691. https://doi.org/10.1016/j.indcrop.2021.113691

Gheysen G, Angenon G, Van Montagu M (1998) Agrobacterium-mediated plant transformation: a scientifically intriguing story with significant applications. In: Lindsey K (ed) Transgenic plant research. Harwood Academic Publishers, Amsterdam, pp 1–33; https://doi.org/10.1016/0092-8674(84)90500-2

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., Michael, T. P., & Schwartz, C. J. (2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist*, *230*(4), 1665–1679. https://doi.org/10.1111/nph.17243

Hollingshead S, Vapnek D (1985) Nucleotide sequence analysis of a gene encoding a Streptomycin/Spectinomycin Adenyltransferase. Plasmid v13 p17-30. https://doi.org/10.1016/0147-619X(85)90052-6

Hirt H, Kögl M, Murbacher T, Heberle-Bors E (1990) Evolutionary conservation of transcriptional machinery between yeast and plants as shown by the efficient expression from the CaMV 35S promoter and 35S terminator. Curr Genet 17: 473–479

Irniger S, Sanfaçon H, Egli CM, Braus GH (1992) Different sequence elements are required for function of the cauliflower mosaic virus polyadenylation site in Saccharomyces cerevisiae compared with in plants. Mol Cell Biol 12: 2322–2330

Kundan, M., Gani, U., Fayaz, Mohd., Angmo, T., Kesari, R., Rahul, V. P., Gairola, S., & Misra, P. (2022). Two R2R3-MYB transcription factors, CsMYB33 and CsMYB78 are involved in the regulation of anthocyanin biosynthesis in Cannabis sativa L. *Industrial Crops and Products*, *188*, 115546. https://doi.org/10.1016/j.indcrop.2022.115546

Lee, J., Durst, R. W., Wrolstad, R. E., Barnes, K. W., Eisele, ; T, Giusti, ; M M, Haché, ; J, Hofsommer, ; H, Koswig, ; S, Krueger, D. A., Kupina, ; S, Martin, ; S K, Martinsen, ; B K, Miller, T. C., Paquette, ; F, Ryabkova, ; A, Skrede, ; G, Trenn, ; U, & Wightman, J. D. (2001). *Determination of Total Monomeric Anthocyanin Pigment Content of Fruit Juices, Beverages, Natural Colorants, and Wines by the pH Differential Method: Collaborative Study*. https://academic.oup.com/jaoac/article/88/5/1269/5657437

Li, M., Wu, Q., Guo, F., Ouyang, Y., Ao, D., You, S., & Liu, Y. (2024). A versatile, rapid Agrobacterium-mediated transient expression system for functional genomics studies in cannabis seedling. *Planta*, *260*(1), 18. https://doi.org/10.1007/s00425-024-04448-5

Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D. C. J., Rensing, K. H., Laflamme-Yonkman, J., Castellarin, S. D., Bohlmann, J., Page, J. E., & Samuels, A. L. (2020). Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *The Plant Journal*, *101*(1), 37–56. https://doi.org/10.1111/tpj.14516

Lynch, R. C., Padgitt-Cobb, L. K., Garfinkel, A. R., Knaus, B. J., Hartwick, N. T., Allsing, N., Aylward, A., Bentz, P. C., Carey, S. B., Mamerto, A., Kitony, J. K., Colt, K., Murray, E. R., Duong, T., Chen, H. I., Trippe, A., Harkess, A., Crawford, S., Vining, K., & Michael, T. P. (2025). Domesticated cannabinoid synthases amid a wild mosaic cannabis pangenome. *Nature*. https://doi.org/10.1038/s41586-025-09065-0

Meyer P, Kartzke S, Niedenhof I, Heidmann I, Bussmann K, Saedler H (1988) A genomic DNA segment from Petunia hybrida leads to increased transformation frequencies and simple integration patterns. Proc Natl Acad Sci U S A 85: 8568–8572

