ELUCIDATING THE GENETIC CONTROL OF QUALITATIVE TRAITS IN HEMP

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By Jacob Alexander Toth

August 2022

ELUCIDATING THE GENETIC CONTROL OF QUALITATIVE TRAITS IN HEMP

Jacob Alexander Toth, Ph.D.

Cornell University 2022

Hemp (*Cannabis sativa* <0.3% tetrahydrocannabinol, THC) is a crop with great potential, but historical blanket bans on *C. sativa* have meant that knowledge and tools are lacking for breeding and production. With the recent legalization of hemp, there has been a wave of interest in the crop, but basic questions about the plant still remain. In this dissertation, data and tools relevant to hemp breeding and production are presented, including an overview of the crop, available genetic tools, and applicable breeding concepts. I describe the development and application of tools to distinguish cannabinoid chemotype and plant sex. The utility of these tools was shown in a marker assisted selection scheme leading to a new cultivar and the determination that a plant with two Y chromosomes and no X chromosomes is likely inviable. A time course analysis of cannabinoid production throughout the course of flowering under several biotic and abiotic stresses was conducted, leading to the somewhat surprising result that in high cannabidiol (CBD) cultivars, the CBD:THC ratio was essentially fixed, information with major importance for CBD production. Finally, I examined the genetics of flowering time and mapped two major flowering time loci on Chromosome 1, named *Autoflower1* and *Early1*. I also analyzed accessions that were induced to flower under continuous light, which suggests the presence of multiple genes controlling photoperiod insensitivity in *C. sativa* germplasm. Possible extensions of this work, some of which are ongoing projects of the Cornell Hemp Breeding team, are discussed. Overall, this work provides useful data for future hemp breeding efforts and high-throughput molecular tools to facilitate breeding.

BIOGRAPHICAL SKETCH

Jacob Alexander Toth was born in Brandon, Manitoba, Canada. Living in the countryside, he was gardening before he could walk, sparking a continued interest in the natural world. Outside of science and related endeavors, he was a distinguished violinist in his youth, regularly winning local competitions. He graduated as valedictorian from Vincent Massey High School in 2010 with the designation of National AP Scholar, and would impress or aggravate friends and family with information about plant identification and taxonomy. Jacob began his undergraduate students at the University of British Columbia, graduating in 2014 with a First Class Bachelor of Science (honors) in Biochemistry and Molecular Biology, with a minor in Philosophy. While not take any plant-related courses until starting his Ph. D., he was closely associated with the Botany Enthusiast's Club and worked with Dr. Lacey Samuels on aspects of xylem cell wall formation in *Arabidopsis* as an undergraduate. He was also active in the iGEM competition for several years, leading to an understanding of the mechanics and limitations of synthetic biology. His honors thesis was on directed protein evolution in Dr. Nobuhiko Tokuriki's lab. After graduating and a short stint working at a tea shop, Jacob began a job as a Research Assistant for the Canadian Government, using molecular tools to assist in breeding wheat and oats for the Canadian Prairie. During this time he began to pursue physical endeavors, being active in powerlifting and climbing mountains including Kilimanjaro, Mauna Kea, and Iztaccihuatl. From there, he began his Ph.D at Cornell in 2018. His work in hemp breeding is detailed in this thesis, which led to the Munger/Murphy Award for Outstanding Graduate Students in Plant Breeding and the Barbara McClintock Graduate Student Award for 2021. Outside of science, he continues to play the violin, mountaineer, and garden.

DEDICATION

Dedicated to Dr. Santosh Kumar of the Brandon Research and Development Centre, for showing me there was more to plant breeding than genomic selection.

ACKNOWLEDGEMENTS

I would like to thank my committee Awais Khan, Gaurav Moghe, Alireza Abbaspourrad, and Larry Smart for their support and encouragement. My major advisor, Larry Smart, has done an astounding job starting the Cornell Hemp program from nothing, and without his acumen there would be very little to report in this thesis. His support in manuscript editing, funding acquisition and management, greenhouse and field administration, and scientific insight were instrumental in the design and execution of the projects detailed here and many more. Other members of his lab, notably current and former students George Stack and Craig Carlson, were also instrumental in developing and testing ideas. The Smart Lab technical staff, especially Lauren Carlson, Jane Petzoldt, Dawn Fishback, Deanna Gentner, Rebecca Wilk, Allison Desario, Teagan Zingg, Alexander Wares, and McKenzie Schessl provided invaluable assistance in the lab, field, and greenhouse. Special thanks goes out to Michael Quade, lab tech par excellence. Several high school and undergraduate students also contributed to this work, including the valuable contributions of Ben DeMoras and Brian Nardone. Outside the Larry Smart Lab, thanks goes out to the "other" Smart Lab, with Ali Cala and Dr. Christine Smart greatly assisting with matters pertaining to pathology. Jamie Crawford, currently in the Moore Lab and formally in the Viands Lab, was an essential contact in Ithaca, assisting in yield trials and management of Ithaca trials. Thanks also to Martin Liu in Alireza Abbaspourrad's lab for a productive collaboration concerning hemp protein. Lastly, I'd like to thank members of the Rose Lab, notably Stephen Snyder and Glenn Philippe, for running the thousands of HPLC samples underlying much of what is presented here.

This work would not be possible without the gracious contribution of germplasm from multiple sources, including Fiacre Seeds, UNISeeds, Hemplogic, Industrial Seed Innovations, Sunrise Genetics, Kayagene, WinterFox Farms, Genesis Hemp Alliance, Boring Hemp, NY Hemp Source, Stem Holdings Agri, AssoCanapa, Joe Calderone, Eric Cerecedes, Go Farm Hemp, Hemplogic, Ultra Rich CBD, Green Lynx Farms, JDMI Limited, CN Kenaf and Hemp Seed Farm, PreProcess, Bish Enterprises, Arcardia Biosciences, Hemp Genetics International, International Hemp, Paul Smith Denver Co., Parkland Industrial Hemp Growers, New West Genetics, Flura, American Hemp Co., Point3 Farma, Ventura Seed Company, Oregon CBD, Phylos Bioscience, Hiliard, Endoca , Legacy Hemp, Schiavi Seed, Phytonyx, Ryes Creek, and Atlas Seeds.

This work was partially supported by New York State Department of Agriculture and Markets through grants AC477 and AC483 from Empire State Development Corporation, as well as the Scotts Corporation through a grant from the Foundation for Food and Agriculture Research Hemp Research Consortium, a sponsored research agreement with Pyxus International, and Thai Leaf.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

AAFCO: Association of American Feed Control Officials

BSA: Bulk segregant analysis

CBC: Cannabichromene

CBCA: Cannabichromenic acid

CBCAS: Cannabichromenic acid synthase

CBD: Cannabidiol

CBDA: Cannabidiolic acid

CBDAS: Cannabidiolic acid synthase

CBDVA: Cannabidivaric acid

CBG: Cannabigerol

CBGA: Cannabigerolic acid

CBGVA: Cannabigerovarinic acid

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleoside triphosphate

GBS: Genotyping-by-sequencing

GOT: Geranyl-diphosphate:olivetolate geranyltransferase

GPP: Geranyl pyrophosphate

GWAS: Genome-wide association study

HPLC: High-performance liquid chromatography

KASP™ : Kompetitive Allele-Specific PCR

LAMP: Loop-mediated isothermal amplification

MADC: Male Associated DNA in Cannabis

NIRS: Near infrared spectroscopy

PACE: PCR Allele Competitive Extension

PCR: Polymerase chain reaction

PVE: Percent variance explained

SNP: Single nucleotide polymorphism

SSR: Simple sequence repeat

STS: Silver thiosulfate

THC: Tetrahydrocannbinol

THCA: Tetrahydrocannabinolic acid

THCAS: Tetrahydrocannabinolic acid synthase

THCVA: Tetrahydrocannabinovaric acid

Chapter 1: Background information and hypotheses

1.1 Introduction to hemp

What is hemp?

Hemp (*Cannabis sativa* L.) is a multi-use crop, with over 50, 000 potential applications (Cherney & Small, 2016). Hemp is usually dioecious (XY), diploid (2n=10), and photoperiod sensitive (short-day), although exceptions to all of these traits are known and used in various production systems (Razumova et al., 2016; Sharma et al., 2015; Zhang et al., 2021). The haploid genome size of female (XX) *C. sativa*. is estimated to be 736 Mb (Grassa et al., 2021), with the Y chromosome being longer than the X chromosome (Divashuk et al., 2014). Hemp has been historically understudied due to blanket bans on the study of *C. sativa*, but there has been recent renewed interest in the crop (Fike et al., 2020).

Hemp uses

While there are perhaps countless potential uses of hemp, there are three main things for which hemp is grown. These are the grain (often used as food or feed), the stems (often for bast fiber and hurd), and the inflorescences (often for cannabinoids other than tetrahydrocannabinol, THC). The grain has excellent food value being rich in protein and omega-3 fatty acids (Callaway, 2004), but yields are not yet competitive with staple crops in the US (*National Hemp Report*, 2022). The stems can be divided into two commercially

relevant sections, called the bast and the hurd. The bast consists of fine cellulosic fibers useful for textiles and anatomically consists of phloem-related sclerenchyma, while the hurd anatomically is secondary xylem, and can be used like wood in applications such as chips or pulp (Zhao et al., 2021). The prodigious biomass of the stems has also led to interest in developing hemp as a bioenergy crop (Das et al., 2017). Lastly, the inflorescences of the plant contain the highest concentration of cannabinoids and terpenes, high-value compounds with potential medicinal properties (White, 2019).

Hemp legal history

The first documented use of hemp can be traced back to cord-impressed pottery about 10,000 BCE, which is supported by genomic analysis of domestication (Ren et al., 2021). Hemp was widely used as a fiber crop until the early 20th century, and in fact the word "canvas" is derived from the genus name *Cannabis* (Small, 2015). However, due to associations with high-THC marijuana, hemp production was discouraged and eventually banned in the USA, with a short window of cultivation to support the American war effort in WWII (Small, 2015; Fike et al., 2020). With the 2014 Farm Bill it became legal to grow hemp in a research capacity in the United States, and with the 2018 Farm Bill hemp was no longer restricted to grow.

Hemp market history

During the first few years of hemp legalization, the predominant category of hemp production was for the inflorescence, particularly for cannabidiol (CBD) production. However, due to massive overproduction of CBD hemp and falling CBD prices, CBD acreage fell rapidly (Cruz et al., 2021). Despite reductions of CBD hemp acreage,

inflorescence production remained the major market class for hemp in the USA in 2021, grown on 15,980 acres and valued at $623 million. By way of contrast, hemp for fiber was grown on 12,690 acres and was valued at $41.4 million and hemp for grain was grown on 8,255 acres and valued at $5.99 million (*National Hemp Report*, 2022).

Hemp yields

The useable part of hemp varies by market class, leading to very different yields per acre. For fiber production, as most of the plant is used, yields can be above 20 Mg ha^{-1}, competitive with bioenergy crops (Tang et al., 2016). However, US production averaged 2.94 Mg ha^{-1} in 2021, potentially due to a lack of adapted cultivars. For grain production, yields of up to 3,000 pounds per acre have been recorded (Horner et al., 2019), but US production averaged 530 pounds per acre in 2021. This is again likely due to a lack of adapted cultivars, as most modern grain hemp cultivars were bred at high latitudes with a shorter growing season, such as Canada and Finland (Zhang et al., 2021). Estimations of the upper bounds of CBD hemp production are not readily available, but US production averaged 1,235 pounds per acre in 2021. The concentration of CBD in the inflorescence was not noted in the 2021 National Hemp Report, but higher CBD concentrations generally command higher market prices.

Hemp production systems

Given the varied nature of the market classes of hemp, it is not surprising that varied production systems are used. For high-cannabinoid hemp, an intensive horticultural system is usually used (Stack et al., 2021), involving raised beds, wide spacing, weed control, and hand-harvesting (Figure 1.1A). Grain and fiber production tend to be more similar to

standard row crops (Figure 1.1B), but harvesting tends to be more difficult due to a lack of specialized equipment to deal with the tall crops with strong fibers (Fike, 2016). There have also been efforts to create dual-purpose or even tri-purpose cropping systems. However, these efforts are somewhat stymied by inherent tradeoffs. Notably, fiber quality and cannabinoid content are lower in plants with mature seeds, and the preferred all-female cropping system for cannabinoid production is unsuitable for fiber and grain production due to cost and pollination concerns (Kurtz et al., 2020; Westerhuis et al., 2019). Despite lower yields and quality, multiple-use cropping systems may be more profitable and sustainable due to the ability to access multiple markets and weather changes in demand and price (Tang et al., 2016).

Figure 1.1. Hemp production systems A) Intensive CBD production system. B) Fiber row crop production system, author for scale.

Hemp processing

The market class of hemp also affects the post-harvest processing requirements. Inflorescences may be processed through simple drying and then smoked, although the legality of this is variable by location. Cannabinoids and terpenes may be extracted from inflorescences through various systems, including organic solvents such as ethanol and other methods such as supercritical CO_2 (Grijó et al., 2018). Grain can be harvested with the use of a combine and threshers, especially with ones specialized for hemp, such as machinery from Hemp Harvest Works, Formation Ag, or Hempflax. Grain harvest can be

difficult due to losses form shattering during combining of wet biomass and also complications when strong fibers wrap around moving parts (Small & Marcus, 2002). These issues could be addressed through development of cultivars less likely to shatter as the crop dries at the end of the season and improvements in machinery. Once the grain has been harvested and dried, it may undergo further processing to separate the hull (pericarp) from the heart (embryo). Oil may also be extracted through pressing either the whole seed or the dehulled heart. The press cake after the oil is removed is high in protein, which can be further purified (Tang et al., 2006). Fiber processing involves cutting plants and allowing them to ret, which results in the separation of the bast from the hurd due to enzymatic action from bacteria and fungi. Retting can occur on the ground (dew retting) or in still or moving water (water retting). Dew retting is general preferred, as it returns nitrogen and other mobile nutrients to the soil, does not require large amounts of water, and is less likely to result in decomposition (Jankauskienė et al., 2015). After retting, the stems are further processed to separate the bast from the hurd, using a decorticator.

Hemp market class competitors

Hemp's three major market classes have similarities to other crops or species. Hemp fiber is a bast fiber, similar to flax, jute, or kenaf. Further work is needed to fully characterize these crops and identify optimal end uses for each, but hemp has the benefit of yielding prodigious amounts of fiber at temperate latitudes unlike other crops (Ramesh, 2018; Salentijn et al., 2015). Hemp fiber also has the potential to be a more sustainable and economical alternative to cotton, although advances in processing infrastructure are needed (Schumache et al., 2020). As a grain crop, hemp has the potential to compete with canola and soy as an oilseed crop, especially if the yields of new cultivars can reach those detailed

in Chapter 2. As a source of plant protein, hemp has among the greatest protein content of any crop, exceeding most legumes while also having a more complete amino acid profile (Callaway, 2004). As a whole seed, the grain currently has a market as a health food beside other nuts and seeds, but breeding efforts to increase yield, such as those detailed in Chapter 2, will be necessary to lower costs and gain more mainstream acceptance. Hemp seed also has a potential market as livestock feed, but this is currently not permitted in the USA due to AAFCO guidelines ("AAFCO Guidelines on Hemp in Animal Food," 2017). For cannabinoid production, no other crop produces cannabinoids to the level that *Cannabis* does, although advances in heterologous cannabinoid production in yeast may affect the CBD market (Luo et al., 2019). At present, CBD derived from hemp may have a different legal status than chemically identical CBD derived from marijuana (USDA, 2014). A homolog of cannabidiolic acid synthase (CBDAS) has been found in the closely related *Humulus lupulus*, but hop is not known to produce cannabinoids to a level justifying extraction (Padgitt-Cobb et al., 2021). Outside of the cannabinoids found in hemp, the liverwort *Radula marginata* produces the cannabinoid perrottetinene, which has some functional relationship to THC (Hussain et al., 2019). However, much more work would be required to develop a market for this liverwort.

Hemp genetic history

It is widely accepted that there is population structure within the genus *Cannabis*, but the specifics of subgenus classification are debated. The genus *Cannabis* is in the family Cannabaceae, with the closest living relative being *Humulus*, the genus containing hop. Other genera in the family Cannabaceae include *Celtis* which includes widely grown street trees, and *Parasponia*, the only non-legume to fix nitrogen through association with

rhizobia (Yang et al., 2013). It is widely accepted that there is a single species of *Cannabis*, named *Cannabis sativa*. This is supported by lack of reproductive barriers between any two plants (outside of plant sex) and relatively modest F_{st} values between subgroups (Carlson et al., 2021; Ren et al., 2021). However, some research programs posit multiple species, largely on the grounds of phenotypic differences with special attention paid to THC content (Clarke & Merlin, 2016; Hillig & Mahlberg, 2004). Genetic arguments for multiple species have also been made (Hillig, 2005). The strongest arguments have been made for two subspecies of *Cannabis sativa*, classified as *Cannabis sativa* ssp. *sativa* which includes European (and possibly Chinese) grain and fiber cultivars, and *Cannabis sativa* ssp. *indica*, which includes high-cannabinoid cultivars (McPartland, 2018). Further grouping within these subspecies has been described by multiple research programs, with somewhat contrasting results (Carlson et al., 2021; McPartland & Small, 2020; Ren et al., 2021; Sawler et al., 2015). Various domestication histories have been proposed, but they differ enough from each other to make synthesis difficult (Clarke & Merlin, 2016; McPartland & Small, 2020; Ren et al., 2021). Notably, there is debate based on the center of origin of the crop, with some groups positing an origin in Central Asia based on feral plant distribution (Clarke, 2016), and other positing an origin in China based on genomic arguments (Ren, 2021). The widespread illegality of the crop and associated regulatory and political barriers make plant collecting expeditions difficult, and it is also very likely that true wild forms no longer exist (Small, 2015). Additionally, the ease of crossbreeding erodes what may have been ancestral population structure or even true speciation.

Hemp vs marijuana vs cannabis nomenclature

Hemp is defined as *Cannabis sativa* containing less than 0.3% THC, while marijuana has a greater content of THC (USDA, 2014). This 0.3% threshold can be traced to an arbitrary distinction made in the 1970s, where it was observed that the CBD:THC ratio tended to fall in three groups, and there did not appear to be any plants in the low THC group that exceeded 0.3% THC (Small & Beckstead, 1973). While it may seem more pertinent to base distinctions between marijuana and hemp on more than this arbitrary divider, this limit has become codified in US law (USDA, 2014). Further work validated the distinct grouping of CBD:THC ratios and found it to behave as a simple codominant trait (de Meijer et al., 2003), but also found that samples from plants in the group with greatest CBD:THC ratio (known as cannabinoid chemotype III) could exceed 0.3% THC (Toth et al., 2020). Within hemp production systems, most grain and fiber cultivars would be considered *C. sativa* ssp. *sativa*, while most cultivars grown for the cannabinoids in their inflorescences would be considered *C. sativa* ssp. *indica*. Some published work refers to anything in the *indica* subspecies as "marijuana" and anything in the *sativa* subspecies as "hemp", but this conception often conflict with legal constructs and definitions, as members of either subspecies may be above or below 0.3% THC depending on the cannabinoid chemotype and total cannabinoids produced (Grassa et al., 2021; Sawler et al., 2015; Toth et al., 2020). Further complicating matters is the widespread use, including in legislation and regulations, of the term "cannabis" to refer strictly to high-THC plants and products (Steigerwald et al., 2018), with a lack of an accepted term to refer to all products produced by a plant in the genus *Cannabis*.

1.2 Hemp molecular genetics

Genomic resources

The first draft genome and transcriptome assemblies of *C. sativa* were published in 2011 and were derived from a high THC accession ('Purple Kush') and a grain cultivar ('FINOLA') (Van Bakel et al., 2011). These genome assemblies were of sufficient quality to identify the chemotype-determining genes encoding CBDAS and tetrahydrocannabinolic acid synthase (THCAS), but did not contain chromosome-level scaffolds. Extension of that work resulted in creation of chromosome-level pseudomolecule assemblies from the same cultivars through the use of long read third-generation sequencing. These assemblies allowed better resolution of the locus containing the genes encoding CBDAS and THCAS and discovery of structural variation among haplotypes which are characterized by severely reduced recombination (Laverty et al., 2019). Around this time, another chromosome-scale genome assembly derived from a high-cannabinoid cultivar with active CBDAS and no active THCAS was released on bioRxiv, which was chosen to be the representative genome in the NCBI database and was eventually published in a peer-reviewed journal a few years after the pre-print was posted (Grassa et al., 2021; Grassa et al., 2018). Various groups identified the largest chromosome as the X chromosome and developed transcriptome assemblies derived from various tissue types including male and female flowers (Braich et al., 2019). However, to date, there is no publicly available assembly of the Y chromosome. The loci containing the genes encoding CBDAS and THCAS are highly divergent and polymorphic, as are the X and Y chromosomes (Sakamoto et al., 2000). Their phenotypic effects are not highly evident until the plant is mature, so the development of molecular markers and assays allowing

early scoring of these traits would be highly beneficial for breeding and production (Toth et al., 2018). Chapters 2 and 4 of this dissertation address the development of molecular markers for these and other relevant traits.