Morita, Y., Saitoh, M., Hoshino, A., Nitasaka, E., & Iida, S. (2006). Isolation of cDNAs for R2R3-MYB, bHLH and WDR Transcriptional Regulators and Identification of c and ca Mutations Conferring White Flowers in the Japanese Morning Glory. *Plant and Cell Physiology*, *47*(4), 457–470. https://doi.org/10.1093/pcp/pcj012

Naing, A. H., & Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiologia Plantarum*, *172*(3), 1711–1723. https://doi.org/10.1111/ppl.13373

Norris SR, Meyer SE, Callis J (1993) The intron of Arabidopsis thaliana polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Molecular Biology v21 p895-906

Ryu, B.-R., Gim, G.-J., Shin, Y.-R., Kang, M.-J., Kim, M.-J., Kwon, T.-H., Lim, Y.-S., Park, S.-H., & Lim, J.-D. (2024). Chromosome-level Haploid Assembly of Cannabis sativa L. cv. Pink Pepper. *Scientific Data*, *11*(1), 1442. https://doi.org/10.1038/s41597-024-04288-8

Saigo, T., Wang, T., Watanabe, M., & Tohge, T. (2020). Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. *Physiology and Metabolism*, *55*, 93–99. https://doi.org/10.1016/j.pbi.2020.04.001

Schwinn, K. E., Ngo, H., Kenel, F., Brummell, D. A., Albert, N. W., McCallum, J. A., Pither-Joyce, M., Crowhurst, R. N., Eady, C., & Davies, K. M. (2016). The Onion (Allium cepa L.) R2R3-MYB Gene MYB1 Regulates Anthocyanin Biosynthesis. *Frontiers in Plant Science*, *Volume 7-2016*. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.01865

Singh, A., Kumar, A., Yadav, S., & Singh, I. K. (2019). Reactive oxygen species-mediated signaling during abiotic stress. *Plant Gene*, *18*, 100173. https://doi.org/10.1016/j.plgene.2019.100173

Tarkowski, P., Kopecký, P., Riboulet-Zemouli, K., & Frébort, I. (2025). Where are you heading to, Cannabis research? *Phytochemistry Reviews*. https://doi.org/10.1007/s11101-024-10052-y

Verma, N., & Burma, P. K. (2017). Regulation of tapetum-specific A9 promoter by transcription factors AtMYB80, AtMYB1 and AtMYB4 in Arabidopsis thaliana and Nicotiana tabacum. *The Plant Journal*, *92*(3), 481–494. https://doi.org/10.1111/tpj.13671

Wahby, I., Caba, J. M., & Ligero, F. (2017). Hairy Root Culture as a Biotechnological Tool in C. sativa. In S. Chandra, H. Lata, & M. A. ElSohly (Eds.), *Cannabis sativa L. - Botany and Biotechnology* (pp. 299–317). Springer International Publishing. https://doi.org/10.1007/978-3-319-54564-6_14

Wang, T., Tohge, T., Ivakov, A., Mueller-Roeber, B., Fernie, A. R., Mutwil, M., Schippers, J. H. M., & Persson, S. (2015). Salt-Related MYB1 Coordinates Abscisic Acid Biosynthesis and Signaling during Salt Stress in Arabidopsis. *Plant Physiology*, *169*(2), 1027–1041. https://doi.org/10.1104/pp.15.00962

Wang, Y., Li, S., Shi, Y., Lv, S., Zhu, C., Xu, C., Zhang, B., Allan, A. C., Grierson, D., & Chen, K. (2024). The R2R3 MYB Ruby1 is activated by two cold responsive ethylene response factors, via the retrotransposon in its promoter, to positively regulate anthocyanin biosynthesis in citrus. *The Plant Journal*, *119*(3), 1433–1448. https://doi.org/10.1111/tpj.16866

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/journal.pone.0016765

Xu, Y., Zhang, J., Tang, Q., Dai, Z., Deng, C., Chen, Y., Cheng, C., Yang, Z., Zhang, X., Chen, J., Luan, M., & Su, J. (2024). Integrated metabolomic and transcriptomic analysis revealed the regulation of yields, cannabinoid, and terpene biosynthesis in Cannabis sativa L. under different photoperiods. *South African Journal of Botany*, *174*, 735–746. https://doi.org/10.1016/j.sajb.2024.09.046