Cannabinoid chemotypes

The cannabinoid chemotype of a *C. sativa* plant refers to the major type of cannabinoid the plant produces. There are five commonly accepted cannabinoid chemotypes, as follows: Chemotype I, characterized by a profile with mostly THC; Chemotype II, containing about equal CBD and THC; Chemotype III, dominated by CBD; Chemotype IV, with a profile predominantly CBG; and Chemotype V, with very low or no cannabinoid accumulation (Stack et al., 2021). Chemotypes I, II, and III result from the expression of a simple codominant locus, called the *B* locus, which contains genes encoding either active CBDAS or active THCAS (Mandolino et al., 2003). As demonstrated in various genome assemblies, there are several other genes with significant sequence similarity to those encoding CBDAS and THCAS, but the function and relevance of these genes is largely unknown (Laverty et al., 2019). At least some of these genes encode functional cannabichromenic acid synthases (CBCAS), but most *in planta* cannabichromenic acid (CBCA) production can be explained through product promiscuity of the CBDAS enzyme (Laverty et al., 2019; Stack et al., 2021). Product promiscuity, the phenomenon of one enzyme producing two or more products, is also the likely major source of THC in chemotype III plants, further discussed in Chapters 2 and 3 (Toth et al., 2021; Toth et al., 2020; Zirpel et al., 2018). However, as discussed in Chapter 3, there may be additional sources of CBCA outside product promiscuity of CBDAS in certain genotypes of high-cannabinoid hemp.

Decarboxylation

Cannabinoids are originally synthesized in their acidic forms, as they are generated from cannabigerolic acid, CBGA. (Figure 1.2). CBGA may also be decarboxylated into cannabigerol (CBG), and CBCA may be decarboxylated to form cannabichromene (CBC). It is common to report the total potential concentration of a given cannabinoid, which is calculated by determining the potential concentration of a neutral cannabinoid if there was complete conversion of the acidic form to the neutral form. The name of neutral form of each cannabinoid is often used as a shorthand for this potential combined total, for instance with "CBD hemp" in fact containing predominantly cannabidiolic acid (CBDA) *in planta*.

CBG dominance

Chemotype IV plants produce predominantly CBG(A), which is the first cannabinoid that is synthesized. In chemotype I, II, and III plants, the CBGA that is produced is converted to THCA or CBDA through the action of THCAS or CBDAS, and later decarboxylated to THC and CBD (Figure 1.2). Chemotype IV plants therefore have a different allele at the *B* locus without functional CBDAS or THCAS. This allele may have the gene encoding canonical THCAS inactivated (Garfinkel et al., 2021), the gene encoding canonical CBDAS inactivated (Onofri et al., 2015), or may lack genes encoding either enzyme (Ren et al., 2021). Given the mechanism of CBG(A) dominance, this is expected to be a recessive trait and some data have demonstrated this (Garfinkel et al., 2021). Additionally, as discussed in Chapter 3, in Chemotype III plants there is some variance in CBD:CBG ratio, which increases as the inflorescence matures.

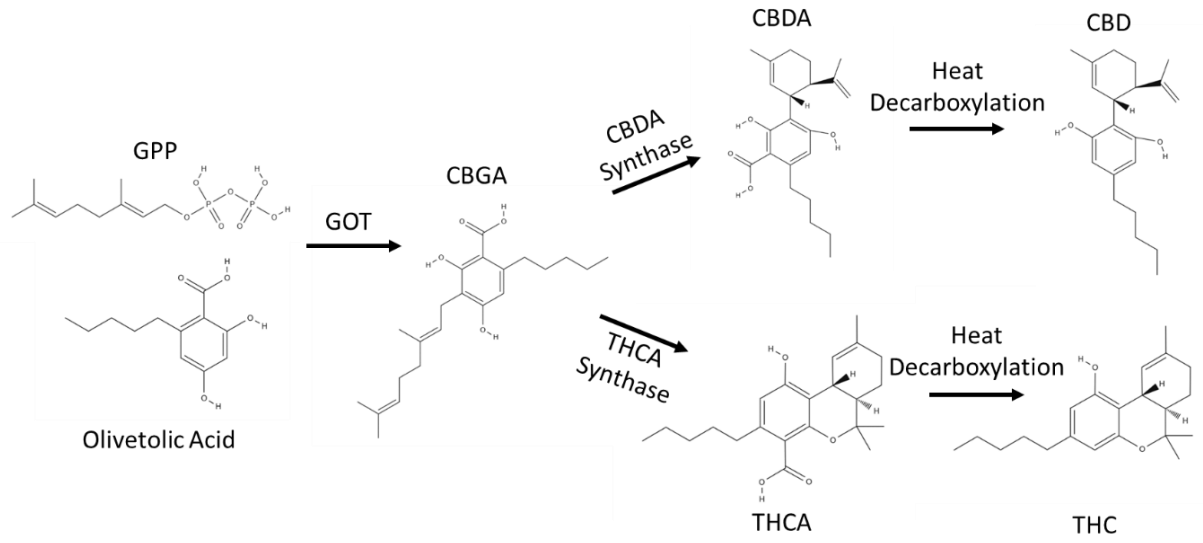

Figure 1.2. Biosynthesis of THC and CBD. GPP is geranyl pyrophosphate. GOT is geranyl-diphosphate:olivetolate geranyltransferase. Reproduced from (Toth et al., 2020).

Cannabinoid-free plants

Chemotype V plants do not produce any cannabinoids and may have any combination of the alleles described above at the *B* locus. The genetic basis of the cannabinoid-free trait has previously been considered a simple qualitative trait governed by the *O* locus acting upstream of CBGA synthesis (de Meije et al., 2009), although some other research disputes this simple genetic model (Woods et al., 2021).

Varins

While the cannabinoid chemotyping system captures much of the variation found in *C. sativa* accessions, it does not address varin cannabinoids. Varin cannabinoids have an alkyl chain with an altered length of 3 compared to 5 carbons, and they are traditionally referred to as variants of the more common parent cannabinoid. The chemotyping system still generally holds when varins are included, with the varin variant of CBGA

(Cannabigerovarinic acid, CBGVA) being converted into the varin variants of CBDA and THCA (Cannabidivaric acid or CBDVA and tetrahydrocannabinovaric acid or THCVA, respectively). The genetics of varin proportion inheritance do not appear to follow a simple genetic model, but a two gene model may capture most of the variance (Welling et al., 2019). The molecular determinants of varin production are as of yet unknown, but likely are involved with fatty acid synthesis upstream of olivetolic acid synthesis (Welling et al., 2020; Welling et al., 2019). Beyond the 3 carbon alkyl chain varin variants, variants with other alkyl chain lengths have been identified in natural samples (Citti et al., 2019; Linciano et al., 2019).

Sexual systems

Hemp is normally dioecious, with female plants having two X chromosomes and male plants having an X and a Y chromosome (Figure 1.3). However, monoecious plants that produce both female and male flowers despite having two X chromosomes have been identified (Faux et al., 2014. To date little is known about the genetic basis of monoecious trait, as it is highly environment dependent and likely polygenic (Faux et al., 2014). This trait is sometimes referred to as hermaphroditism, but this is botanically incorrect as that term refers to flowers with male and female parts, while in nearly all cases monoecious *C. sativa* expresses separate male and female flowers on the same plant (Green, 2005; Lebel-Hardenack & Grant, 1997).

Figure 1.3. *Cannabis sativa* flowers. A) Female (XX) terminal inflorescence. B) Male (XY) terminal inflorescence.

Genetic sex determination

Little is known about the molecular determinants of sex determination in *C. sativa*. It has been suggested based on work with polyploids that the XY system in *C. sativa* functions in an X:autosome ratio system (Ming et al., 2011), but the data do not preclude an active Y element or a Y:autosome ratio system (Warmke, 1944). Various genes have been found to be differentially expressed between male and female plants, but further analysis to find causal genes has not been successfully completed (Adal et al., 2021; Prentout et al., 2020).

Chemical sex determination

While *C. sativa* has a genetic sex determination system, XX plants will produce male flowers upon application of silver thiosulfate (STS), and XY plants will produce female flowers with application of ethephon (Lubell & Brand, 2018; Ram et al., 1970). The application of STS to female plants often results in viable pollen without Y chromosomes and subsequently all XX female offspring, a technique that is widely used in the high-

cannabinoid industry to produce “feminized seed” (Lubell & Brand, 2018). Ethephon treatment of male plants is less widely practiced, but has the potential to result in a YY plant, a putative “supermale” that would only produce male or supermale offspring. An attempt to make supermale plants is detailed in Chapter 2, with the result that it is unlikely that supermales are viable.

Flowering time

It is a well-known phenomenon that male hemp plants flower earlier than female plants, usually with a difference in timing of two weeks (Mediavilla et al., 1998). Appropriate flowering time is essential for any cultivar of *C. sativa*, late enough to allow sufficient time for biomass accumulation for productive agronomic yields before flowering and not so late as to risk crop loss to damaging frost or persistent rainfall. Determination of flowering time across different plants tends to have highly conserved mechanisms, (Jung & Müller, 2009; Jung et al., 2017). There are generally several conserved pathways across plants including photoperiod, age, gibberellin, and temperature sensing which interplay with a host of transcription factors and integrator genes to activate floral meristem identity genes, causing production of flowers (Wang, et al., 2019).

Flowering time in *C. sativa*

It is well established that there is a range of flowering time across *C. sativa* accessions with strong genetic basis (Carlson et al., 2021; Petit et al., 2020c; Stack et al., 2021). A genome-wide association study (GWAS) approach for flowering time in fiber accessions implicated multiple photosensory and thermosensory genes, as well as multiple transcription factors and known floral signal integrator genes (Petit et al., 2020a). Other

work has identified the likely existence of large-effect loci within elite high-cannabinoid CBD cultivars (Stack et al., 2021). Grey literature sources suggest the existence of a single photoperiod insensitivity (day neutral) locus, resulting in what are colloquially known as "Autoflower" plants (Green, 2005). A patent on the region expected to contain this locus has previously been filed (Phylos Bioscience, International Patent WO 2021/097496 A2). Some grain cultivars such as 'FINOLA' have been labeled as "Autoflowering" as well, but appear to behave differently at different latitudes with different daylengths (Callaway, 2002; Van Bakel et al., 2011). Chapter 4 discusses mapping efforts and development of molecular markers linked to two large-effect flowering time loci as well as other investigations into photoperiod insensitivity in hemp. Chapter 3 details a time-series of cannabinoid accumulation after the start of flowering under different conditions.

Molecular markers in hemp

Molecular markers have the potential to increase breeding efficiency by reducing the need for phenotyping, especially for hard to assay traits or traits only expressed at maturity (Toth et al., 2018). Molecular markers can also aid in the determination of zygosity of a gene, permitting selection of heterozygotes which may not have a unique phenotype in recessive/dominant systems. Molecular markers may consist of single nucleotide polymorphisms (SNPs), simple sequence repeats (SSRs), indels, or other genomic features, and associated molecular assays allow easy assessment of these markers. Three major traits important for hemp breeding amenable to the development and application of molecular markers are plant sex, cannabinoid chemotype, and flowering time. Cannabinoid phenotyping usually requires time-consuming and expensive assays, while plant sex and flowering time are generally only expressed at maturity. What's more, these traits are

known to be controlled by one or a few known major effect loci, so very few markers are required to explain the majority of the variance in some instances. Previous work, outlined in Chapter 2, has described molecular markers and assays for distinguishing chemotypes I, II, and III as well as plant sex, but these have generally been low-throughput assays. Chapters 2 and 4 detail development of high-throughput assays for cannabinoid chemotype, sex, and two major effect flowering time loci. Other work has developed molecular assays for one type of cannabinoid chemotype IV (Garfinkel et al., 2021), and there have also been studies published to find linked molecular markers for monoecy, total cannabinoids, and other quality and agronomic traits (Faux et al., 2016; Grassa et al., 2021; Petit et al., 2020b; Woods et al., 2021), although broadly applicable molecular assays were not developed from this work.

1.3 Breeding hemp

Types of cultivars

In plant breeding, a cultivar is a population that is stable, distinct, and uniform. Cultivars may consist of clones, inbred lines, open pollinated populations, or hybrids, each with unique benefits and detriments. Each of these types of cultivars have been used in hemp production, often differing based on production methods (Smart et al., 2022). Clones, single genotypes that are reproduced vegetatively, have been used for high-cannabinoid production as they have the benefit of uniform maturity and quality (Stack et al., 2021). Clones can also be useful for breeding purposes as a single plant may not make enough seed, but multiple clones of the same genotype could be used without sacrificing genetic uniformity. Inbred lines are the result of selfing a plant or crossing with close relatives for multiple generations, but as hemp exhibits severe inbreeding depression, inbreds are not commercially viable (Kurtz et al., 2020). However, the homogeneity and homozygosity of inbred lines makes them suitable as parents of F_1 hybrids. F_1 hybrids, the result of crosses between two genetically unique inbred parents, often show hybrid vigor and provide a modicum of intellectual property protection since it is very difficult to recapitulate the characteristics of an F_1 hybrid from its seed progeny. While clones and F_1 hybrids are common among high-cannabinoid cultivars, open pollinated populations are more common for grain and fiber production (Campbell et al., 2019). Open pollinated populations are genetically heterogenous, are usually highly heterozygous, and are easier and much less expensive to multiply than clones or hybrids.

Breeding methods

The breeding method applied depends on the intended type of cultivar to be produced. For clones, simple identification of an elite individual can suffice, assuming sufficient rooting capacity. The source of this elite individual may come from an open pollinated population, hybridization between two plants, or other methods that generate genetic diversity. For open pollinated populations, recurrent cycles of phenotypic selection over several cycles can improve traits, especially those with high heritability (Holland et al., 2003). Potential genetic gain is increased when both the male and female parents are selected, but this is difficult for certain traits, such as for selection for seed traits in male plants. The development of F_1 hybrids involves inbreeding to increase homozygosity, followed by evaluation of the progeny of two inbred lines. Often hybrid vigor can result from crosses between heterotic groups, but the existence of heterotic groups has not yet been shown in *C. sativa* (Carlson et al., 2021). Due to inbreeding depression, true F_1 hybrids are not easy to produce in hemp, leading some companies to market F_1 hybrids that are in fact segregating for major traits such as flowering time. An example of this and subsequent mapping of the major flowering time loci is presented in Chapter 4.

The breeder's equation

The breeder's equation (Equation 1.1) relates the genetic gain from selection for a trait (R) to the heritability (h^2), selection differential (S), and cycle time (T). Efforts to improve heritability, increase selection differential, and reduce cycle time have the potential to increase gain from selection over some unit time. The heritability of a trait depends on the relative environmental and genetic components of the variance of a trait, as well as the error involved in phenotyping the trait. In the case where one parent cannot be

phenotyped, as is the case of males for grain traits, the heritability is halved. The selection differential depends on the population size and phenotypic variance of a trait, with larger values allowing greater selection differentials. The cycle time is an important aspect to consider as well, as performing two cycles of selection per unit time effectively doubles the gain from selection over time. For seeded populations, the narrow sense heritability, the part of the heritability resulting from additive rather than dominant or epistatic variance, is generally more important as dominance and epistasis effects may not be transmitted to the next generation. In hybrid breeding, the general combining ability, or how well the offspring of a certain individual crossed with a population perform, is also related to the additive genetic variance.

$$R = \frac{h^2 S}{T}$$ **(Equation 1.1)**

Marker assisted breeding

Molecular markers can increase gain from selection in several ways. They can increase heritability as long as phenotypic variance is linked to known genetic variance, as no phenotyping is required beyond analysis of assay results. This means that environmental influence and phenotyping error are no longer factors in selection. They can also increase heritability through selection of male and female parents even when males do not express the trait. The use of molecular markers also allows for increased population size, as in most breeding programs, it is more expensive to gather a phenotype than it is to conduct high-throughput marker assays. Cycle time can also be decreased as selection on marker call rather than phenotype allows growth and selection of the plants under conditions which are

not related to the ultimate production area, such as a cycle of selection in a winter greenhouse for summer field production.

Drawbacks of marker assisted breeding

While there is great potential in marker assisted breeding, there are several drawbacks as well, mostly due to a lack of information of concordance of genotype and phenotype. Some traits with simple genetic architecture, such as cannabinoid chemotype and XY maleness in *C. sativa*, form discrete, qualitative groups associated with the genotypes at single loci. Once these loci are known, application of molecular markers can greatly aid breeding. However, most traits of interest, such as yield, fiber quality, cannabinoid content, and monoecious expression do not have such a simple genetic architecture and association of genotype to phenotype is much more difficult (Faux et al., 2016; Grassa et al., 2021; Petit et al., 2020a). As demonstrated in Chapter 2, the use of molecular markers can reduce costs and decrease selection cycle time, but recurrent phenotypic selection remains an effective method of population improvement.

Genomic selection

Knowledge of the molecular function of a particular variant can lead to effective molecular markers, but generating this knowledge is a difficult process. The technique of genomic selection uses statistical marker-trait associations across the whole genome without regard for molecular function, and has been used in many breeding programs with varying degrees of success, with continued concerns about cost, human and technical resources, and prediction accuracies, especially across widely varying environments (Wartha & Lorenz, 2021).

1.4 Methods of genotyping

Overview

The development of molecular markers relies on generating relevant DNA sequences. There have been three major generations of methods of DNA sequencing, with hosts of different specific technologies and consistently declining costs over time (Heather & Chain, 2016). Each sequencing generation has found a niche in modern genetics applications, depending on the throughput and information required. With declining costs per base pair, sequencing large amounts of DNA is practical and useful for many breeding programs, and the development and application of molecular markers and assays from this sequencing data can aid in the development of new cultivars. Chapter 4 details the use of second-generation sequencing and subsequent analysis to develop new useful molecular assays.

First-generation (Sanger) sequencing

The first generation of sequencing involved determining the sequence of pure populations of a single relatively small stretch of DNA, often on the order of 1000 bp or less. The technique which has become synonymous with first generation sequencing is Sanger sequencing, named after Frederick Sanger who invented the method in 1977 (Heather & Chain, 2016). Briefly, this technique is similar to polymerase chain reaction (PCR) in that it involves a primer, the DNA sequence of interest, dNTPs, and a DNA polymerase. However, unlike in PCR where the whole sequence is amplified exponentially, in Sanger sequencing the incorporation of a dideoxynucleotide results in arrest of the reaction and an indication that the complementary base to that dideoxynucleotide was present in the

template strand. In its original embodiment, four reactions could be run, each with a low concentration of dideoxynucleotide analogs of A, T, C, and G. In each reaction, the species present therefore have a known terminal nucleotide, allowing sequence determination through close examination of sequence length. This general technique is still widely used today, being a quick and inexpensive way to determine short sequences.

Second-generation (Illumina) sequencing

The second generation of sequencing involves massive multiplexing of sequencing reactions from a multi-species pool. There are several technologies associated with this second generation of sequencing, but in brief, they involve creating small segments of DNA (on the order of tens to hundreds of base pairs), binding these segments to unique locations and amplifying them into a colony or cluster to increase signal, and then determining the sequence of the DNA in the cluster. The actual determination of the DNA sequence can vary, but a common method is to flow through fluorescently labeled dNTPs with reversible terminators over the clusters in the presence of a DNA polymerase, assessing the dNTP incorporated through examination of the type of fluorescence, removing the terminator, and repeating. The read length of this technique is generally lower than for Sanger sequencing, but it can be automated with thousands or millions of reactions occurring at once. This type of sequencing was used in Chapter 4 to generate sequencing data from plants contrasting for phenotype.

Third-generation (PacBio or Oxford Nanopore) sequencing

The third generation of sequencing involves real-time analysis of nucleotides from single molecules. There are two major technologies in this space, and it continues to advance

rapidly. The first, PacBio sequencing, utilizes zero-mode waveguides to directly examine which fluorescently labeled nucleotide gets incorporated into a growing strand in real time. The second, Oxford Nanopore sequencing, involves threading DNA through a pore protein and tracking deformations in the pore protein, which differ based on nucleotide. To date third-generation sequencing platforms can generate extremely long reads, on the order of millions of nucleotides (Payne et al., 2019). However, accuracy tends to be lower than with other techniques. Despite this, the long reads can be useful as scaffolding, especially when combined with more accurate short reads (Grassa et al., 2021; Heather & Chain, 2016)

Reduced representation libraries

It is not always necessary or desirable to sequence the whole genome. There are multiple techniques that reduce the total DNA that gets sequenced, only including desirable DNA regions. There are multiple ways to achieve this. One commonly used method is genotyping-by-sequencing (GBS), which uses a methylation-sensitive restriction enzyme and subsequent size filtering to enrich for euchromatic regions (Elshire et al., 2011). Other methods commonly used include SNP chips which only include certain previously defined SNPs (which may or may not be relevant to the trait at hand), amplicon-based sequencing which involves sequencing regions of interest amplified through multiplex PCR, and hybridization-based sequencing which involves using predefined probes for regions of interest and only sequencing genomic regions that bind to these probes.

Analysis of marker-trait associations

Once sequencing has been performed, it is not always trivial to find causative variants due to genetic linkage and insufficient population size. A common technique to analyze

associations between differences in DNA sequences and traits is to SNPs with respect to a reference genome, and the allelic state of these SNPs can be tested for statistical association with a trait across multiple individuals. This can be done in a related population or across unrelated cultivars, and both approaches have been used to identify marker-trait associations in hemp (Petit et al., 2020b; Woods et al., 2021).

Bulk segregant analysis

Bulk segregant analysis, discussed in greater length in Chapter 4, is a technique that involves sequencing, SNP calling, and further statistical analysis of marker-trait associations, but instead of retrieving DNA sequences from each individual plant, the DNA from plants with common phenotype (or extreme tails) for a contrasting trait are pooled and the DNA of each pool is then sequenced. This has the benefit of reducing costs of sequencing, as libraries are constructed for only the pools, rather than from each individual plant. One must include a sufficient number of individuals in each pool to randomize the genotypes across the genome except at the locus controlling the trait in common within the pool. This technique is useful for qualitative traits where the genetic architecture of the trait is known. Here, rather than compare the allelic state of each individual, the read number or read type of each bulk can be compared. As demonstrated in Chapter 4, in the case of an F_2 population segregating for a recessive trait, one bulk will be homozygous, while the other will have a read frequency of about 33% for the allele homozygous in the recessive group. With enough individuals contributing to the bulk, all other regions of the genome should not differ between the two pools. Various statistical tests can be done to determine significance limits for regions of the genome associated with the trait using this method (Zhang & Panthee, 2020).

High-throughput genotyping methods

Once a significant marker has been determined, it is often useful to develop high-throughput assays that can be used on larger populations. These assays should be cheaper and quicker than whole-genome sequencing. One early type of molecular marker assay was gel-based determination of size differences in SSRs. SSRs are regions of the genome that have a short sequence that varies in copy number, resulting in different sized bands when flanking primers are used in a PCR reaction, which can be visualized by running reaction products on a gel. More modern high-throughput genotyping methods are gel-free, greatly increasing throughput and reducing waste (Toth et al., 2018). These methods include LAMP (Loop-Mediated Isothermal Amplification) assays, TaqMan™ assays, and PACE or KASP™. LAMP assays are dominant, room temperature colorimetric assays useful for detecting presence or absence of some DNA (Niessen & Vogel, 2010). TaqMan™ probes are modified PCR reactions that include two labeled probes which result in different fluorescent patterns depending which allele is present in the PCR reaction. PACE (PCR Allele Competitive Extension) and KASP™ (Kompetitive allele-specific PCR) are similar in process to TaqMan™ probes, but they do not require labeled probes, but rather standardized nucleotide tags on primers with alternative SNP alleles of interest at the 3' end of the primer, and result in different fluorescent patterns depending on which primer is used in the reaction. An example genotyping assay for 96 samples is shown in Figure 1.4.

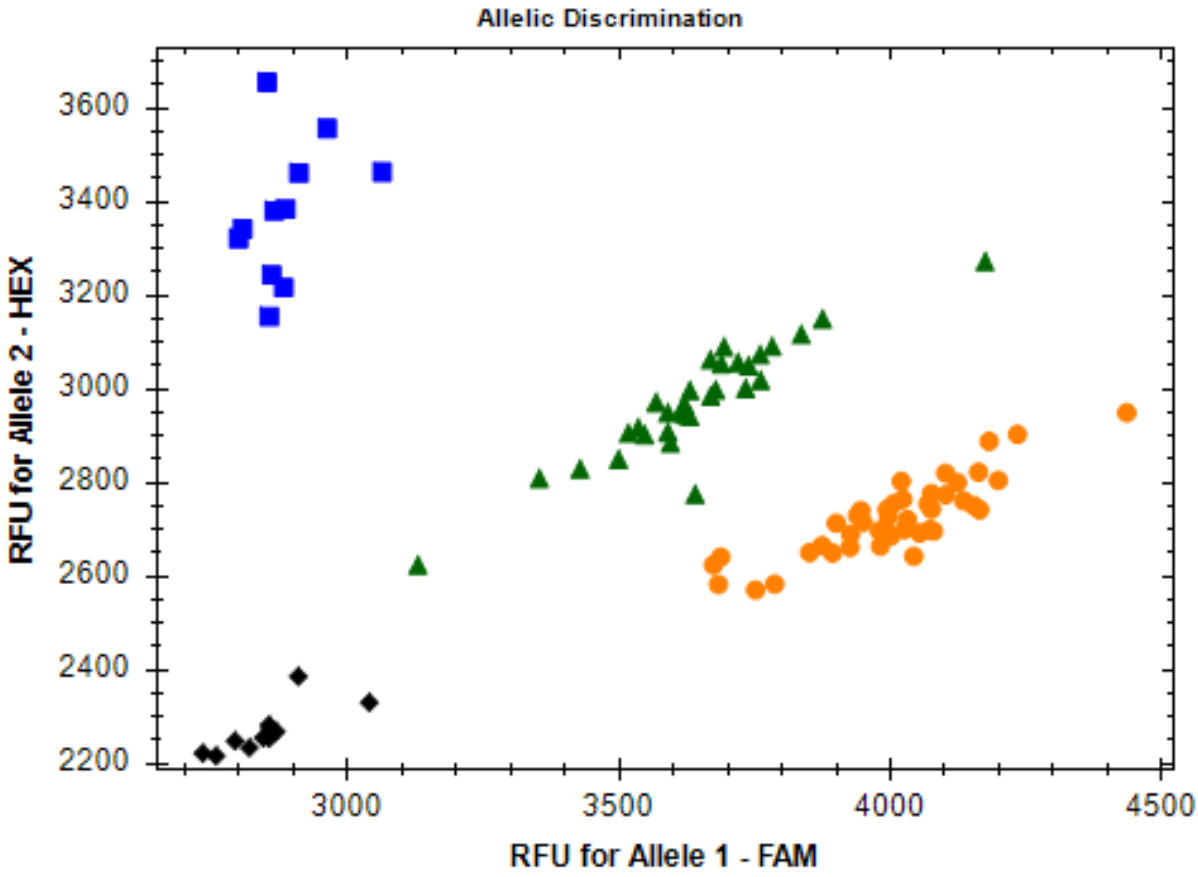

Figure 1.4. Example PACE reaction result. Each colored group corresponds to a unique allelic state, related to a distinct phenotype.

1.5 Methods of phenotyping hemp

Flowering

Cannabis sativa produces two main types of flowers: axillary or solitary flowers, which are produced as a function of age (Spitzer-Rimon et al., 2019), and terminal flowers, which are produced at the apex of the plant in inflorescences when the plant is induced to flower. The latter is more relevant for production of any market class, but to date there are competing ideas about what exact morphological stage constitutes "flowering" (Stack et al., 2021; Yang et al., 2020). A decimal growth stage for hemp has been developed (Mediavilla et al., 1998), but it offers little guidance on the definition of what a flowering plant looks like. In this dissertation, the flowering phenotype is described as in (Stack et al., 2021), where the authors noted as occurring when the internodes at the apex of the plant shorten, and in the case of female plants, pistils are visible at the apex of the plant.

Cannabinoid determination

Chapters 2 and 3 involve the determination and analysis of cannabinoids through high-performance liquid chromatography (HPLC). This technique requires drying the tissue, followed by milling, weighing, solvent extraction, filtering, and then running the filtrate on a HPLC column with liquid solvents then comparing retention times with those of known standards. The flow of solvents and column matrix allows separation of different chemical species, which can be visualized by a far-UV absorbance detector, here working at 214 nm. The absorbance can be compared to internal standards, allowing accurate determination of cannabinoid concentration. This method can resolve neutral from acidic forms of cannabinoids, unlike gas chromatography (Nahar et al., 2020). Gas chromatography works

by a similar principle as HPLC, but heat is applied to volatilize the cannabinoids, leading to potential breakdown and decarboxylation of acidic forms of cannabinoids (Lazarjani et al., 2020). As it is possible that multiple chemical species have identical chromatographic retention times, mass spectroscopy is sometimes used in conjunction with liquid or gas chromatography for more precise chemical identification.

Other methods of cannabinoid determination

While HPLC is an effective way to determine cannabinoids, it is resource intensive and requires specialized equipment and personnel. There is potential for more facile cannabinoid determination through the use of near infrared spectroscopy (NIRS). This technique involves examining the absorbance or reflectance of wavelengths in the near infrared range and using multivariate models to relate this to phenotype, in this case cannabinoid content. This technique has been successfully applied to *C. sativa* samples, but chemotype-specific models are likely needed to improve determination of minor cannabinoids (Callado et al., 2018).

1.6 Hypotheses

This work addresses several important basic questions in hemp breeding and genetics. Central hypotheses driving the work include the following:

1) Cannabinoid chemotype can be explained through a simple genetic model.
2) Supermale plants are not viable.
3) CBD:THC ratio is fixed across environments in cannabinoid chemotype III plants.
4) The “Autoflower” trait is recessive with respect to photoperiod insensitivity.
5) There are major flowering time loci outside the “Autoflower” trait.

Following are experimental results to address each of these hypotheses.

References

AAFCO Guidelines on Hemp in Animal Food. (2017, July 16 2020).

Adal, A. M., Doshi, K., Holbrook, L., & Mahmoud, S. S. (2021). Comparative RNA-Seq analysis reveals genes associated with masculinization in female *Cannabis sativa*. *Planta, 253*(1), 1-17.

Braich, S., Baillie, R. C., Jewell, L. S., Spangenberg, G. C., & Cogan, N. O. (2019). Generation of a comprehensive transcriptome atlas and transcriptome dynamics in Medicinal cannabis. *Scientific Reports, 9*(1), 1-12.

Callado, C. S., Núñez-Sánchez, N., Casano, S., & Ferreiro-Vera, C. (2018). The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: A comparative study. *Talanta, 190*, 147-157.

Callaway, J. (2002). Hemp as food at high latitudes. *Journal of Industrial Hemp, 7*(1), 105-117.

Callaway, J. (2004). Hempseed as a nutritional resource: An overview. *Euphytica, 140*(1), 65-72.

Campbell, B., Dong, Z., & McKay, J. K. (2019). Hemp genetics and genomics. *Industrial hemp as a modern commodity crop*, 92-106.

Carlson, C. H., Stack, G. M., Jiang, Y., Taşkıran, B., Cala, A. R., Toth, J. A., Philippe, G., Rose, J.K., Smart, C.D., & Smart, L. B. (2021). Morphometric relationships and their contribution to biomass and cannabinoid yield in hybrids of hemp (*Cannabis sativa*). *Journal of Experimental Botany, 72*(22), 7694-7709.

Cherney, J. H., & Small, E. (2016). Industrial Hemp in North America: Production, Politics and Potential. *Agronomy, 6*(4), 58.

Citti, C., Linciano, P., Russo, F., Luongo, L., Iannotta, M., Maione, S., Laganà, A., Capriotti, A.L., Forni, F., Vandelli, M.A. & Gigli, G. (2019). A novel phytocannabinoid isolated from *Cannabis sativa* L. with an in vivo cannabimimetic activity higher than Δ9-tetrahydrocannabinol: Δ9-Tetrahydrocannabiphorol. *Scientific Reports, 9*(1), 1-13.

Clarke, R. C., & Merlin, M. D. (2016). Cannabis domestication, breeding history, present-day genetic diversity, and future prospects. *Critical Reviews in Plant Sciences, 35*(5-6), 293-327.

Cruz, J. C., House, L. A., & Blare, T. D. (2021). Global overview of hemp production and the market of hemp-derived CBD in the US. 2021 Annual Meeting, August 1-3, Austin, Texas 312927, Agricultural and Applied Economics Association. DOI: 10.22004/ag.econ.312927

Das, L., Liu, E., Saeed, A., Williams, D.W., Hu, H., Li, C., Ray, A.E., & Shi, J. (2017). Industrial hemp as a potential bioenergy crop in comparison with kenaf, switchgrass and biomass sorghum. *Bioresource Technology, 244*, 641-649.

de Meijer, E., Hammond, K., & Sutton, A. (2009). The inheritance of chemical phenotype in *Cannabis sativa* L.(IV): cannabinoid-free plants. *Euphytica, 168*(1), 95-112.

de Meijer, E. P., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. C., Ranalli, P., & Mandolino, G. J. G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics, 163*(1), 335-346.

Divashuk, M. G., Alexandrov, O. S., Razumova, O. V., Kirov, I. V., & Karlov, G. I. (2014). Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. *PLoS One, 9*(1), e85118.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One, 6*(5), e19379.

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.

Faux, A.-M., Draye, X., Flamand, M.-C., Occre, A., & Bertin, P. (2016). Identification of QTLs for sex expression in dioecious and monoecious hemp (*Cannabis sativa* L.). *Euphytica, 209*(2), 357-376.

Fike, J. (2016). Industrial hemp: renewed opportunities for an ancient crop. *Critical Reviews in Plant Sciences, 35*(5-6), 406-424.

Fike, J., Darby, H., Johnson, B. L., Smart, L., & Williams, D. W. (2020). Industrial hemp in the USA: a brief synopsis. In *Sustainable Agriculture Reviews 42* (pp. 89-109). Springer, Cham.

Garfinkel, A. R., Otten, M., & Crawford, S. (2021). SNP in potentially defunct tetrahydrocannabinolic acid synthase is a marker for cannabigerolic acid dominance in *Cannabis sativa* L. *Genes, 12*(2), 228.

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

Grassa, C.J., Wenger, J.P., Dabney, C., Poplawski, S.G., Motley, S.T., Michael, T.P., Schwartz, C.J., & Weiblen, G. D. (2018). A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *BioRxiv*, 458083.

Green, G. (2005). *The Cannabis Breeder's Bible.* San Fransisco, USA: Green Candy Press.

Grijó, D. R., Osorio, I. A. V., & Cardozo-Filho, L. (2018). Supercritical extraction strategies using CO2 and ethanol to obtain cannabinoid compounds from *Cannabis* hybrid flowers. *Journal of CO2 Utilization, 28*, 174-180.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics, 107*(1), 1-8.

Hillig, K. W. (2005). Genetic evidence for speciation in *Cannabis* (Cannabaceae). *Genetic Resources and Crop Evolution, 52*(2), 161-180.

Hillig, K. W., & Mahlberg, P. G. (2004). A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). *American Journal of Botany, 91*(6), 966-975.

Holland, J. B., Nyquist, W. E., Cervantes-Martínez, C. T., & Janick, J. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breeding Reviews, 22*.

Horner, J., Milhollin, R., Roach, A., Morrison, C., & Schneider, R. (2019). Comparative analysis of the industrial hemp industry: guide to the evolution of the US industrial hemp industry in the global economy. Retrieved from https://extension.missouri.edu/publications/mx71

Hussain, T., Espley, R. V., Gertsch, J., Whare, T., Stehle, F., & Kayser, O. (2019). Demystifying the liverwort *Radula marginata*, a critical review on its taxonomy, genetics, cannabinoid phytochemistry and pharmacology. *Phytochemistry Reviews, 18*(3), 953-965.

Jankauskienė, Z., Butkutė, B., Gruzdevienė, E., Cesevičienė, J., & Fernando, A. L. (2015). Chemical composition and physical properties of dew-and water-retted hemp fibers. *Industrial Crops and Products, 75*, 206-211.

Jung, C., & Müller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends in Plant Science, 14*(10), 563-573.

Jung, C., Pillen, K., Staiger, D., Coupland, G., & Von Korff, M. (2017). Recent advances in flowering time control. *Frontiers in Plant Science, 7*, 2011.

Kurtz, L. E., Brand, M. H., & Lubell-Brand, J. D. (2020). Production of tetraploid and triploid hemp. *HortScience, 55*(10), 1703-1707.

Kurtz, L. E., Mahoney, J. D., Brand, M. H., & Lubell-Brand, J. D. (2020). Comparing genotypic and phenotypic variation of selfed and outcrossed progeny of hemp. *HortScience, 55*(8), 1206-1209.

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. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Research, 29*(1), 146-156.

Lazarjani, M. P., Torres, S., Hooker, T., Fowlie, C., Young, O., & Seyfoddin, A. (2020). Methods for quantification of cannabinoids: A narrative review. *Journal of Cannabis Research, 2*.

Lebel-Hardenack, S., & Grant, S. R. (1997). Genetics of sex determination in flowering plants. *Trends in Plant Science, 2*(4), 130-136.

Linciano, P., Citti, C., Luongo, L., Belardo, C., Maione, S., Vandelli, M.A., Forni, F., Gigli, G., Laganà, A., Montone, C.M., & Cannazza, G. (2019). Isolation of a high-affinity cannabinoid for the human CB1 receptor from a medicinal *Cannabis sativa* variety: Δ9-tetrahydrocannabutol, the butyl homologue of Δ9-tetrahydrocannabinol. *Journal of Natural Products, 83*(1), 88-98.

Lubell, J. D., & Brand, M. H. (2018). Foliar sprays of silver thiosulfate produce male flowers on female hemp plants. *HortTechnology, 28*(6), 743-747.

Luo, X., Reiter, M., d'Espaux, L., Wong, J., Denby, C., Lechner, A., Zhang, Y., Grzybowski, A., Harth, S., Lin, W., Lee, H., Yu, C., Shin, J., Deng, K., Benites, V., Wang, G., Baidoo, E., Chen, Y., Dev, I., Petzold, C., & Keasling, J. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature, 567*(7746), 123.

Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., & de Meijer, E. (2003). Qualitative and quantitative aspects of the inheritance of chemical phenotype in *Cannabis*. *Journal of Industrial Hemp, 8*(2), 51-72.

McPartland, J. M. (2018). *Cannabis* systematics at the levels of family, genus, and species. *Cannabis and Cannabinoid Research, 3*(1), 203-212.

McPartland, J. M., & Small, E. (2020). A classification of endangered high-THC cannabis (*Cannabis sativa subsp. indica*) domesticates and their wild relatives. *PhytoKeys, 144*, 81.

Mediavilla, V., Jonquera, M., Schmid-Slembrouck, I., & Soldati, A. (1998). Decimal code for growth stages of hemp (*Cannabis sativa* L.). *Journal of the International Hemp Association, 5*(2), 65.

Ming, R., Bendahmane, A., & Renner, S. S. (2011). Sex chromosomes in land plants. *Annual Review of Plant Biology, 62*, 485-514.

Nahar, L., Onder, A., & Sarker, S. D. (2020). A review on the recent advances in HPLC, UHPLC and UPLC analyses of naturally occurring cannabinoids (2010–2019). *Phytochemical Analysis, 31*(4), 413-457.

National Hemp Report. (2022). Retrieved from https://downloads.usda.library.cornell.edu/usda-esmis/files/gf06h2430/xd07hw825/v692v917t/hempan22.pdf

Niessen, L., & Vogel, R. F. (2010). Detection of *Fusarium graminearum* DNA using a loop-mediated isothermal amplification (LAMP) assay. *International Journal of Food Microbiology, 140*(2-3), 183-191.

Onofri, C., de Meijer, E. P., & Mandolino, G. (2015). Sequence heterogeneity of cannabidiolic-and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry, 116*, 57-68.

Padgitt-Cobb, L.K., Kingan, S.B., Wells, J., Elser, J., Kronmiller, B., Moore, D., Concepcion, G., Peluso, P., Rank, D., Jaiswal, P., Henning, J., & Hendrix, D. A. (2021). A draft phased assembly of the diploid Cascade hop (*Humulus lupulus*) genome. *The Plant Genome, 14*(1), e20072.

Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics, 35*(13), 2193-2198.

Petit, J., Salentijn, E. M., Paulo, M.-J., Denneboom, C., & Trindade, L. M. (2020a). Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa* L.): A genome-wide association study. *Frontiers in Plant Science, 11*, 1704.

Petit, J., Salentijn, E. M., Paulo, M.-J., Denneboom, C., van Loo, E. N., & Trindade, L. M. (2020b). Elucidating the genetic architecture of fiber quality in hemp (*Cannabis sativa* L.) using a genome-wide association study. *Frontiers in Genetics, 11*, 1101.