Xu, Z.-S., Yang, Q.-Q., Feng, K., Yu, X., & Xiong, A.-S. (2020). DcMYB113, a root-specific R2R3-MYB, conditions anthocyanin biosynthesis and modification in carrot. *Plant Biotechnology Journal*, *18*(7), 1585–1597. https://doi.org/10.1111/pbi.13325

Yadav, S. P. S., Kafle, M., Ghimire, N. P., Shah, N. K., Dahal, P., & Pokhrel, S. (2023). An overview of phytochemical constituents and pharmacological implications of Cannabis sativa L. *Journal of Herbal Medicine*, *42*, 100798. https://doi.org/10.1016/j.hermed.2023.100798

Yang, T., Ma, H., Li, Y., Zhang, Y., Zhang, J., Wu, T., Song, T., Yao, Y., & Tian, J. (2021). Apple MPK4 mediates phosphorylation of MYB1 to enhance light-induced anthocyanin accumulation. *The Plant Journal*, *106*(6), 1728–1745. https://doi.org/10.1111/tpj.15267

Yu, J., Deng, S., Huang, H., Mo, J., Xu, Z.-F., & Wang, Y. (2023). Exploring the Potential Applications of the Noninvasive Reporter Gene RUBY in Plant Genetic Transformation. *Forests*, *14*(3). https://doi.org/10.3390/f14030637

Zhang, X., Xu, G., Cheng, C., Lei, L., Sun, J., Xu, Y., Deng, C., Dai, Z., Yang, Z., Chen, X., Liu, C., Tang, Q., & Su, J. (2021). Establishment of an Agrobacterium-mediated genetic

transformation and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (Cannabis Sativa L.). *Plant Biotechnology Journal*, *19*(10), 1979–1987. https://doi.org/10.1111/pbi.13611

Supplementary Tables

Supplementary Table 3.1 CsMYB Coding Sequences. CsMYB nucleotide coding sequences from Kundan et al. (2022) which were used in the construction of GAANTRY vectors. Sequences were 98.81% identical and 95.40% identical to cs10 reference transcripts, XM_030636983.1 and XM_030622841.1, respectively.