Petit, J., Salentijn, E. M., Paulo, M.-J., Thouminot, C., van Dinter, B. J., Magagnini, G., Gusovius, H., Tang, K., Amaducci, S., Wang, S., Uhrlaub, B., Müssig, J., & Trindade, L. M (2020c). Genetic variability of morphological, flowering, and biomass quality traits in hemp (*Cannabis sativa* L.). *Frontiers in Plant Science, 11*, 102.

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

Ram, H. M., & Jaiswal, V. S. (1970). Induction of female flowers on male plants of *Cannabis sativa* L. by 2-chloroethanephosphonic acid. *Experientia, 26*(2), 214-216

Ramesh, M. (2018). Hemp, jute, banana, kenaf, ramie, sisal fibers. In *Handbook of Properties of Textile and Technical Fibres* (pp. 301-325): Elsevier.

Razumova, O. V., Alexandrov, O. S., Divashuk, M. G., Sukhorada, T. I., & Karlov, G. I. (2016). Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution. *Protoplasma, 253*(3), 895-901.

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.

Sakamoto, K., Ohmido, N., Fukui, K., Kamada, H., & Satoh, S. (2000). Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Molecular Biology, 44*(6), 723-732.

Salentijn, E. M., Zhang, Q., Amaducci, S., Yang, M., & Trindade, L. M. (2015). New developments in fiber hemp (*Cannabis sativa* L.) breeding. *Industrial Crops and Products, 68*, 32-41.

Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E. & Myles, S. (2015). The genetic structure of marijuana and hemp. *PLoS One, 10*(8), e0133292.

Schumacher, A. G. D., Pequito, S., & Pazour, J. (2020). Industrial hemp fiber: A sustainable and economical alternative to cotton. *Journal of Cleaner Production, 268*, 122180.

Sharma, V., Srivastava, D. K., Gupta, R. C., & Singh, B. (2015). Abnormal meiosis in tetraploid (4x) *Cannabis sativa* (L.) from Lahaul-Spiti (cold desert higher altitude Himalayas): A neglected but important herb. *Journal of Biological and Chemical Chronicles, 2*, 38-42.

Small, E. (2015). Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. *The Botanical Review, 81*(3), 189-294.

Small, E., & Beckstead, H. (1973). Cannabinoid phenotypes in *Cannabis sativa*. *Nature, 245*(5421), 147-148.

Small, E., & Marcus, D. (2002). Hemp: A new crop with new uses for North America. *Trends in new crops and new uses, 24*(5), 284-326.

Smart, L.B., Toth, J.A., Stack, G.M., Monserrate, L., and Smart, C.D. (2022) "Breeding of hemp (*Cannabis sativa*)." In: Goldman, I. ed. Plant Breeding Reviews Vol. 46, New York, NY; Wiley, pp. XX-XX. (in press).

Spitzer-Rimon, B., Duchin, S., Bernstein, N., & Kamenetsky, R. (2019). Architecture and florogenesis in female *Cannabis sativa* plants. *Frontiers in Plant Science, 10*, 350.

Stack, G. M., Toth, J. A., Carlson, C. H., Cala, A. R., Marrero-González, M. I., Wilk, R. L., Gentner, D. R., Crawford, J.L, Philippe, G,. Rose, J. K., Viands, D.R., Smart, C.D., & Smart, L.B. (2021). Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy, 13*(4), 546-561.

Steigerwald, S., Wong, P. O., Khorasani, A., & Keyhani, S. (2018). The form and content of cannabis products in the United States. *Journal of General Internal Medicine, 33*(9), 1426-1428.

Tang, C.-H., Ten, Z., Wang, X.-S., & Yang, X.-Q. (2006). Physicochemical and functional properties of hemp (*Cannabis sativa* L.) protein isolate. *Journal of Agricultural and Food Chemistry, 54*(23), 8945-8950.

Tang, K., Struik, P., Yin, X., Thouminot, C., Bjelková, M., Stramkale, V., & Amaducci, S. (2016). Comparing hemp (*Cannabis sativa* L.) cultivars for dual-purpose production under contrasting environments. *Industrial Crops and Products, 87*, 33-44.

Toth, J., Pandurangan, S., Burt, A., Mitchell Fetch, J., & Kumar, S. (2018). Marker-assisted breeding of hexaploid spring wheat in the Canadian prairies. *Canadian Journal of Plant Science, 99*(2), 111-127.

Toth, J. A., Smart, L. B., Smart, C. D., Stack, G. M., Carlson, C. H., Philippe, G., & Rose, J. K. (2021). Limited effect of environmental stress on cannabinoid profiles in high-cannabidiol hemp (*Cannabis sativa* L.). *GCB Bioenergy, 13*(10), 1666-1674.

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K. & Smart, L.B. (2020). Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy, 12*(3), 213-222.

USDA. (2014). *Farm Bill. Legitimacy of Industrial Hemp Research.* Washington, DC: US Gov. Printing Office.

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.

Wang, X., Liu, Q., He, W., Lin, C., & Wang, Q. (2019). Characterization of Flowering Time Mutants. In *Phytochromes* (pp. 193-199): Springer.

Warmke, H. E. (1944). Polyploid investigation. *Year book of the Carnegie Institute of Washington, 43*, 135-139.

Wartha, C. A., & Lorenz, A. J. (2021). Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. *Crop Breeding and Applied Biotechnology, 21*.

Welling, M. T., Liu, L., Kretzschmar, T., Mauleon, R., Ansari, O., & King, G. J. (2020). An extreme-phenotype genome-wide association study identifies candidate cannabinoid pathway genes in *Cannabis*. *Scientific Reports, 10*(1), 1-14.

Welling, M. T., Liu, L., Raymond, C. A., Kretzschmar, T., Ansari, O., & King, G. J. (2019). complex patterns of cannabinoid Alkyl Side-chain inheritance in *Cannabis*. *Scientific Reports, 9*(1), 1-13.

Westerhuis, W., van Delden, S., van Dam, J., Marinho, J. P., Struik, P., & Stomph, T. (2019). Plant weight determines secondary fibre development in fibre hemp (*Cannabis sativa* L.). *Industrial Crops and Products, 139*, 111493.

White, C. M. (2019). A review of human studies assessing cannabidiol's (CBD) therapeutic actions and potential. *The Journal of Clinical Pharmacology, 59*(7), 923-934.

Woods, P., Campbell, B. J., Nicodemus, T. J., Cahoon, E. B., Mullen, J. L., & McKay, J. K. (2021). Quantitative trait loci controlling agronomic and biochemical traits in *Cannabis sativa*. *Genetics, 219*(2), iyab099.

Yang, M.-Q., van Velzen, R., Bakker, F. T., Sattarian, A., Li, D.-Z., & Yi, T.-S. (2013). Molecular phylogenetics and character evolution of Cannabaceae. *Taxon, 62*(3), 473-485.

Yang, R., Berthold, E., McCurdy, C. R., da Silva Benevenute, S., Brym, Z. T., & Freeman, J. H. (2020). Development of Cannabinoids in Flowers of Industrial Hemp (*Cannabis sativa* L.)—a Pilot Study. *Journal of Agricultural and Food Chemistry*.

Zhang, J., & Panthee, D. R. (2020). PyBSASeq: a simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data. *BMC Bioinformatics, 21*(1), 1-10.

Zhang, M., Anderson, S. L., Brym, Z. T., & Pearson, B. J. (2021). Photoperiodic Flowering Response of Essential Oil, Grain, and Fiber Hemp (*Cannabis sativa* L.) Cultivars. *Frontiers in Plant Science*, *12*(1498). doi:10.3389/fpls.2021.694153

Zhao, X., Wei, X., Guo, Y., Qiu, C., Long, S., Wang, Y., & Qiu, H. (2021). Industrial Hemp—an Old but Versatile Bast Fiber Crop. *Journal of Natural Fibers*, 1-14.

Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *Journal of Biotechnology, 284*, 17-26.

Chapter 2: Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L.[1]

2.1 Chapter Overview

High-throughput molecular markers are an important part of modern plant breeding, allowing more facile analysis of difficult to assay traits (Toth et al., 2018). They are especially useful for traits with underlying major effect genes, including qualitative traits, which form discrete groups. They may be less useful for more quantitative traits governed by many distinct loci, but can improve speed and accuracy of breeding.

There are relatively few molecular markers available for *Cannabis*, and those that are available have not been generally validated over broad germplasm. (Pacifico et al., 2006). This is likely due in a large part to the previously illicit nature of the crop, although with increasing legalization there is great potential for the development of new Cannabis cultivars using modern tools (*Global Cannabis Report: 2019 Industry Outlook*, 2019).

There are several qualitative traits that have been previously identified in hemp, forming discrete groups. This include cannabinoid chemotype (de Meijer et al., 2003) and plant sex (Faux et al., 2014). While some molecular markers have been previously developed (as

[1] Most of Chapter 2 was published as Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K. & Smart, L.B. (2020). Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy*, *12*(3), 213-222.

described in Chapter 2) these markers are not high-throughput or widely validated. What's more, they were not always codominant, making scoring of heterozygotes difficult.

The development of high-throughput molecular markers for sex and chemotype opens several doors, including assessment of ratios of X and Y chromosomes to test the existence of YY "supermale" plants, explanation of "spikes" in THC content in high CBD varieties, and development of novel germplasm that is exclusively chemotype 3 from a large segregating base population.

To date there has been little research into distinguishing the genetic and environmental effects that contribute to THC production. In the following work, a major genetic component for THC production is shown to be segregating in a plurality of cultivars, including cultivars grown for CBD, grain, and fiber. Interestingly, a comparison of two sites of a trial utilizing cultivars grown for CBD production showed little difference in THC content or CBD:THC ratio, suggesting differences in these important traits are primary genetic, even when grown under contrasting field conditions. The environmental effect on THC production is further explored in Chapter 3.

2.2 Abstract

Hemp (*Cannabis sativa* L.) is an emerging dioecious crop grown primarily for grain, fiber, and cannabinoids. There is good evidence for medicinal benefits of the most abundant cannabinoid in hemp, cannabidiol (CBD). For CBD production, female plants producing CBD but not tetrahydrocannabinol (THC) are desired. I developed and validated high-throughput PACE (PCR Allele Competitive Extension) assays for *C. sativa* plant sex and cannabinoid chemotype. The sex assay was validated across a wide range of germplasm and resolved male plants from female and monoecious plants. The cannabinoid chemotype assay revealed segregation in hemp populations, and resolved plants producing predominantly THC, predominantly CBD, and roughly equal amounts of THC and CBD. Cultivar populations that were thought to be stabilized for CBD production were found to be segregating phenotypically and genotypically. Many plants predominantly producing CBD accumulated more than the current US legal limit of 0.3% THC by dry weight. These assays and data provide potentially useful tools for breeding and early selection of hemp. The utility of this work was shown through the application of molecular markers in the development of a new high-yielding cultivar, as well as to the identification of sex chromosome distortion in an XY × XY genotypic male intercross.

2.3 Introduction

Hemp (*Cannabis sativa* L.) is a multi-use crop grown primarily for grain, fiber, and cannabinoids. In recent years, there has been a resurgence in the study of hemp in the United States and across the world in fields including genomics (Grassa et al., 2021; Laverty et al., 2019), agronomics (Campbell et al., 2019), and novel end-uses (Turner et al., 2019; Wang & Xiong, 2019). The market for hemp (defined as <0.3% tetrahydrocannabinol; THC by dry weight) and marijuana (>0.3% THC), is expected to surpass $26 billion in the US by 2025 (*Global Cannabis Report: 2019 Industry Outlook*, 2019). The major market for *C. sativa* is as a source of cannabinoids, the two most abundant of which are THC and cannabidiol (CBD). There is growing evidence of the different ways that CBD and THC interact with human endocannabinoid signaling pathways (Citti et al., 2015). THC is a psychoactive compound and is currently listed as a Schedule 1 controlled substance in the United States, and CBD is now formulated as an approved drug marketed as Epidiolex by GW Pharmaceuticals (Cambridge, UK).

Plant sex and cannabinoid chemotype are essentially qualitative traits that are important for hemp producers and breeders. Maximal production of CBD occurs in unpollinated female hemp plants. To grow and develop legally compliant hemp cultivars (<0.3% THC by dry weight), genetic propensity for THC production must be known. However, sex and cannabinoid chemotype are difficult to phenotype in young plants. Until the onset of flowering, male and female plants are phenotypically indistinguishable, and immature plants produce relatively small quantities of cannabinoids. Cannabinoid chemotype of immature plants may also not reflect the cannabinoid profile of mature plants (de Meijer et

al., 2009; Pacifico et al., 2008). Molecular markers can address these challenges, since DNA from very young plants can be used in reliable genotype assays.

The stalked capitate trichomes of unpollinated female inflorescences have the highest concentration of cannabinoids (Livingston et al., 2019; Mahlberg & Kim, 2004). CBDA and THCA, the acidic precursors to CBD and THC (also referred to as Δ^9-THC), are produced from cannabigerolic acid (CBGA) by CBDA synthase (CBDAS) and THCA synthase (THCAS), respectively. CBGA is produced from the condensation of olivetolic acid and geranyl pyrophosphate (GPP) by the enzyme geranyl-diphosphate:olivetolate geranyltransferase (also known as aromatic prenyltransferase, AP, or GOT) (Fellermeier et al., 2001; Fellermeier & Zenk, 1998). CBD and THC are generally derived from their corresponding acids through non-enzymatic decarboxylation, enhanced by heat and light (Figure 2.1) (Hanuš et al., 2016; Smith & Vaughan, 1977). The decarboxylated forms (CBD and THC) are biologically active for medicinal or recreational use, while the acidic precursors do not share the same activity (Citti et al., 2018).

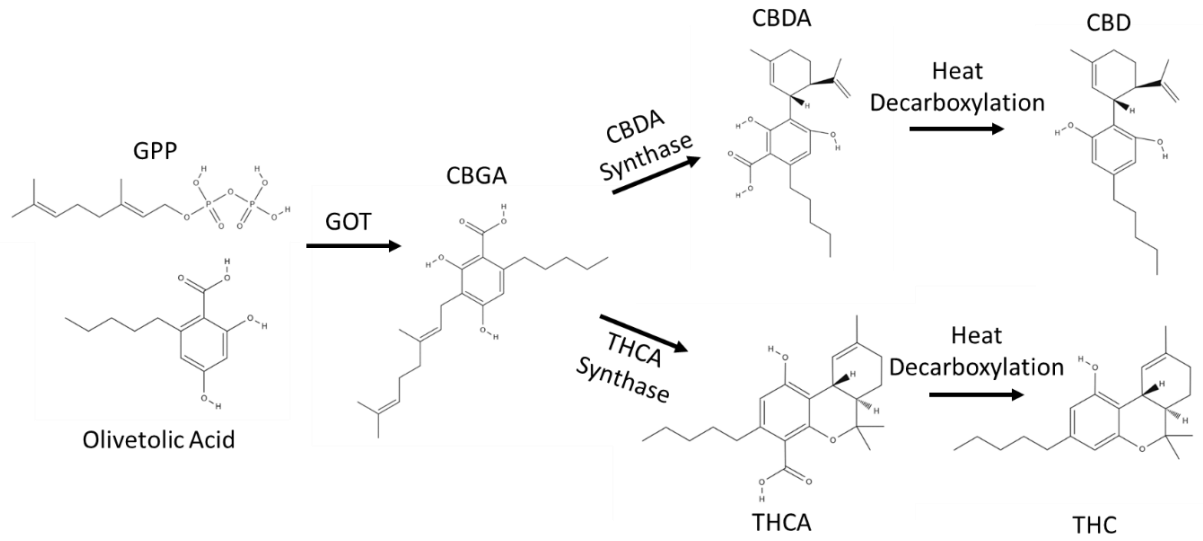

Figure 2.1. Biosynthesis of THC and CBD. GPP is geranyl pyrophosphate. GOT is geranyl-diphosphate:olivetolate geranyltransferase.

The genetic structure of CBDAS and THCAS have been recently elucidated (Grassa et al., 2021; Laverty et al., 2019). While the genes are highly alike, sharing 84% amino acid identity (Onofri, et al., 2015), they are not allelic or at equivalent loci. However, chromosomal scaffolds containing these genes are physically linked in repulsion and not highly homologous, leading to low recombination (Laverty et al., 2019). Consequently, cannabinoid chemotype inheritance can largely be modeled as monogenic, with plants producing predominantly THC (chemotype I, B_T/B_T), about equal amounts of CBD and THC (chemotype II, B_T/B_D), or predominantly CBD (chemotype III, B_D/B_D) (de Meijer et al., 2003; Mandolino et al., 2003; Small & Beckstead, 1973).

Cannabis sativa is usually dioecious, having male and female flowers produced on separate plants. Male plants have heteromorphic X and Y sex chromosomes, while female and monecious plants have two X chromosomes (Divashuk et al., 2014; Razumova et al, 2016). Genetic sex determination is believed to function through the X: autosome ratio (Ainsworth, 2000; Vyskot & Hobza, 2004), although this mechanism is not fully

understood (Divashuk et al., 2014; Ming, et al., 2011). Despite having well-defined heteromorphic sex chromosomes (Sakamoto et al., 1998), environment can play a large role in sex determination in *Cannabis sativa* (Schaffner, 1921). Factors such as altered hormones (Lubell & Brand, 2018; Ram & Jaiswal, 1970), daylength (Schaffner, 1921), and autosomal genes (Faux et al., 2014) have been shown to influence sex expression. Through manipulation of hormones, it is possible to create all-female progeny using pollen produced by a female (XX) plant induced to produce male flowers and pollen (Lubell & Brand, 2018). From a commercial perspective, these so-called "feminized" seeds are generally more expensive to produce than normal dioecious seeds due to the additional work required for their production and the market demand for all-female seed lots. Female flowers can also be produced on male plants using Ethephon, with resultant viable seeds (Ram & Jaiswal, 1970). The seeds from such a plant, assuming the other parent was an XY male, theoretically should segregate ¼ XX, ½ XY, ¼ YY, although it is not known if YY plants are viable. If such a plant were viable, it would be a "supermale", having only male (XY) or supermale (YY) offspring.

The absence of a Y chromosome does not appear to be sufficient to ensure a total lack of production of male flowers (Faux et al., 2014; Menzel, 1964; Razumova et al., 2016). Some *Cannabis* plants are monoecious, producing both male and female flowers on the same plant (Menzel, 1964). While monoecious plants are commonly referred to as hermaphrodites, the botanical definition of hermaphrodite requires staminate and carpellate parts on the same flower (Lebel-Hardenack & Grant, 1997), a phenomenon rarely seen in *C. sativa*.

There are several published marker assays for sex in *C. sativa*, but to date none are sufficiently high-throughput to be used efficiently in a breeding program. For sex determination, the male-associated sequences MADC1, MADC2, MADC3, MADC4, MADC5, and MADC6 have been published (Mandolino et al., 1999; Sakamoto et al., 2005; Sakamoto et al., 1995; Törjék et al., 2002). MADC1 is a hybridization-based probe, which is not tractable for either breeders or producers (Sakamoto et al., 1995). MADC3 and MADC4 are Random Amplified Polymorphic DNA (RAPD) markers, which have issues with reproducibility and interpretation (Sakamoto et al., 2005). MADC5 and MADC6 were reported to be non-diagnostic and are gel-based, which significantly reduces throughput (Törjék et al., 2002). MADC2 is a marker that had been shown to be diagnostic, but is also gel-based (Divashuk et al., 2014; Faux et al., 2014; Mandolino et al., 1999). Male-associated Amplified Fragment Length Polymorphism (AFLP) markers have also been reported (Flachowsky et al., 2001; Peil et al., 2003). Another report described female-specific markers, but given the XY sex-determination system and low number of samples tested, this was likely a false positive (Shao et al., 2003). None of the Y-associated markers previously developed were reported to be codominant with the X chromosome, making evaluation of the existence of YY plants difficult.

Multiple marker assays to determine cannabinoid chemotype have been described involving a range of assay technologies (Borna et al., 2017; Kojoma et al., 2006; Pacifico et al., 2006; Rotherham & Harbison, 2011). However, these published assays are low-throughput or require expensive instrumentation. There have been extensive studies on sequence and copy number variation of THCAS and CBDAS, but sequence-function relationships are not always clear (Onofri et al., 2015). An inactive form of CBDAS

appears to be conserved among chemotype I plants (Van Bakel et al., 2011; Weiblen et al., 2015); however the published assay conditions poorly resolve chemotype II plants, limiting the diagnostic use of this protocol.

PACE (PCR Allele Competitive Extension, 3CR Bioscience Ltd, Essex, UK) is a high-throughout, fluorescence-based marker system that can interrogate SNPs, indels, or other polymorphic DNA features. Fluorescence-based marker systems such as PACE, KASP, TaqMan, and the recently developed RhAMP are estimated to be 45 times faster than gel-based systems (Rasheed et al., 2016; Toth et al., 2018). PACE has the further advantage of lower cost while retaining simple codominance, unlike loop-mediated isothermal amplification (LAMP) (Notomi et al., 2000) or direct sequencing (Weiblen et al., 2015). PACE assays, unlike some other marker systems, require a fluorescent plate reader or qPCR system to score, limiting field-based testing. However, the ease and speed of PACE assays is well suited for a breeding or advanced production systems. Here, I used publicly available sequence information to develop reliable, high throughput PACE assays that are highly predictive of sex and cannabinoid chemotype phenotypes in *C. sativa*.

2.4 Materials and methods

For sex testing, dioecious seeds from CBD cultivars of *C. sativa* were started in a greenhouse in plugs of soilless mix in 2019 and cultivated under a 18L:6D light regime. DNA was extracted from leaves harvested from 2-week old plants. Genotyped female plants were planted in outdoor field trials, while males were transplanted to two-gallon pots and kept in the greenhouse. Plant sex was noted at the onset of flowering.

For cannabinoid chemotype marker testing, cannabinoid data and tissue from the 2018 Cornell CBD Hemp Cultivar Trials was used. More information about the 2018 Cornell CBD Hemp Cultivar Trial is available at https://hemp.cals.cornell.edu/resources/reports-factsheets/

Hemp seeds were obtained from multiple sources (Table 2.6). Plants for CBD production were started in the greenhouse under a 18L:6D light regime with males or monoecious plants removed based on phenotype. The trials were located on Cornell University farms in upstate New York: one at Bluegrass Lane Turf and Ornamental Research Farm (Ithaca, NY) and the other at Cornell AgriTech Gates West Farm (Geneva, NY). Late season rainfall lead to saturated field conditions in the Geneva location during flowering.

The top 10 cm of mature female plants were harvested by hand at maturity and dried in a greenhouse. The inflorescence was then milled using a Magic Bullet food grinder (Homeland Housewares, Los Angeles, CA) and stored at 4°C until analysis. For each sample, 50 mg of dried, milled tissue was mixed with 1.5mL ethanol by high-speed shaking at room temperature with a Tissuelyser (Qiagen), and filtered through a SINGLE StEP PTFE Filter Vial (Thomson, Oceanside, CA). The resultant liquid was directly subjected to HPLC analysis (Dionex UltiMate 3000; Thermo Fisher, Waltham, MA) with biphenyl-4-carboxylic acid (BPCA) as an internal standard, using a Phenomenex (Torrance, CA) Kinetex 2.6 µm Polar 100 Å column 150 x 4.6 mm heated at 35°C. Samples were injected and eluted at 1.2 mL min^{-1} over a 6 min gradient, from 65% acetonitrile, 0.1% formic acid, to 80% acetonitrile, 0.1% formic acid, followed by a 4 min isocratic step. Absorbance was measured at 214 nm. The following standards were used as calibrants: THCA, Δ^9-THC, CBDA, CBD, cannabichromenic acid (CBCA),

cannabichromene (CBC), CBGA, CBG, cannabinol (CBN), and Δ^8-THC (Sigma Aldrich, St. Louis, MO).

DNA was isolated using a high-throughput modified CTAB method utilizing PALL (Port Washington, NY) DNA binding plates (modified from Doyle & Doyle, 1987). PACE reactions were run according to the manufacturer's (3CR Bioscience Ltd, Essex, UK) instructions with five extra final cycles on a Bio-Rad (Hercules, CA) C1000 Touch thermocycler. A Bio-Rad CFX96 qPCR machine was used as a fluorescent plate reader, and the data were analyzed using Bio-Rad CFX Maestro software. The primers (Table 2.1) were designed to have male and THC-dominant plants result in HEX fluorescence.

To create a XY × XY population, a population of 'Logan×OP' was screened using CSP-1 for male plants and grown under 16:8 light. About 6 weeks after planting, the male plants had any male primordia pinched back, and were moved to 12:12 light. 12 male plants were treated with 1920 ppm aqueous Ethephon (2-chloroethylphosphonic acid) spray at the onset of flowering (Ram & Jaiswal, 1970), while 12 male plants were untreated. 12 female plants were also retained. The Ethephon-treated plants produced female flowers and subsequently set seed, which were grown under greenhouse conditions in plugs. DNA was harvested and tested with CSP-2 for existence of YY supermale plants. DNA from XY plant-derived seeds was also tested, as were seedlings from the XX females from the same population.

Table 2.1. Primer sequences for sex and cannabinoid PACE assays.

Name	Sequence	Allele
CSP-1-FAM	GAAGGTGACCAAGTTCATGCTAGC TTGAAATGAGATGTCAAACC	Female
CSP-1-HEX	GAAGGTCGGAGTCAACGGATTGAG CTTGAAATGAGATGTCAAACT	Male
CSP-1-COMMON	GCAGCAGACCTGGGCATATAG	
CSP-2-FAM	GAAGGTGACCAAGTTCATGCTT GCAGATTCGTATGTGGCAACA	Y Chromosome
CSP-2-HEX	GAAGGTCGGAGTCAACGGATTT GCAGATTCGTATGTGGCAACG	X Chromosome
CSP-2-COMMON	ATAGCCGCTGCTGGAGTT	
CCP-1-FAM	GAAGGTGACCAAGTTCATGCTA TTAGACTGGTTGCTGTCCCAAA	B_D
CCP-1-HEX	GAAGGTCGGAGTCAACGGATTA TTAGACTGGTTGCTGTCCCAAC	B_T
CCP-1-COMMON	ACTTGACAAGCTCATGTATCTCCA	

Multiple linear regression analysis was performed in RStudio version 1.1.463 running R version 3.5.1 (R Core Team, 2018). Cannabinoid chemotype allele score was numerically coded as [-1,0,1] while all other variables were coded as factors. Total potential CBD and THC were calculated by summing the concentration of the decarboxylated form with the concentration of CBDA or THCA multiplied by 0.877.

For application of CCP-1 in a breeding context, approximately 700 individuals of the cultivar population ‘Han-NW’ (CN Kenaf and Hemp Seed Farm) were direct seeded in July 2020 in a field location in Geneva NY. DNA was extracted from each plant and CCP-1 was run on each individual, with all chemotype I and II plants discarded before flowering. Seed and a shoot tip for cannabinoid content were collected at the end of the

season, and a selected bulk based on yield, seed size, and varin content was grown in the 2021 Cornell Hemp Grain and Fiber Trial in Ithaca, NY in a replicated plot design. Plants from this population were harvested by hand and seed was dried to 6-8% moisture for yield determination. Grain from other cultivars was harvested by combine, a technique noted to reduce harvested yield by ~30% (Vogl et al., 2004). More data concerning the 2021 Cornell Hemp Grain and Fiber Trial will be available at https://hemp.cals.cornell.edu/resources/reports-factsheets/.

2.5 Results

Sex determination assay development

A novel high-throughput assay for the Y chromosome was developed based on the previously identified male-specific MADC6 sequence (Genbank AF364955.1). To develop the assay, the MADC6 sequence was compared to the 'FINOLA' (Genbank GCA_003417725.1) and 'Purple Kush' (Genbank GCA_000230575.1) genomes (Van Bakel et al., 2011) using BLAT (BLAST-Like Alignment Tool) on the *C. sativa* Genome Browser Gateway (UCSC Genome Browser, University of California). A PACE assay, named CSP-1, was designed based on a SNP between the sequences (Figure 2.2).

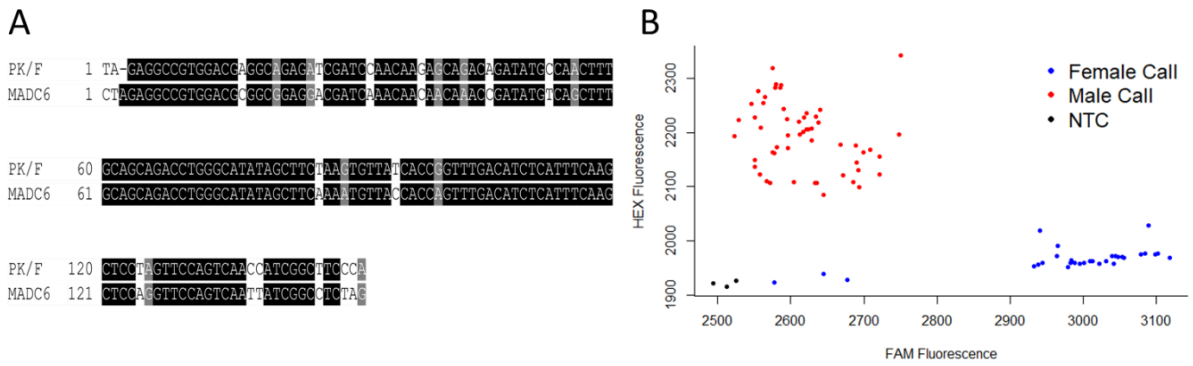

Figure 2.2. CSP-1 development. A) Sequence comparison of closest BLAT result of the 'Purple Kush'/'FINOLA' (PK/F) genomes and MADC6. B) Representative chromatogram of CSP-1. Some females clustered with the No Template Control (NTC).

Sex assay validation

The CSP-1 assay was used to test a total of 2,170 individual plants of 14 cultivars. In all but one population the genetic male:female ratio fit the expected 1:1 model (Chi-square p >0.05, Table 2.1). The individuals genetically scored as females were planted in field trials and the individuals genetically scored as males were discarded or moved to greenhouse conditions.

Table 2.1. Sex distribution and germination percentage. Value in red has Chi-square p <0.05.

	Males	Females	% Female	Chisquared (vs 1:1)	Germination %	Monoecious individuals
'NY Cherry'	64	33	34.0%	0.002	69%	No
'RN17'	39	55	58.5%	0.099	97%	Yes
'RN19'	40	48	54.5%	0.394	97%	No
'Cherry 307'	46	50	52.1%	0.683	45%	No
'Cherry 308'	52	38	42.2%	0.140	88%	No
'RN16'	46	47	50.5%	0.917	98%	No
'Cherry 5'	44	37	45.7%	0.437	83%	No
'Otto II'	153	142	48.1%	0.522	93%	No
'ACDC'	155	149	49.0%	0.731	97%	No
'Nebraska Feral'	53	54	50.5%	0.923	46%	Yes
'A2R4'	143	152	51.5%	0.600	68%	Yes
'RNF'	198	203	50.6%	0.803	85%	No

'RN13a'	56	38	40.4%	0.063	73%	No
Lithuanian	14	21	60.0%	0.237	70%	No
Total (N=2170)	1103	1067	49.2%	0.440		

Approximately 98% of the screened genetic females were phenotypically female. Approximately 1% of the screened genetic females were monoecious, including individuals from three cultivars (Table 2.1). Two screened plants were phenotypically male, and when retested, shown to be originally miscalled. 270 plants genetically scored as male from four hemp cultivars were allowed to flower in greenhouse conditions, and all were phenotypically male (Table 2.2). Monoecious plants (20 plants each of the cultivars 'Anka', 'Hlesia', and 'USO-31') were also examined with this assay, and all monoecious plants were scored as female.

Table 2.2. Validation of the *C. sativa* sex assay CSP-1. Lithuanian is a grain type. Nebraska is a grain/fiber type with monoecious individuals. 'RN16' and 'RNF' are CBD types.

	Lithuanian	'Nebraska'	'RN16'	'RNF'
Male Plants	14	53	46	157
Female/Monoecious Plants	21	54	47	157
Marker Accuracy	100%	100%	100%	100%

A second sex determination assay was developed as it was noted that several female individuals had null results, making it not truly codominant and not well suited for distinguishing YY supermales. To address this, CSP-2 was developed through identifying genes expressed only in males from a male-derived transcriptome (Braich et al 2019), aligning them to a male contig-level genome (Jamaican Lion Father, unpublished), and aligning this contig (JAATIQ010000695.1) to the X chromosome (Grassa et al., 2021).

SNPs were identified in Geneious Prime and validated to give identical results as CSP-1 on 500 plants, but to be strictly codominant and lacking nulls.

Sex assay application

The development of CSP-2 allowed the distinction between XY and putative YY plants through assaying the relative dosage of the X and Y chromosomes. It is unknown whether YY plants would be viable (perhaps due to essential genes located on the X chromosome lacking on the Y chromosome), or if the development of haploid ovules with a Y chromosome would be possible. Seeds and seedlings from XY plant-produced seed were grown and genotyped with CSP-2. Two distinct groups were found in XY plant-produced seed, phenotypically corresponding to male and female/monoecious plants. The genotype calls from each population are summarized in Table 2.4.

Table 2.3. Number of plants with CSP-2 genotype group calls.

	XY Mother (Seeds)	XY Mother (Seedlings)	XY Mother (Combined)	XX Mother
Female Group	17	16	33	17
Male Group	22	11	33	22

Cannabinoid chemotype assay development

A PACE assay to predict cannabinoid chemotype was generated through comparison of marijuana-type CBDAS (B_T) and hemp-type CBDAS (B_D), which were previously found to correspond to high-THC and high-CBD chemotypes, respectively (Figure 2.3; Weiblen et al., 2015). While THCAS and CBDAS are not the same gene, their close linkage in

repulsion suggests that they are inherited monogenically as a cannabinoid chemotype locus (de Meijer et al., 2003; Laverty et al., 2019). This assay was named CCP-1.

```
THCAS(Purple Kush)   ATTGCAGCATGGAAAATCAAACTGGTTGCTGTCCCATCAAAGTCTACTATATTCAGTGTT
CBDAS(Finola)        ATTGTAGCATGGAAAATTAGACTGGTTGCTGTCCCAAA---GTCTACTATGTTTAGTGTT
CBDAS(Purple Kush)   ATTGTAGCATGGAAAATTAGACTGGTTGCTGTCCCAACAAAGTCTACTATGTTTAGTGTT

THCAS(Purple Kush)   AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
CBDAS(Finola)        AAAAAGATCATGGAGATACATGAGCTTGTCAAGTTAGTTAACAAATGGCAAAATATTGCT
CBDAS(Purple Kush)   AAAAAGATCATGGAGATACATGAGCTTGTCAAGTGAGTTAACAAATGGCAAAATATTGCT
```

Figure 2.3. Alignment of CCP-1 primers to cannabinoid synthase genes. The blue sequence is specific to B_D alleles while the red sequence is specific to B_T alleles. The sequence in purple is common to both The 'Purple Kush' THCAS is found on scaffold19603, the 'FINOLA' CBDAS is found on scaffold 14546436, and the 'Purple Kush' CBDAS is found on scaffold 39155 (Van Bakel et al., 2011).

Cannabinoid chemotype assay validation

217 plants from 14 hemp cultivars grown for CBD in two locations were tested with the cannabinoid chemotype (CCP-1) assay and phenotyped for cannabinoids using HPLC. Of these, two were homozygous for the marijuana-type allele (B_T/B_T) 65 were heterozygous (B_T/B_D), and 150 were homozygous for the hemp-type allele (B_D/B_D). Most cultivar populations were segregating for this allele, which was consistent with the phenotypic data (Figure 2.4). The genotypic data corresponded with three apparent chemotypes, in terms of total potential CBD and THC (Figure 2.5a, ANOVA p<1e-4). This indicates that the CCP-1 assay identifies previously established B_T and B_D alleles (de Meijer et al., 2003).

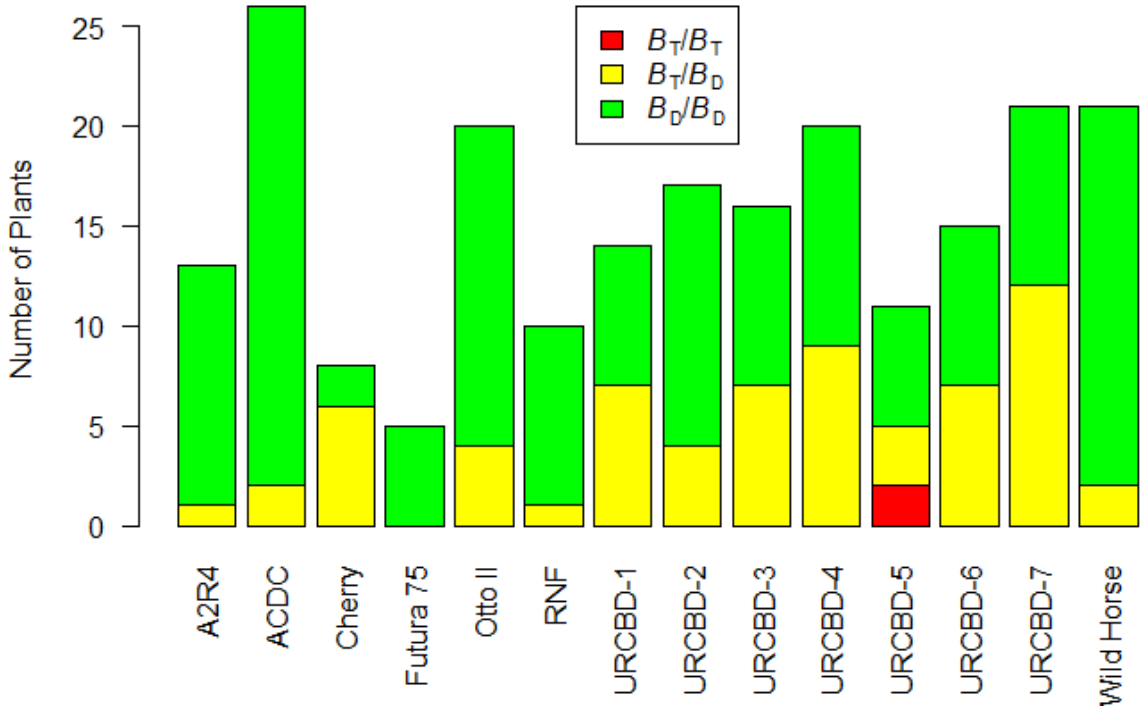

Figure 2.4. Distribution of cannabinoid chemotype alleles across cultivar populations. Additional allele frequency data can be found in Table 2.5.

Mean Δ^9-THC and total potential THC differed across genotypic groups (ANOVA p<1e-4). Within the genotypic groups there was a strong correlation between total potential CBD and total potential THC concentrations (Figure 2.5a; B_T/B_D r=0.72 p<1e-4, B_D/B_D r=0.86 p<1e-4).

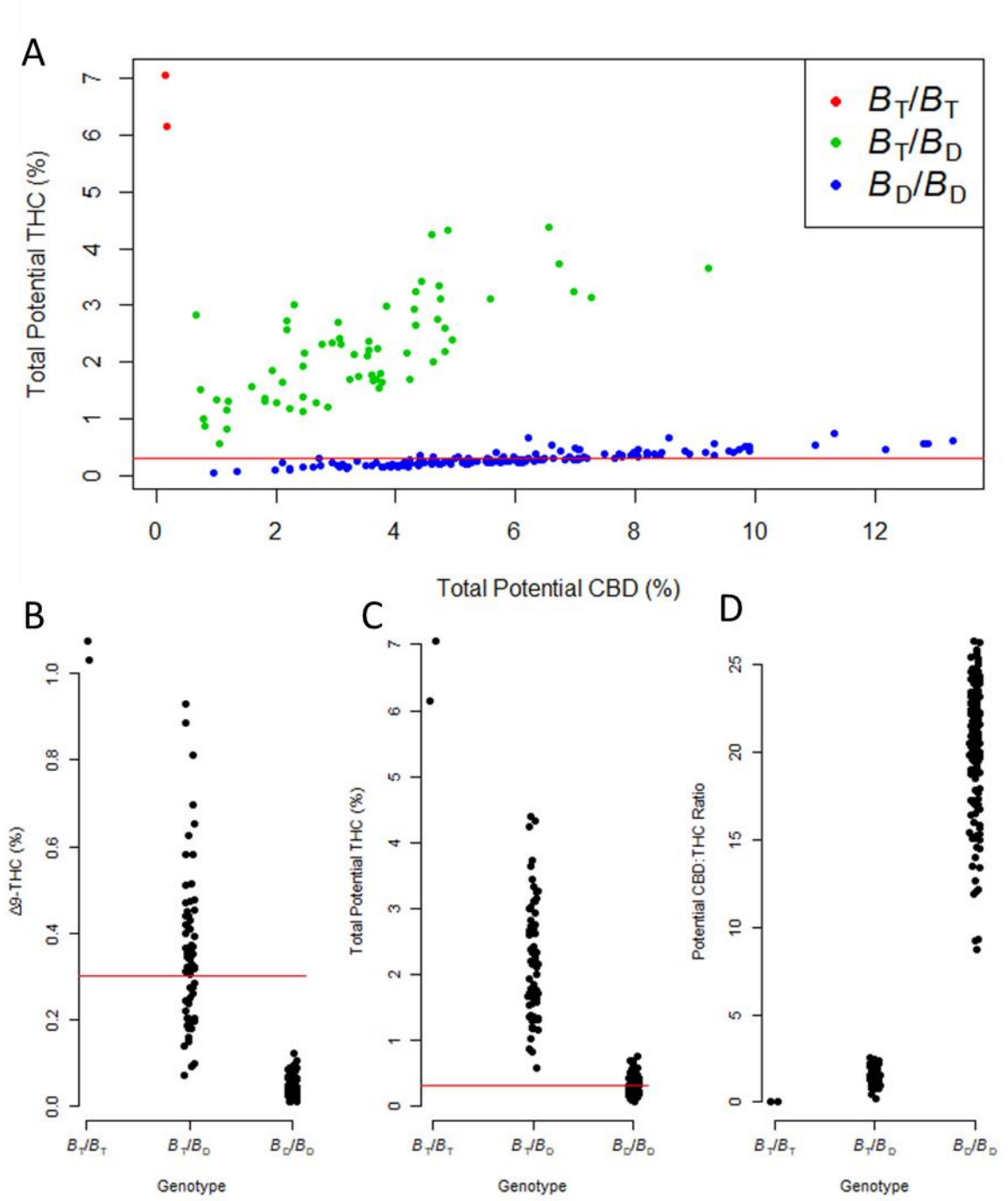

Figure 2.5. Genotype to phenotype relationships. A) Total potential THC and CBD concentration (% dry mass) by genotype determined by CCP-1. The red line indicates 0.3% total potential THC B) Δ^9-THC concentration by genotype. The red line indicates 0.3% dry weight Δ^9-THC. C) Total potential THC concentration by genotype. The red line indicates 0.3% dry weight total potential THC. D) Total potential CBD:THC concentration ratio. All means differ (ANOVA p<1e-4). Tabular data can be found in Table 2.4.