Gene	Coding Sequence
CsMYB33	ATGCCGCGGAGCAATATTAATATGGTGACAGTATTAGTAACGGACCAAAGGGAGGGCGGTTCCTCTGGTGGGAGAGCGGTCAGGAAAGG CGCATGGACCAGAGAAGAAGATGATCTTCTTAAAGATTGCGTAGACAAGTATGGTGAAGGCAAATGGCATCTTGTTCCACTTAGAGCTGGAT TGAATAGATGTAGGAAGAGTTGTAGACTGAGGTGGTTGAATTATTTGAAGCCCGATATCAAACGAGGAGAATTTACGGCCGATGAAGTTGAC CTTATTCTTAGACTTCATAAGCTTTTAGGCAACAGATGGTCCCTAATCGCAGGCAGAATACCAGGAAGGACAGCCAATGATGTGAAAAATTA TTGGAACACACACTTGGGAAAGAAGGTAATGATGTCCTACGATAAACTTAATCAAAAACACTACCTCAGAAACAAAGATGATCATCACAAGT TTATCCCTGAATCCAACATCATTAAACCTCGTCCTCGGACCTTCAGAAACCAATTACCCTTGACGAATACTACCACTGTTTCAAAAAACAATA TTAGCTCTGCACCAGTACCACATGTAGTAGCCTGTACACTTAGCCAGCATGGGCCTGTTTTAGAAATGAATAATAAGAACTGGGGCTGGGA GAGCTTATTAGACATGATCGACAGCGAGGATCAACAATTTTTGAGCTCTGGTTTTGATGAAGAATTTAATAATGTTGCCCATAACTCTGAATTTG AACCTGTTTCTCATCGACCTTCAGAATCAAAAATCAAAGACATTAGTACTACTACTACTAATAACAGTGGTATTTTTGTCATTGATGGCTATCAA AATTGCTCAACTGACTTTACCATGGACATGGACCTATGGGATCTCGGAATCACTGCCACATCAGTACCAGACAGGGAAATGCATGAGCAA CCTTAA
CsMYB78	ATGGTGAGAGTATTAGTAAAGAAAAGTAGTTTGAATGTGAAGAAAGGTGCATGGAGTAGAGAAGAAGACATTGTTCTTAGGGAATGCATAGAC AAATATGGTGAAGGAAAATGGCATCTTGTTCCTCTTAGAGCTGGATTGAGTAGATGTAGGAAGAGTTGTAGATTGAGGTGGTTGAACTATTTGA AGCCTGATATTAAACGTGGAGAATTTGAGGAGGACGAAGTTGATTTAGTTCTTAGACTTCATAAACTTTTAGGCAACAGATGGTCATTGATTGC CGGTAGAATCCCAGGAAGAACAGCTAATGATGTGAAAAATTATTGGAATACTCATCAGAGAAAAAAGATGGTTATTAATATTGATATTCATCAA AGTAGTAAACTCAATAATAATAATAATAACCATGATAAAGCATGGTCCAACATGGTAACAAAGCCTCGTCCAACTCGAATCATCAGAAGCCAA TTATCAGCCATGATGAATAATAATAAACAACCTGTTATAGATCATGACCAAACAAATAATAAGAATTATTGTTGGAGGGAATTATTAAACAACAAT ATCACAGAACATGGTGACCAAATAATTCAAGGATATATAGAGAAACCCATAATAAATAGTGGTTTGGCTAATCATGATCATGTAGTTTCAGAATG CAATGATTATTTCAGGGGCCATAGTACAAGTGATGATCAATTTGTCAATAATGATCATGACTACCACAATTTTTTGACAGACTTCAACATTAATG ACATCGACATTCACCTTTGGGACCTTTGA

Supplementary Table 3.2. GAANTRY Vector Design Components. Table is reproduced with the permission of Wisconsin Crop Innovation Center.

Long name	Short name	Source Organism	Locus	Description	Reference
Right Border	RB	Agrobacterium sp.		Right Border - site of T-strand	Gheysen et al., 1998

				formation by virD2 nicking	
Arabidopsis thaliana Ubiquitin 10 promoter	AtUBQ10p	Arabidopsis thaliana	AT4G05320	Strong Constitutive Promoter	Norris et al., 1993
Nopaline synthase terminator	NOSt	Agrobacterium tumefaciens strain C58		Sequence directing transcription termination and polyadenylation of transcript	Bevan et al., 1983a
Transformation Booster Sequence	TBS-1	Petunia hybrida	LOC_Os02g06640.1	Insulator used to protect promoters from potential adjacent enhancers	Meyer et al., 1988
Cauliflower Mosaic Virus 35S promoter	2XCaMV35S pTMV Ω	plant virus		Promoter for constitutive expression of aadA1a	Odell et al., 1985 Kay et al., 1987 Gallie and Walbot, 1992 Gallie, 2002
Aminoglycoside 3"-adenyltransferase	aadA1a	Escherichia coli		resistance to spectinomycin	Hollinghead and Vapnek, 1985
Cauliflower Mosaic Virus 35S terminator	CaMV35St	Cauliflower Mosaic Virus		Sequence directing transcription termination and polyadenylation of transcript	Hirt et al., 1990 Irniger et al., 1992
Left Border	LB	Agrobacterium sp.		25 bp sequence delineating	Gheysen et al., 1998

				the end of the Transfer DNA	

Supplementary Table 3.3. ddPCR Primers

Primer Name	Primer Sequence	Description
P0140-CsUBQ_F7	TGCTGCTGTGTTCTTGATGA	Cannabis Ubiquitin internal control
P0141-CsUBQ_R7	GTGGTACCTTCTGATCTGCC	
D0080-aadA1a_F107	CTGCTGTTGATGGAGGACTT	aadA/Spectinomycin selection marker
D0081-aadA1a_R235	CTCCTGGAGAAGCAGAAGTT	