The Δ^9-THC concentration for B_D/B_D samples was consistently <0.3% (dry weight), while 35% of the B_T/B_D samples had a Δ^9-THC concentration <0.3% (Figure 2.5b). Only 39% of the B_D/B_D samples had total potential THC concentration <0.3% (Figure 2.5b). The mean ratio of total potential CBD:THC was 0.02, 1.6, and 20.3 for B_T/B_T, B_T/B_D, and B_D/B_D lines, respectively (Figure 2.5d, Table 2.4).

Table 2.4. Genotype class data.

	Samples	Chemotype	Min	Max	Mean	SD
Potential CBD:THC Ratio	2	B_T / B_T	0.02	0.03	0.02	0.00
	65	B_T / B_D	0.23	2.53	1.55	0.53
	150	B_D / B_D	8.75	26.34	20.28	3.67
THC (%)	2	B_T / B_T	1.03	1.08	1.05	0.03
	65	B_T / B_T	0.07	0.93	0.36	0.18
	150	B_D / B_D	0.01	0.12	0.04	0.02
Potential THC (%)	2	B_T / B_T	6.15	7.06	6.61	0.64
	65	B_T / B_D	0.57	4.38	2.20	0.88
	150	B_D / B_D	0.06	0.75	0.30	0.12
CBD (%)	2	B_T / B_T	0.07	0.08	0.08	0.01
	65	B_T / B_D	0.06	0.64	0.26	0.13
	150	B_D / B_D	0.07	1.13	0.26	0.15
Potential CBD (%)	2	B_T / B_T	0.14	0.17	0.16	0.02
	65	B_T / B_D	0.65	9.23	3.39	1.72
	150	B_D / B_D	0.96	13.30	5.92	2.34

A total of 2156 plants from 51 cultivars were tested with the CCP-1 assay. These cultivars were from multiple sources and grown for CBD, grain, or grain/fiber. The THC-associated B_T allele frequency varied by cultivar, from 0% in some clones grown for CBD, up to 98% in a Chinese grain/fiber cultivar (Table 2.5).

Cultivar	Source	Type	Usage	Alleles Tested	B_D %	B_T %	Plants Tested	B_D / B_D	B_T / B_D	B_T / B_T
Hlesia	Fiacre Seeds	Seed	Grain/Fiber	40	100%	0%	20	100%	0%	0%
USO-31	UNISeeds	Seed	Grain	40	100%	0%	20	100%	0%	0%
Cherry 307	Hemplogic	Seed	CBD	16	100%	0%	8	100%	0%	0%
Cherry 5	Hemplogic	Seed	CBD	162	100%	0%	81	100%	0%	0%
Deschutes	Industrial Seed Innovations	Feminized Seed	CBD	20	100%	0%	10	100%	0%	0%
First Light 49	Sunrise Genetics	Clone	CBD	2	100%	0%	1	100%	0%	0%
First Light 58	Sunrise Genetics	Clone	CBD	2	100%	0%	1	100%	0%	0%
First Light 70	Sunrise Genetics	Clone	CBD	2	100%	0%	1	100%	0%	0%
First Light 71	Sunrise Genetics	Clone	CBD	2	100%	0%	1	100%	0%	0%
First Light 80	Sunrise Genetics	Clone	CBD	2	100%	0%	1	100%	0%	0%
KG9201	Kayagene	Feminized Seed	CBD	14	100%	0%	7	100%	0%	0%
KG9202	Kayagene	Feminized Seed	CBD	14	100%	0%	7	100%	0%	0%
Nebraska Feral	WinterFox Farms	Seed	Grain/Fiber/CBD	184	100%	0%	92	100%	0%	0%
NY Cherry	Genesis Hemp Alliance	Seed	CBD	138	100%	0%	69	100%	0%	0%

Rogue	Industrial Seed Innovations	Feminized Seed	CBD	20	100%	0%	10	100%	0%	0%
T2	Boring Hemp	Feminized Seed	CBD	20	100%	0%	10	100%	0%	0%
Tangerine	NY Hemp Source	Clone	CBD	2	100%	0%	1	100%	0%	0%
TJs CBD	Stem Holding	Clone	CBD	4	100%	0%	2	100%	0%	0%
Umpqua	Industrial Seed Innovations	Feminized Seed	CBD	20	100%	0%	10	100%	0%	0%
Late Sue	NY Hemp Source	Clone	CBD	2	100%	0%	1	100%	0%	0%
Futura 75	AssoCanapa	Seed	CBD	10	100%	0%	5	100%	0%	0%
Fedora 17	Calderone	Seed	CBD	20	100%	0%	10	100%	0%	0%
RNF	Eric Cerecedes	Seed	CBD	202	100%	0%	101	99%	1%	0%
RN13a	Go Farm Hemp	Seed	CBD	174	99%	1%	87	98%	2%	0%
Cherry 308	Hemplogic	Seed	CBD	64	98%	2%	32	97%	3%	0%
RN16	Go Farm Hemp	Seed	CBD	190	98%	2%	95	97%	3%	0%
RN19	Go Farm Hemp	Seed	CBD	94	98%	2%	47	96%	4%	0%
A2R4	WinterFox Farms	Seed	CBD	120	96%	4%	60	92%	8%	0%
ACDC	WinterFox Farms	Seed	CBD	150	95%	5%	75	91%	9%	0%
Anka	UNISeeds	Seed	Grain/Fiber	40	95%	5%	20	90%	10%	0%

Winterlake: R4 x Cherry Wine	Calderone	Seed	CBD	20	95%	5%	10	90%	10%	0%
Wild Horse	WinterFox Farms	Seed	CBD	42	93%	7%	21	86%	14%	0%
URCBD-2	Ultra Rich CBD	Seed	CBD	34	91%	9%	17	82%	18%	0%
Picolo	Hemp Genetics International	Seed	Grain/Fiber	32	91%	9%	16	81%	19%	0%
Otto II	WinterFox Farms	Seed	CBD	218	89%	11%	109	77%	23%	0%
RN17	Go Farm Hemp	Seed	CBD	78	81%	19%	39	72%	18%	10%
URCBD-4	Ultra Rich CBD	Seed	CBD	38	79%	21%	19	58%	42%	0%
URCBD-3	Ultra Rich CBD	Seed	CBD	32	78%	22%	16	56%	44%	0%
URCBD-1	Ultra Rich CBD	Seed	CBD	26	77%	23%	13	54%	46%	0%
URCBD-6	Ultra Rich CBD	Seed	CBD	30	77%	23%	15	53%	47%	0%
URCBD-7	Ultra Rich CBD	Seed	CBD	42	71%	29%	21	43%	57%	0%
URCBD-5	Ultra Rich CBD	Seed	CBD	22	68%	32%	11	55%	27%	18%
Brilliance	Green Lynx Farms	Feminized Seed	CBD	74	68%	32%	37	32%	65%	3%
Cherry	JDMI Limited	Seed	CBD	16	63%	38%	8	25%	75%	0%

Puma	CN Kenaf & Hemp Seed Farm	Seed	Grain/Fiber	210	60%	40%	105	36%	49%	15%
Bama	CN Kenaf & Hemp Seed Farm	Seed	Grain/Fiber	156	57%	43%	78	33%	47%	19%
Han-NW	CN Kenaf & Hemp Seed Farm		Grain/Fiber	1074	49%	51%	537	22%	58%	20%
Sterling Gold	WinterFox Farms	Seed	Grain/Fiber	180	21%	79%	90	14%	12%	73%
R4	Calderone	Seed	-	10	20%	80%	5	20%	0%	80%
SC-1	PreProcess	Seed	Grain/Fiber	168	7%	93%	84	1%	11%	88%
Si-1	CN Kenaf & Hemp Seed Farm	Seed	Grain/Fiber	40	3%	98%	20	0%	5%	95%

Cannabinoid chemotype assay application

It was noted that the cultivar population ‘Han-NW’ (CN Kenaf and Hemp) had a large seed size and good yield potential (2019 Cornell Hemp Trials for New York State). Unfortunately, it also produced total THC in excess of 0.3%, and flowered very late to the point of not maturing in New York State. However, segregation for chemotype was expected given previous results, and segregation for flowering time was also noted in the field. To use this population for the breeding of a cultivar, marker assisted selection for chemotype was accomplished using CCP-1 and selection for earliness was done phenotypically in the field. The effect of selection on cannabinoid profile, the yield, and maturity date of the resultant cultivar (GVA-H-20-1179) is shown in Figure 2.6.

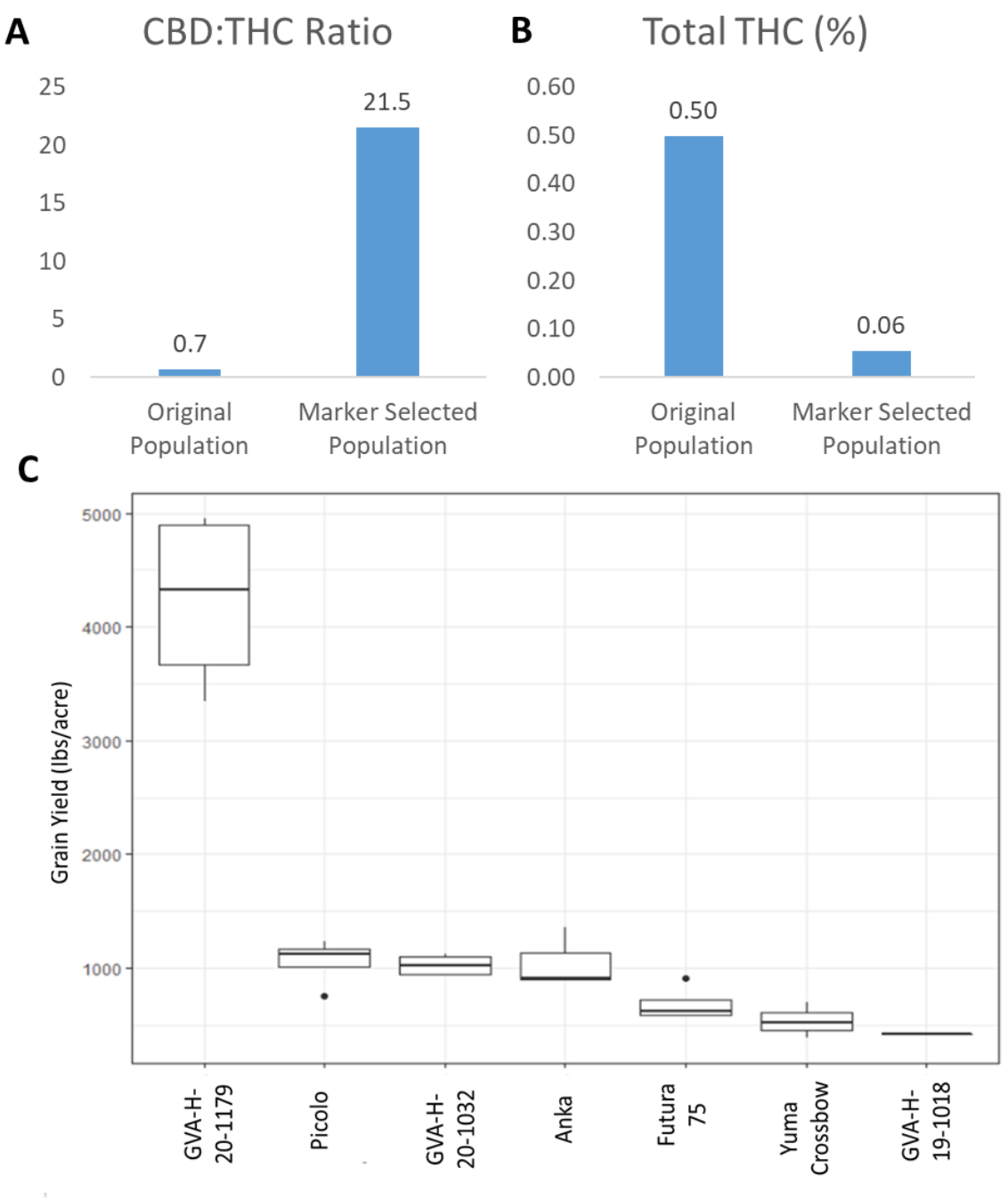

Figure 2.6. Application of CCP-1 to a breeding population. A) Effect of selection on CBD:THC ratio. B) Effect of selection on THC content. C) Yield of selected population (GVA-H-20-1179) and various checks.

Other factors affecting cannabinoid production

Genotypic group, cultivar, and trial were used to create models explaining the potential CBD:THC concentration ratio as well as the concentrations of Δ^9-THC, CBD, potential THC, potential CBD, and total potential cannabinoids (Table 2.6). Total potential cannabinoids included CBD, THC, CBC, CBG, and their corresponding acids. Genotypic group explained the most variance in the CBD:THC ratio, as well as Δ^9-THC and potential THC levels, but not total potential cannabinoids. Cultivar was an important factor in total potential cannabinoid abundance, as well as the concentration of CBD and Δ^9-THC. The cultivar explained ~3% of the variation in the potential CBD:THC ratio when the genotypic group was taken into consideration, and the trial was a poor predictor of all measured variables.

Table 2.6. Linear regression R^2 values of various models predicting cannabinoid data. "+" indicates the variable was included in the model and "–" indicates that the variable was not included in the model. Light grey cells are $p<0.01$. Dark grey cells are $p<1e\text{-}4$.

	Trial	Marker Coding	Cultivar	Potential CBD:THC	Δ^{9}-THC (%)	CBD (%)	Potential THC (%)	Potential CBD (%)	Total Potential Cannabinoids (%)
Model 1	+	+	+	0.89	0.77	0.21	0.81	0.38	0.19
Model 2	+	+	–	0.86	0.74	0.03	0.78	0.25	0.01
Model 3	+	–	+	0.26	0.23	0.19	0.20	0.18	0.19
Model 4	+	–	–	0.01	0.01	0.03	0.00	0.01	0.01
Model 5	–	+	+	0.89	0.76	0.18	0.81	0.38	0.19
Model 6	–	+	–	0.86	0.73	0.00	0.77	0.25	0.01
Model 7	–	–	+	0.25	0.22	0.16	0.20	0.17	0.18

2.6 Discussion

CSP-1 sex assay

The CSP-1 assay was found to be a reliable predictor of plant sex. There was a 50:50 segregation ratio in nearly all tested dioecious populations including CBD types, grain types, and grain/fiber types. As expected, monoecious plants were scored as female (Divashuk et al., 2014). Given these data, it is likely that the CSP-1 assay distinguishes a non-recombining part of the Y chromosome. This is somewhat surprising, given that the original MADC6 marker assay was not diagnostic for plant sex, with 2/75 reported recombinants (Törjék et al., 2002). It is possible that this was due to PCR failure, monoecious plants with a quantitatively male phenotype, or that the CSP-1 assay in fact examines a different DNA sequence than the original MADC6 assay. Recent *C. sativa* whole-genome sequencing (Laverty et al., 2019) showed six unassembled scaffolds in the male genome with >99% identity to the MADC6 sequence in *C. sativa*, possibly contributing to the empirical success of this assay. As MADC6 shows some sequence relationship to retrotransposons, it is possible that the sequence was subject to copy number increase in the recent past (Sakamoto et al,. 2000; Törjék et al., 2002). It is well known that in the development of sex-determining regions of plants, an absence of recombination between male- and female-specific sequences can lead to an expansion of retrotransposon copy number repeats, which are not lost through a Muller's ratchet-type mechanism (Sakamoto et al., 2000; Vyskot & Hobza, 2004). The assay CSP-2 was found to give identical results to CSP-1 but be easier to score and lacked null genotype calls.

Application of CSP-2

CSP-2 resulted in only two distinct groups when tested on XY plant-produced seeds, corresponding to female and male plants. Interestingly, the data better suggests a 1:1 segregation ratio of male:female plants in the XY plant-produced seeds and seedlings than a 2:1 segregation ratio which might have been expected if YY plants alone were non-viable. This could potentially be due to selection against Y-only haploid female germ cells, with most or all Y chromosomes from the XY mother plant coming from the pollen parent. This could be further assayed through tagged Y chromosomes, or through the use of pollen from XX plants.

CCP-1 cannabinoid chemotype assay

The CCP-1 cannabinoid chemotype marker assay detected three genotypic groups that corresponded to three phenotypic groups, reflecting previously described chemotypes (de Meijer et al., 2003). Since the CCP-1 assay examines CBDAS only and CBDAS and THCAS are not allelic, a recombinant plant that would be scored incorrectly is possible. However, I did not detected this in any of my samples, and the tight linkage in repulsion between CBDAS and THCAS is well established (de Meijer et al., 2003; Grassa et al., 2021; Weiblen et al., 2015).

The mean Δ^9-THC and total potential THC concentrations as percent dry matter were significantly different in each chemotype. If Δ^9-THC concentration alone as assayed by HPLC was used as the criterion for legal compliance at the level of 0.3%, then all B_D/B_D samples and 35% of the B_T/B_D samples would be below the threshold. It is possible that past breeding material chosen for low THC was in fact heterozygous, leading to

segregation in released cultivars (Table 2.6). If total potential THC is used as the legal criterion, then 61% of the B_D/B_D samples would be above the legal threshold of 0.3%, and therefore be non-compliant. The close correlation between potential CBD and THC concentrations in the B_D/B_D class suggests that it might be difficult to develop a cultivar that accumulates high CBD concentrations, while maintaining low total potential THC. The average potential CBD:THC ratio was about 20:1, which suggests that accumulation of greater than 6% CBD will result in total potential THC rising above 0.3%. A clear target for breeders developing high CBD hemp cultivars is to raise the ratio of total potential CBD:THC.

There are several lines of evidence to suggest that the concomitant increase in THC concentration with that of CBD in the B_D/B_D group is due to promiscuous activity of the active CBDAS. Despite attempts, no demonstrably active transcribed THCAS has been isolated from a confirmed chemotype III plant (Kojoma et al., 2006; Laverty et al., 2019; Onofri et al., 2015). Other research found a *C. sativa* plant with a B_D/B_D genotype and a catalytically inactive CBDAS that accumulates CBGA and essentially no THCA, although mutations in CBDAS and a putative THCAS would also explain this observation (Onofri et al., 2015). A purported active THCAS from CBD-dominant fiber-type hemp was later shown to be a cannabichromenic acid synthase (CBCAS) (Kojoma et al., 2006; Laverty et al., 2019). Lastly, *in vitro* expression of wild-type CBDAS leads to production of CBDA:THCA in a ratio very close to 20:1 at optimal pH (Zirpel, Kayser, & Stehle, 2018). Future breeding efforts should be informed by this promiscuous activity.

Some studies have found sequence and copy number variation in CBDAS and THCAS and correlated them to differences in cannabinoid production (Weiblen et al., 2015). These

differences were not assayed here, but could have conceivably contributed to some of the variation within groups.

Application of CCP-1

As shown in Figure 2.6, CCP-1 was effective in in selecting chemotype III individuals from the 'Han-NW' population, leading to greatly increased CBD:THC ratio and greatly decreased total THC content, below the 0.3% threshold. Phenotypic selection on flowering time also resulted in a cultivar population harvestable in Upstate New York, although with a maturity date of October 15 it runs the risk of being affected by early frost (NRCC First Frost). With first fall frost date on average getting later in the region (Dobson et al., 2020), it is possible this cultivar will become better suited to the region as it changes, or may be more suited to more southern latitudes or locations with longer growing seasons. While harvest of this cultivar may be difficult, it displays excellent yield potential, almost 9 times the current average hemp seed yield in the USA (4300 lbs/acre vs 530 lbs/acre, USDA, 2022).

Other factors affecting cannabinoid production

Neither trial nor cultivar *per se* explained variations in the CBD:THC ratio. Trial did not appear to have much of an effect on any measured parameters, despite flooding stress in one trial location. The explanatory power of models 3 and 7, including cultivar but not marker coding, are likely due to differences in allele frequency in each cultivar population. While cultivar *per se* poorly explained CBD:THC ratio, cultivar was the best predictor of total potential cannabinoid concentration. It has previously been demonstrated that the

factors affecting cannabinoid chemotype are not linked to total cannabinoid content (Grassa et al., 2021).

In testing 14 different cultivars, with 217 plants total across two locations, I found that none of the B_D/B_D samples had a Δ^9-THC concentration >0.3% dry weight. However, I did find that most cultivar populations were segregating for the B_T allele. It is possible that differences in cannabinoid production ascribed to changes in environment may in fact be due to sampling of individual plants with B_T alleles. Additional studies of the influence of environment on cannabinoid production coupled with individual plant genotyping may lead to a better understanding of the regulation of cannabinoid production.

2.7 Acknowledgements

Other individuals who made substantial intellectual contributions to this work include George M. Stack, Ali Cala, Rebecca Wilk, Jamie Crawford, Craig H. Carlson, Glenn Philippe, Donald R. Viands, Christine D. Smart, Jocelyn K.C. Rose, and Lawrence B. Smart. I am grateful to the Smart, Viands, and Rose research teams for their support, especially Lauren Carlson, Jane Petzoldt, Dawn Fishback, Ben DeMoras, and Sarah Wright. This work was supported by New York State Department of Agriculture and Markets through grants AC477 and AC483 from Empire State Development Corporation.

2.8 References

2019 Cornell Hemp Trials for New York State. Cornell Hemp. (2020). Retrieved March 15, 2022, from https://hemp.cals.cornell.edu/resources/reports-factsheets/

Ainsworth, C. (2000). Boys and girls come out to play: the molecular biology of dioecious plants. *Annals of Botany*, *86*(2), 211-221.

Borna, T., Salami, S. A., & Shokrpour, M. (2017). High resolution melting curve analysis revealed SNPs in major cannabinoid genes associated with drug and non-drug types of cannabis. *Biotechnology & Biotechnological Equipment*, *31*(4), 839-845.

Braich, S., Baillie, R. C., Jewell, L. S., Spangenberg, G. C., & Cogan, N. O. (2019). Generation of a comprehensive transcriptome atlas and transcriptome dynamics in medicinal cannabis. *Scientific reports*, *9*(1), 1-12.

Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype× Environment Interactions of Industrial Hemp Cultivars Highlight Diverse Responses to Environmental Factors. *Agrosystems, Geosciences & Environment*, *2*(1).

Citti, C., Braghiroli, D., Vandelli, M. A., & Cannazza, G. (2018). Pharmaceutical and biomedical analysis of cannabinoids: a critical review. *Journal of Pharmaceutical and Biomedical Analysis*, *147*, 565-579.

de Meijer, E., Hammond, K., & Micheler, M. (2009). The inheritance of chemical phenotype *in Cannabis sativa* L.(III): variation in cannabichromene proportion. *Euphytica*, *165*(2), 293-311.

de Meijer, E., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. C., Ranalli, P., & Mandolino, G. J. G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*, *163*(1), 335-346.

DEA Drug Scheduling. URL https://www.dea.gov/drug-scheduling

Divashuk, M. G., Alexandrov, O. S., Razumova, O. V., Kirov, I. V., & Karlov, G. I. (2014). Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. *PLoS One*, *9*(1), e85118.

Dobson, K. C., Beaty, L. E., Rutter, M. A., Hed, B., & Campbell, M. A. (2020). The influence of Lake Erie on changes in temperature and frost dates. *International Journal of Climatology*, *40*(13), 5590-5598.

Doyle, J. J., & Doyle, J. L. (1987). *A rapid DNA isolation procedure for small quantities of fresh leaf tissue*. *Phytochemical Bulletin*, *19*(1):11:15.

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.

NRCC First Frost. (2000). Retrieved March 15, 2022, from https://www.nrcc.cornell.edu/services/blog/2009/10/06_first_frost/index.html

Fellermeier, M., Eisenreich, W., Bacher, A., & Zenk, M. H. (2001). Biosynthesis of cannabinoids: Incorporation experiments with 13C-labeled glucoses. *European Journal of Biochemistry*, *268*(6), 1596-1604.

Fellermeier, M., & Zenk, M. H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS letters*, *427*(2), 283-285.

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

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

Lubell, J. D., & Brand, M. H. (2018). Foliar sprays of silver thiosulfate produce male flowers on female hemp plants. *HortTechnology*, *28*(6), 743-747.

Mahlberg, P. G., & Kim, E. S. (2004). Accumulation of cannabinoids in glandular trichomes of *Cannabis* (Cannabaceae). *Journal of Industrial Hemp*, *9*(1), 15-36.

Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., & de Meijer, E. (2003). Qualitative and quantitative aspects of the inheritance of chemical phenotype in *Cannabis*. *Journal of Industrial Hemp*, *8*(2), 51-72.

Mandolino, G., Carboni, A., Forapani, S., Faeti, V., & Ranalli, P. (1999). Identification of DNA markers linked to the male sex in dioecious hemp (*Cannabis sativa* L.). *Theoretical and Applied Genetics*, *98*(1), 86-92.

Menzel, M. Y. (1964). Meiotic chromosomes of monoecious Kentucky hemp (*Cannabis sativa*). *Bulletin of the Torrey Botanical Club*, 193-205.

Ming, R., Bendahmane, A., & Renner, S. S. (2011). Sex chromosomes in land plants. *Annual Review of Plant Biology*, *62*, 485-514.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, *28*(12), e63-e63.

Onofri, C., de Meijer, E. P., & Mandolino, G. (2015). Sequence heterogeneity of cannabidiolic-and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry*, *116*, 57-68.

Pacifico, D., Miselli, F., Carboni, A., Moschella, A., & Mandolino, G. (2008). Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica*, *160*(2), 231-240.

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., & Mandolino, G. (2006). Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. *Molecular Breeding*, *17*(3), 257-268.

Peil, A., Flachowsky, H., Schumann, E., & Weber, W. (2003). Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). *Theoretical and Applied Genetics*, *107*(1), 102-109.

Ram, H. M., & Jaiswal, V. S. (1970). Induction of female flowers on male plants of *Cannabis sativa* L. by 2-chloroethanephosphonic acid. *Experientia*, *26*(2), 214-216

Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., Guo, Q., Zhang, Y., Dreisigacker, S., Xia, X. and He, Z. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theoretical and Applied Genetics*, *129*(10), 1843-1860.

Razumova, O. V., Alexandrov, O. S., Divashuk, M. G., Sukhorada, T. I., & Karlov, G. I. (2016). Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution.

Protoplasma, 253(3), 895-901.
Rotherham, D., & Harbison, S. (2011). Differentiation of drug and non-drug *Cannabis* using a single nucleotide polymorphism (SNP) assay. *Forensic Science International, 207*(1-3), 193-197.
Sakamoto, K., Abe, T., Matsuyama, T., Yoshida, S., Ohmido, N., Fukui, K., & Satoh, S. (2005). RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. *Genome, 48*(5), 931-936.
Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H., & Satoh, S. (1998). Characterization; genome sizes and morphology of sex chromosomes in hemp (*Cannabis sativa* L.). *Cytologia, 63*(4), 459-464.
Sakamoto, K., Ohmido, N., Fukui, K., Kamada, H., & Satoh, S. (2000). Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Molecular Biology, 44*(6), 723-732.
Sakamoto, K., Shimomura, K., Komeda, Y., Kamada, H., & Satoh, S. (1995). A male-associated DNA sequence in a dioecious plant, *Cannabis sativa* L. *Plant and Cell Physiology, 36*(8), 1549-1554.
Schaffner, J. H. (1921). Influence of environment on sexual expression in hemp. *Botanical Gazette, 71*(3), 197-219.
Shao, H., Song, S.-J., & Clarke, R. C. (2003). Female-associated DNA polymorphisms of hemp (Cannabis sativa L.). *Journal of Industrial Hemp, 8*(1), 5-9.
Small, E., & Beckstead, H. (1973). Common cannabinoid phenotypes in 350 stocks of *Cannabis*. *Lloydia*. *36*(2), 144-65.
Smith, R., & Vaughan, C. (1977). The decomposition of acidic and neutral cannabinoids in organic solvents. *Journal of Pharmacy and Pharmacology, 29*(1), 286-290.
R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna,Austria. URL https://www.R-project.org/
Törjék, O., Bucherna, N., Kiss, E., Homoki, H., Finta-Korpelová, Z., Bócsa, I., Nagy, I., & Heszky, L. E. (2002). Novel male-specific molecular markers (MADC5, MADC6) in hemp. *Euphytica, 127*(2), 209-218.
Toth, J., Pandurangan, S., Burt, A., Mitchell Fetch, J., & Kumar, S. (2018). Marker-assisted breeding of hexaploid spring wheat in the Canadian prairies. *Canadian Journal of Plant Science, 99*(2), 111-127.
Turner, B. D., Sloan, S. W., & Currell, G. R. (2019). Novel remediation of per-and polyfluoroalkyl substances (PFASs) from contaminated groundwater using *Cannabis sativa* L. (hemp) protein powder. *Chemosphere, 229*, 22-31.
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.

USDA. (2022). Value of hemp production totaled $824 million in 2021. Retrieved March 15, 2022, from https://www.nass.usda.gov/Newsroom/2022/02-17-2021.php#:~:text=The%20average%20yield%20for%20hemp,production%20totaled%2027.6%20million%20pounds.
Vogl, C. R., Lissek-Wxsolf, G., & Surböck, A. (2004). Comparing hemp seed yields (*Cannabis sativa* L.) of an on-farm scientific field experiment to an on-farm agronomic evaluation under organic growing conditions in lower Austria. *Journal*

of Industrial Hemp, *9*(1), 37-49.
Vyskot, B., & Hobza, R. (2004). Gender in plants: sex chromosomes are emerging from the fog. *TRENDS in Genetics, 20*(9), 432-438.
Wang, Q., & Xiong, Y. L. (2019). Processing, Nutrition, and Functionality of Hempseed Protein: A Review. *Comprehensive Reviews in Food Science and Food Safety, 18*(4), 936-952.
Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., & Marks, M. D. J. N. p. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytologist*, *208*(4), 1241-1250.
Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *Journal of Biotechnology, 284*, 17-26.

Chapter 3: Limited effect of environmental stress on cannabinoid profiles in high-cannabidiol hemp (*Cannabis sativa* L.)[2]

3.1 Chapter overview

As discussed in Chapter 2, there is a strong genetic component to THC production in the *B* locus, which controls which major cannabinoids are synthesized. A major result from that work was that the ratio of CBD:THC was apparently fixed across chemotype III genotypes, with little or no environmental variance. This is contrary to public understanding of how THC is produced, with a New York Times article published in 2019 discussing how environmental stress is a major cause of THC production in hemp (Nir, 2019). The work described in Chapter 2 gives a convincing account of what might cause THC levels to "spike" (be very high, especially compared to CBD levels), in the demonstrated segregation of B_T alleles conferring low CBD:THC ratios regardless of environment.

Despite the strong genetic influence on CBD:THC ratios, it is conceivable that this ratio could change upon environmental stress within a given cannabinoid chemotype. If the mechanistic source of THC is off-target activity from the CBDAS (Zirpel et al., 2018, Toth

[2] Chapter 3 was published as Toth, J. A., Smart, L. B., Smart, C. D., Stack, G. M., Carlson, C. H., Philippe, G., & Rose, J. K. (2021). Limited effect of environmental stress on cannabinoid profiles in high-cannabidiol hemp (*Cannabis sativa* L.). *GCB Bioenergy, 13*(10), 1666-1674.

et al., 2020), then by altering the environment of the CBDAS, perhaps by trichome acidification or expression of dirigent proteins, it is possible that the ratio may change. To date the published research on effect of environment in chemotype III plants that produce high levels of cannabinoids is very limited, and the work described below fills an important niche and emphasizes the somewhat surprising result of Chapter 2, that CBD:THC ratios are essentially genetically fixed in chemotype III plants.

3.2 Abstract

Hemp (*Cannabis sativa*) is a burgeoning crop, but research-based information about genetic and environmental effects of cannabinoid production is limited and will be essential for expanded cultivation. There are limited data available about the effect of environmental stressors on cannabinoid content, particularly for tetrahydrocannabinol (THC) in high-cannabidiol (CBD) hemp. To address this, five stress treatments were applied in a replicated field trial with three high-CBD hemp cultivars and cannabinoid content was assayed over a three-week time-course spanning floral maturation. Cannabinoid production in terminal inflorescence shoot tip samples of three cultivars was measured under stress imposed by flooding, ethephon, powdery mildew, herbicide, and physical wounding in a split plot design. The treatments had limited effects on cannabinoid levels, with the exception of herbicide treatment which resulted in decreased cannabinoid content. Notably, there was no evidence that any of these stresses caused THC concentration or the ratio of THC to CBD to increase at harvest.

3.3 Introduction

Hemp (*Cannabis sativa* L.) has many potential uses, including grain, fiber, and cannabinoid production. Fundamental knowledge of the genetic and environmental influences on important traits is critical for the breeding of improved, stable, and uniform cultivars that are compliant with regulations of tetrahydrocannabinol (THC) concentration. In many countries, there is a regulatory threshold of THC concentration in dry floral tissue that defines *C. sativa* as hemp. This threshold varies between countries, with a value of 0.2% in most of Europe (Salentijn et al., 2015), 0.3% in the United States (Adesso et al., 2019), and 1% in Australia (Davidson et al., 2018). It has been suggested that various environmental stresses increase the abundance of cannabinoids in hemp, especially THC (Nir, 2019), however there are limited published data to address this idea.

Previous work has determined that the suite of major cannabinoids produced (THC, CBD, and cannabigerol, CBG), also referred to as the cannabinoid chemotype, is a simple genetic trait, but that variation in cannabinoid content is genetically complex and potentially affected by environment (Campbell et al., 2019; de Meijer et al., 2003; Mandolino et al., 2003). Cannabinoid chemotype can be predicted by the allelic state of the *B* locus, with production of mostly THC (chemotype I) characteristic of homozygous B_T individuals, production of about equal THC and CBD (chemotype II) typical of heterozygous B_T/B_D individuals, and production of mostly CBD (chemotype III) typifying homozygous B_D individuals (de Meijer et al., 2003). Many commercially available cultivar populations, including high-cannabinoid hemp as well as grain and fiber types, are segregating at the *B* locus (Chapter 2, Toth et al., 2020). Breeding for homozygous B_D individuals will be essential to stabilize hemp cultivars for THC compliance, but the

degree to which other factors, such as environmental stressors, affect cannabinoid production is not well established.

The major acidic cannabinoids, cannabidiolic acid (CBDA), tetrahydrocannabinolic acid (THCA), and cannabichromenic acid (CBCA), are synthesized from CBGA by CBDA, THCA, or CBCA synthases. These acidic cannabinoids decarboxylate under certain conditions (Perrotin-Brunel et al., 2011), forming CBG, CBD, THC, and cannabichromene (CBC). It is conceivable that stress alters conditions relevant to decarboxylation, such as production of antioxidants (Singh et al., 2020). CBDA and THCA synthases exhibit product promiscuity when heterologously expressed in yeast, meaning they make small amounts of the other cannabinoids in addition to their major products when incubated with CBGA precursor (Zirpel et al., 2018). Notably, CBDA synthase has been reported to synthesize approximately 5% THCA and 5% CBCA normalized to 100% CBDA (Zirpel et al., 2018). This is likely to be the primary source of the THCA and CBCA that has been detected in chemotype III hemp plants that do not express active THCA synthase (Toth et al., 2020, Stack et al., 2021). It is conceivable that allelic variation among CBDA synthases or expression of other cannabinoid synthases could lead to altered CBD:THC ratios. While there was good agreement between *in planta* data and *in vitro* data for this ratio in a previous report (Toth et al., 2020), further testing is required to determine if there is variation of this ratio within chemotype III plants and to determine the environmental effect, if any.

Most studies to date on the effect of stresses on cannabinoid production have focused on cannabinoid chemotype I and II plants grown under controlled environment conditions. For example, in a small study of greenhouse-grown chemotype II plants, drought stress

was associated with increased levels of cannabinoids (THC and CBD) (Caplan et al., 2019). Another study found an increase in THC upon UVB exposure (Lydon et al., 1987) in drug-type (chemotype I) plants, but no increase in any cannabinoids in fiber (chemotype III) plants. Other work linked abscisic acid (ABA) with changes in THC concentration (Mansouri & Asrar, 2012; Mansouri et al., 2009), although the direction of this effect was not consistent.

The effect of stress on field-grown high-cannabinoid chemotype III hemp plants is not well understood and of great potential importance for production systems. If stresses resulted in increased cannabinoid content or variation in CBD:THC ratio, management of stress (induction or avoidance) would play a critical role in production systems. This current study examined the effect of stresses on the accumulation of cannabinoids in three high-cannabinoid CBD cultivars using exclusively female chemotype III plants in a split-plot design in a single outdoor location. Here I examined five stresses, as well as an unstressed control. The five stresses were as follows:

1) Ethephon: Ethephon (2-chloroethyl phosphoric acid) is a plant growth regulator that in converted *in planta* to ethylene, a plant hormone involved in aspects of plant development. Previous work has found an effect of ethephon on cannabinoids (Mansouri et al., 2013; Mansouri et al., 2016), but its effect on field-grown high-CBD plants has yet to be investigated. Ethephon has also been used to induce genetically male plants to produce female flowers (Ram & Jaiswal, 1970). It is possible that ethephon treatment, by inducing female-associated gene expression, could lead to increased trichome numbers on female inflorescences and accordingly increased cannabinoid concentration.

2) Flooding: Flooding is an abiotic stress that can occur following high rainfall, especially in poorly drained soils. Flooding can lead to hypoxia in the roots, leading to reduced nutrient uptake and the production of stress hormones (Colmer & Voesenek, 2009). However, previous work found limited difference in cannabinoid content between a naturally flooded field and one without this stress (Chapter 2, Toth et al., 2020), or between an irrigated and non-irrigated field (Campbell et al., 2019).
3) Herbicide: As hemp acreage grows, it will be important to consider how hemp responds to commonly applied chemicals such as herbicides. While the general effect of herbicides on hemp has been not been rigorously studied, herbicide drift is a relatively common phenomenon that has been found to injure susceptible plants and interfere with secondary metabolism (Ding et al., 2011). For example, the herbicide glyphosate interferes with the shikimate pathway in plants (Duke & Powles, 2008). While the shikimate pathway is not directly involved in cannabinoid biosynthesis, glyphosate-induced stress might alter cannabinoid levels through general stress responses or result in reduced vigor.
4) Powdery Mildew: Powdery mildew, caused by the fungal pathogen *Golovinomyces spadiceus* is a biotic stress which is common in greenhouses and fields with favorable environmental conditions (Szarka et al., 2019; Weldon et al., 2020). Powdery mildew has the potential to reduce yield, especially in greenhouse conditions (Lyu et al., 2019), but can also be severe in outdoor field settings. The effect of powdery mildew on cannabinoid production is largely unknown, but cannabinoids may have evolved to deter pests and pathogens (Gorelick & Bernstein, 2017) and so such a relationship would not be surprising.

5) Wounding: Mechanical damage can be caused by natural sources, such as hail or herbivory, or result from cultivation and mechanical weed removal. It has been suggested that wounding that mimics insect damage might increase cannabinoid levels, and that the resistance of *Cannabis* to insects might be substantially affected by cannabinoids (Gorelick & Bernstein, 2017). In general, wounding has the potential to cause a systemic response, inducing the systemic production of hormones such as jasmonic acid and abscisic acid (Savatin et al., 2014), which have been linked to changes in cannabinoid abundance (Mansouri et al., 2009; Salari & Mansori, 2013).