Supplementary Table 3.4. Anthocyanin Concentrations of Transgenic Plants

Sample	Concentration (mg L^{-1})
***CsMYB33* Plant 1**	7.08
***CsMYB33* Plant 2**	2.53
***CsMYB78* Plant 1**	2.06

***CsMYB78* Plant 2**	0
Non-Pigmented Control	0.13

Conclusion

The regulatory network of anthocyanin accumulation is complex and mainly mediated by an MBW transcription factor complex composed of an R2R3-MYB, a bHLH cofactor, and a WDrepeat scaffold protein. Previous mapping studies from the Ellison lab indicated a region on chromosome 6 from 74.54Mb – 75.77Mb as a primary region of interest for candidate genes influencing anthocyanin accumulation in cannabis (S. Kim, Unpublished). In this work, we performed a genome wide association study (GWAS) to map regions of the cannabis genome affecting anthocyanin accumulation in flowers, leaves, and stems, and were able to pinpoint a significant marker association cluster on chromosome 1 affecting leaf and stem anthocyanin and confirm the boundaries of the cluster on chromosome 6 affecting flower anthocyanin pigmentation. Within the bounds of both major marker clusters were five strong candidate regulators of anthocyanin accumulation: *CsMYB17, CsGLABRA3, CsANTHOCYANINLESS2, CsbHLH143,* and *CsMYB78*. DNA and RNA sequencing of *CsMYB78* revealed several large cultivar-dependent insertions and SNPs in the promoter, intron, and terminator of this gene across accessions.

Two candidate genes were chosen for further functional characterization: *CsMYB78*, an R2R3-MYB found within the bounds of the marker cluster on chromosome 6, and *CsMYB33*, an R2R3-MYB previously identified as an ortholog of *CsRUBY* by Bassolino et al. (2020) and Kundan et al. (2022). Interestingly, *CsMYB33,* localized at 16.48Mb-16.49Mb on chromosome 3, was not found in any of our mapping studies, suggesting that it may be fixed in our populations. The results of *in planta* insertional mutants showed that *CsMYB33* could induce visible anthocyanin pigmentation while *CsMYB78* could not.

Taken together, these results may suggest that the MBW-mediated regulation of anthocyanin accumulation in cannabis is not only tissue specific, but that cultivar or population-dependent alterations exist in insertions (possibly transposable elements containing environmentally inducible TF recruitment motifs) affecting the expression of key regulatory elements. Indeed, in controlled environment experiments it was often seen that cultivars expected to stay green could accumulate anthocyanins under specific conditions (Unpublished data), suggesting that the difference between genotypes of green and purple samples in our mapping population was not in their ability to accumulate anthocyanin, but in their ability to accumulate anthocyanin in response to abiotic stress. This idea is further supported by two studies investigating orthologs of *AtMYB1* in apple, the former of which found that *MdMYB1* was upregulated by the ethylene inducible cofactor *MdEIN3-Like1* and the latter finding that *MdMYB1* required protein level stabilization by another light inducible cofactor, *MdMPK4* (An et al., 2018; Yang et al., 2021). Overall, the regulation of anthocyanin biosynthesis in cannabis appears to be variable between some cultivars or populations based on environmental response capabilities. Future work creating double and triple mutants of candidate MBW complex constituents, investigating the effect of native promoters on environmental response, and developing inbred lines for QTL mapping could lead to new insights into the regulatory network and confirm hypotheses of tissue specific activity. Teasing out the network of interactions affecting transcription factors and confirm function across broad cannabis accessions will be important in developing our knowledge of the molecular basis of secondary metabolite regulation across tissues and open the door for targeted development of aesthetic and competitive cannabis cultivars.