3.4 Materials and methods

Three cultivars of high-cannabinoid hemp were used for this study: 'TJ's CBD' (Stem Holdings Agri, Eugene, OR; clonal), 'T2' (Boring Hemp, Boring, OR; feminized seeds), and Cornell breeding line GVA-H-19-1039 (dioecious seeds). All cultivars were started at a similar time from either cuttings or seeds, and GVA-H-19-1039 was screened using the molecular marker CSP-1 (Chapter 2, Toth et al., 2020) to remove male plants. All selected plants were entirely phenotypically female and no pollination in the field was noted. A split-plot design was used, with the three cultivars randomized within treatment plots, which were randomized in a complete block design with four replicate blocks (Figure 3.1). Each treatment plot contained three plants of each cultivar.

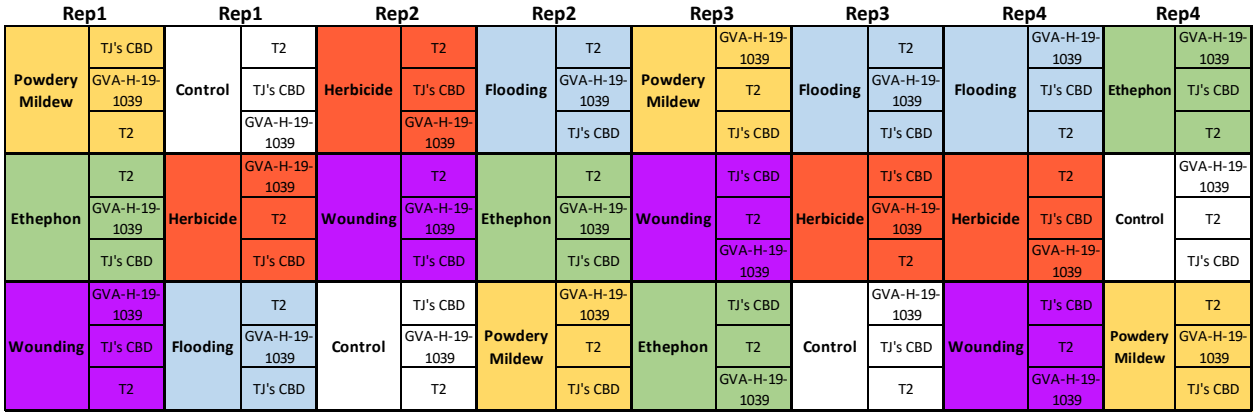

Figure 3.1. Experimental plot layout of hemp stress trial. Three plants per cultivar in each subplot were planted.

Seedlings and cuttings were established in a greenhouse in potting mix (Lambert's LM111) in 50-cell deep trays. These were transplanted into raised beds with black plastic mulch and drip irrigation on July 28, 2019 at the Cornell AgriTech McCarthy Farm, Geneva, NY (42.896300, -77.008062) in a field with well drained Ontario loam soil with more than 2 m depth to a restrictive feature. Conventional fertilizer (19-19-19 N-P-K, Phelps Supply Inc., Phelps, NY) was applied at a rate of 157 kg N ha^{-1} during bed formation. No additional fertilizer was added after transplanting. Soil moisture was monitored in the control and flooded plots using an Onset HOBO RXW-SMD-10HS sensor installed to a depth of 10 cm in the middle of each plot and wirelessly linked to a HOBO RX3000 remote monitoring station (Onset, Bourne, MA). Adequate soil moisture was applied through trickle irrigation during periods with insufficient rainfall to maintain soil volumetric water content >0.27 m^3 m^{-3}. Temperature and rainfall data for this site are reported in Stack et al. (2021).

Stress treatments were initially applied on Sept. 14 and 15, 2019 when the plants had initiated terminal flowering. For the flooding stress, irrigation was applied through trickle

irrigation only on flood treatment plots sufficient to raise soil volumetric water content to field capacity (0.35-0.4 m^3 m^{-3}) and was repeated throughout the sampling period in order to maintain soil volumetric water content >0.32 m^3 m^{-3}, typically two or three times per week. Ethephon (0.5% Ethephon 2, Nufarm, Alsip, IL, 1% active ingredient, 75 mM) was applied as a spray to the entire plant until leaves were fully wet. Ethephon was applied twice, once on September 14 and again on September 22, 2019. Powdery mildew inoculation was accomplished by transferring dry conidia from diseased leaves to shoot tips of treatment plants using a paint brush. Leaves infected with *G. spadiceus* were taken from naturally infected plants cultivar 'TJ's CBD' growing in a variety trial in Geneva, NY. Four shoots of each plant in the powdery mildew treatment plots were marked with flagging tape for subsequent shoot tip sampling, and the terminal five leaves of each shoot were painted with dry conidia. Glyphosate (0.5% Roundup Pro, Monsanto, St. Louis, MO) was applied one time as a spray to the entire plant until leaves were wet. The wounding treatment was accomplished by partially damaging the lower and middle foliage with a grass and weed trimmer (Model FS70R, Stihl Inc, Virginia Beach, VA) in such a way as to remove or wound a majority of the foliage on the outer portion of the plant below the inflorescence. The percentage of damage was not precisely quantified, but since the inner portions of each stem were not affected, the damage was approximately 40-50% of foliage wounded below the inflorescence. The damage was implemented to remove and damage the leaves, but not to break or prune stems. The wounding treatment was applied on September 14 and repeated immediately after the week two sampling on September 29, 2019.

Shoot tips were sampled for cannabinoid extraction and analysis immediately prior to application of the stress treatments and again in one-week intervals for three weeks (September 14, 22, 29, October 6 2019) for a total of four sampling times. The third week after initial stress application was designated as the presumptive harvest date. The plants that received the herbicide treatment began to exhibit necrosis and browning by the week two sampling period, so sampling targeted the healthiest looking shoot tips remaining by week three. One shoot tip sample was collected by harvesting the top 10 cm of an upper canopy shoot from each of the three plants in a plot, and those three shoot tips were combined in a paper bag, air dried at room temperature, and milled in a Nutri Ninja Pro food blender (SharkNinja Operating LLC, Needham, MA). Cannabinoids were extracted and quantified by high pressure liquid chromatography (HPLC) using a previously established method (Stack et al. 2021). Total cannabinoids were calculated by summing the neutral form with the acidic form multiplied by a factor (0.877 for THCA, CBDA, and CBCA; 0.878 for CBGA) to account for decarboxylation.

Statistical analysis was done in R version 3.5.1 (R Core Team, 2013). The library agricolae (de Mendiburu, 2014) was used for Tukey mean separation and significance tests. Split-plot ANOVA was modeled with "Treatment" as the main-plot factor and "Cultivar" as the sub-plot factor using the Satterthwaite approximation for calculating degrees of freedom with the following equation:

$$Trait_{ijk} = \mu + Rep_i + Treatment_k + \eta_{ik} + Cultivar_j + (Cultivar \times Treatment)_{ij} + \varepsilon_{ijk}$$ (**Equation 3.2)**

Neutral:total cannabinoid ratio was calculated as the average of the concentration of neutral forms (CBD, THC, CBC, and CBG) divided by the corresponding total as calculated above.

3.5 Results

Cannabinoid accumulation over time in three *C. sativa* cultivars grown in unstressed conditions

Total potential CBD, THC, CBC, and CBG increased over time in the unstressed control plots, achieving maximum concentrations of 7.5-12% total CBD by week three (Figure 3.2). GVA-H-19-1039 had greater total CBC concentration at week 3 than ‘T2’ (Figure 3.2c, Tukey α=0.05). Total CBG levels followed a pattern of accumulation from the other cannabinoids, reaching a maximum in week 2 for GVA-H-19-1039 and ‘TJ’s CBD’ and then declining slightly, while in ‘T2’ total CBG concentration continued to increase through week 3 (Figure 3.2d).

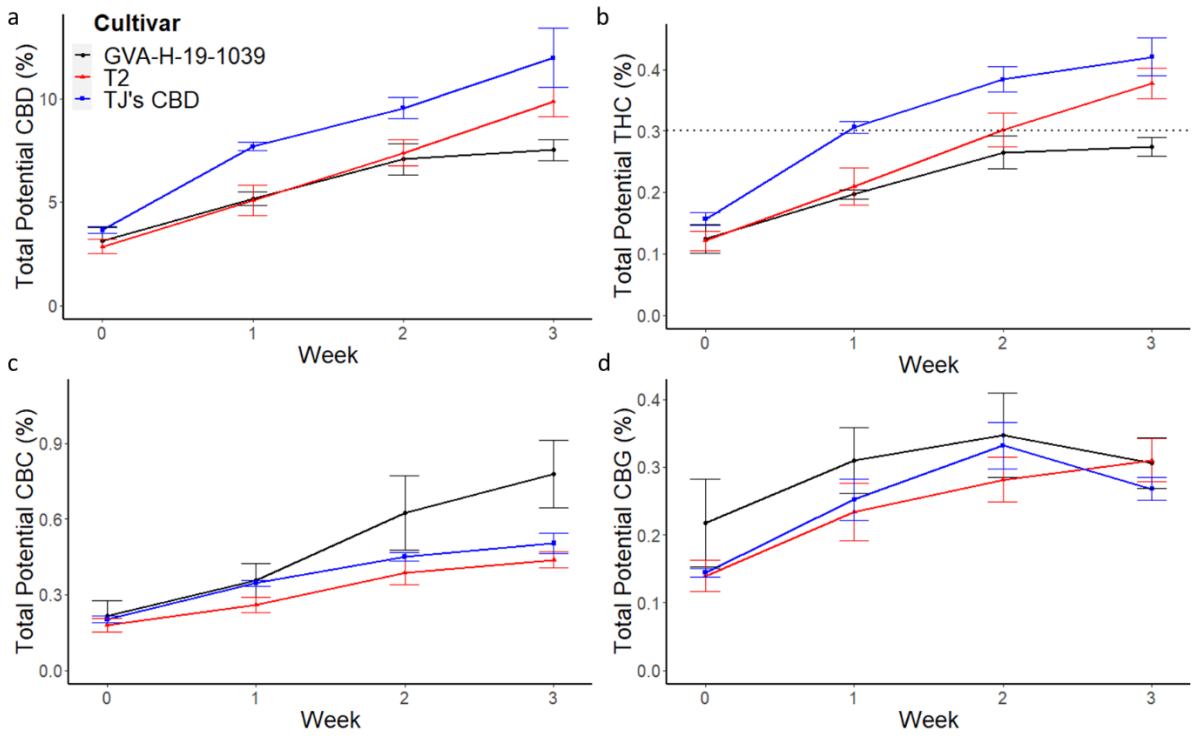

Figure 3.2. Cannabinoid accumulation by cultivar over three weeks in the unstressed control treatment. Week 0 refers to samples harvested immediately prior to stress application in the other treatment blocks. a) Total potential CBD (%), b) total potential THC (%), dotted line is 0.3%, c) total potential CBC (%), and d) total potential CBG (%). Error bars are standard error (n=4).

Cannabinoid ratios over time in three *C. sativa* cultivars grown in unstressed conditions

The ratio of total potential CBD to THC is of great importance to hemp growers as high CBD and low THC is desired to maintain regulatory compliance and maximize yield. The range of mean values of the CBD:THC ratio by cultivar in the unstressed control treatment was 23.3-28.2 (Figure 3.3a). There was a significant effect of cultivar (percent variation explained, PVE=14%, $p<0.01$) and week (PVE=34%, $p<0.01$) on CBD:THC ratio, but no interaction effect ($p=0.19$) (Table 3.1).

The ratios of CBD to CBC were significantly different between cultivars, reflecting differences in total CBC abundance, and had a range of 10.2-28.2 (Figure 3.2c, Figure

3.3b, ANOVA α=0.05). GVA-H-19-1039 had a substantially lower CBD:CBC ratio at harvest, averaging 10.2 vs 22.4 for 'T2' and 23.5 for 'TJ's CBD'. The CBD to CBG ratio across all weeks were significantly different by cultivar (PVE = 36%, $p<0.01$, Figure 3.3c), and the range of mean CBD to CBG ratios was 15.6-45.0. The ratio of neutral:total cannabinoids decreased substantially after week 0, and the range of mean ratios across all time points was 12-52% (Figure 3.3d).

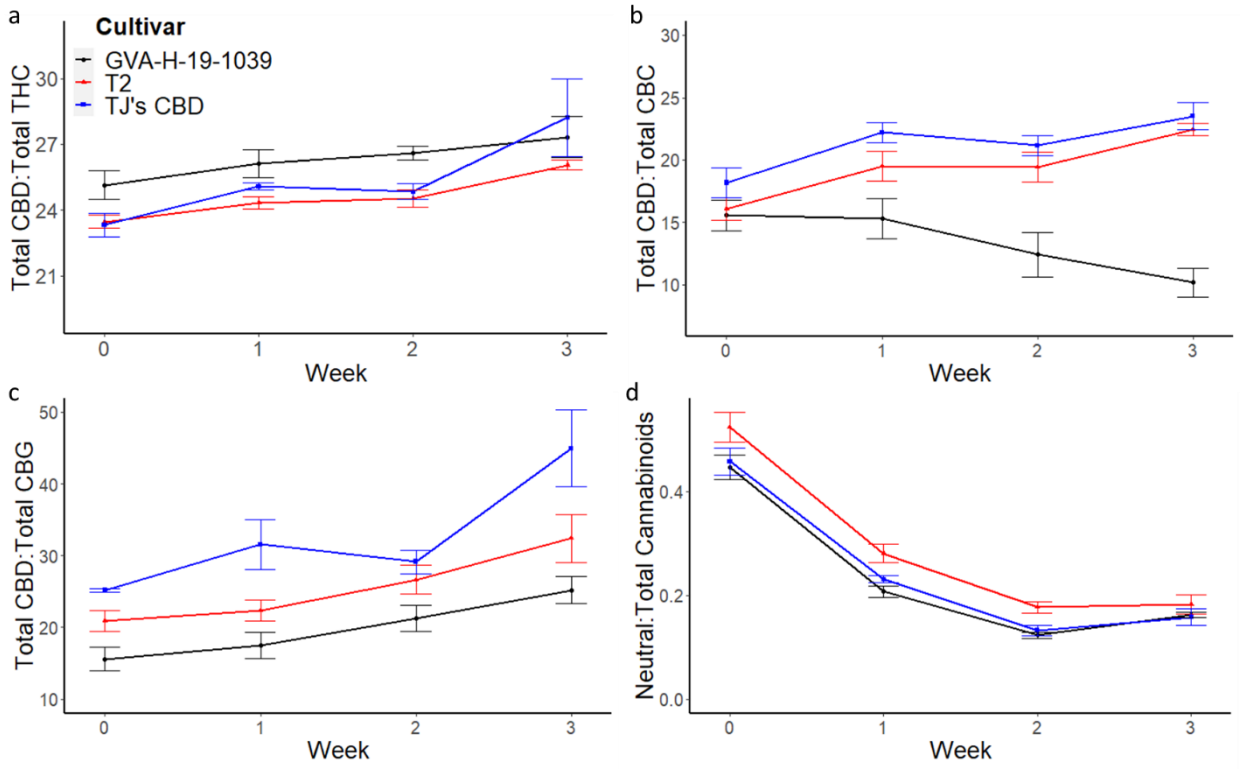

Figure 3.3. Cannabinoid ratios over three weeks with respect to cultivar in the unstressed control treatment. Week 0 refers to samples harvested immediately prior to stress application in the other treatments blocks. a) Total CBD:total THC, b) total CBD:total CBC, c) total CBD:total CBG, and d) neutral:total cannabinoids. Error bars are standard error (n=4).

Table 3.1 ANOVA of CBD:THC ratio in the unstressed control treatment.

Source of variation	Degrees of freedom	Sum of squares	PVE (%)	F value	Pr(>F)
Cultivar	2	23	14	6.1	<0.001
Week	1	58	35	30	<0.001
Cultivar : Week	2	6.9	3.9	1.7	0.19

Genotype-by-environment interaction of cultivar and stress treatment

Using a split-plot mixed linear model in ANOVA, there was no significant (α=0.01) cultivar × treatment effect at any sampling points for total CBD, total THC, total CBC, total CBG; cannabinoid ratios of CBD:THC, CBD:CBC, CBD:CBG; and neutral:total cannabinoid ratio. Due to the lack of cultivar-specific response, cultivars were combined to examine stress effects.

Cannabinoid accumulation in response to stress treatments

When considering all data across all weeks by stress treatment, herbicide application was the only treatment that lead to a statistically significant reduction in total potential CBD (Figure 3.4a). Similar reductions in the herbicide-treated group were seen for THC ($p<0.01$), and CBG ($p<0.01$) compared to the unstressed control, but no significant reduction was found for CBC. The concentration of cannabinoids in each week were similar for each treatment with the exception of herbicide treatment, which was consistently lower (Figure 3.4). Mean cannabinoid concentrations for plants in the wounding treatment was often greater than other treatment blocks but this effect was not statistically significant at any time point ($p>0.05$), except against herbicide-treated blocks.

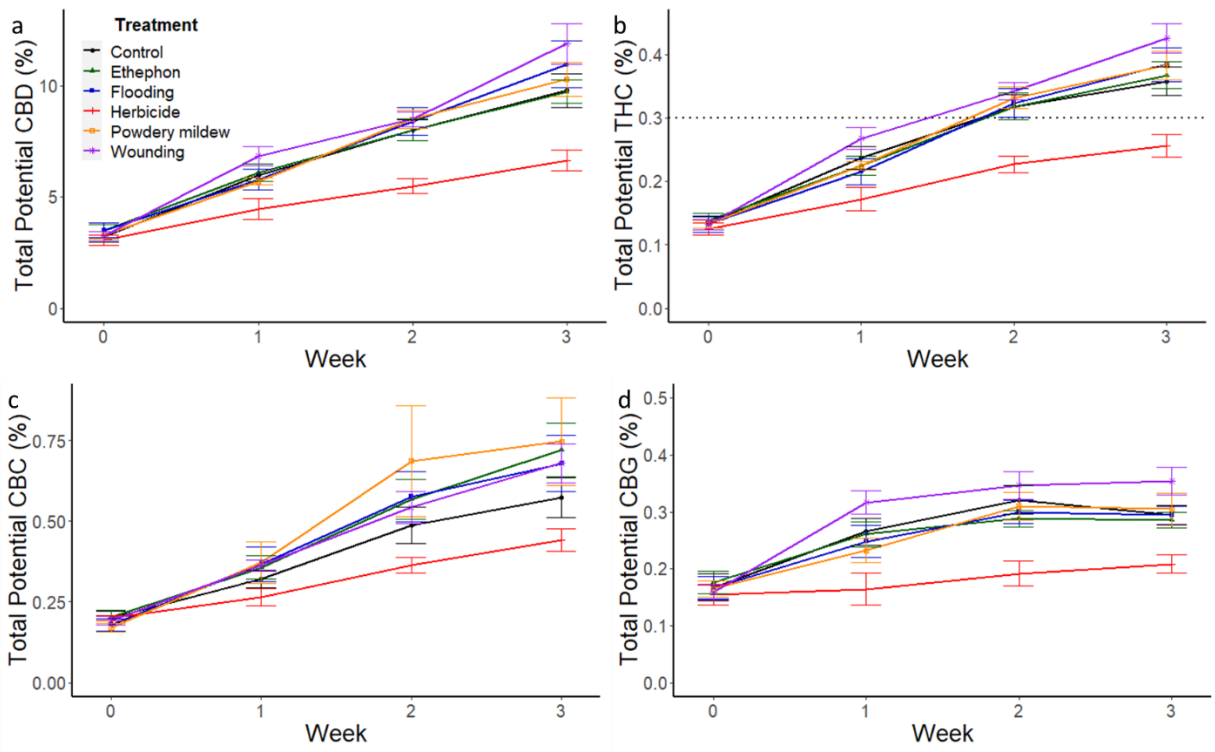

Figure 3.4. Cannabinoid accumulation in response to stress treatments over three weeks. Week 0 refers to samples harvested immediately prior to initial stress application. a) Total potential CBD (%), b) total potential THC (%), dotted line is 0.3%, c) total potential CBC (%), and d) total potential CBG (%). Means represent all cultivars combined. Error bars are standard error (n=12).

The effects of stress treatments on cannabinoid ratios

The range of mean CBD:THC ratios in stress-treated plants over the course of the trial were similar to the range of mean ratios in the cultivars grown in the control treatment: 24.0-28.2 (Figure 3.5a). There was an unexplained treatment block effect on CBD:THC ratio before treatment was applied (week 0, ANOVA $p<0.01$), largely due to a high CBD:THC ratio in the group that was intended to be flooded compared to the other treatment blocks (Tukey $\alpha=0.05$). There was also a treatment block effect on CBD:CBC ratio at week 0 and harvest only (Figure 3.5b, Table 3.2). There was no effect of treatment on CBD:CBG ratio at any timepoint except week 1, where herbicide treatment was

significantly higher than wounding treatment (Tukey α=0.05), although neither was significantly different than the control treatment. The neutral:total cannabinoid ratio was unaffected by the stress treatments (Figure 3.5d, p>0.05) at any time point.

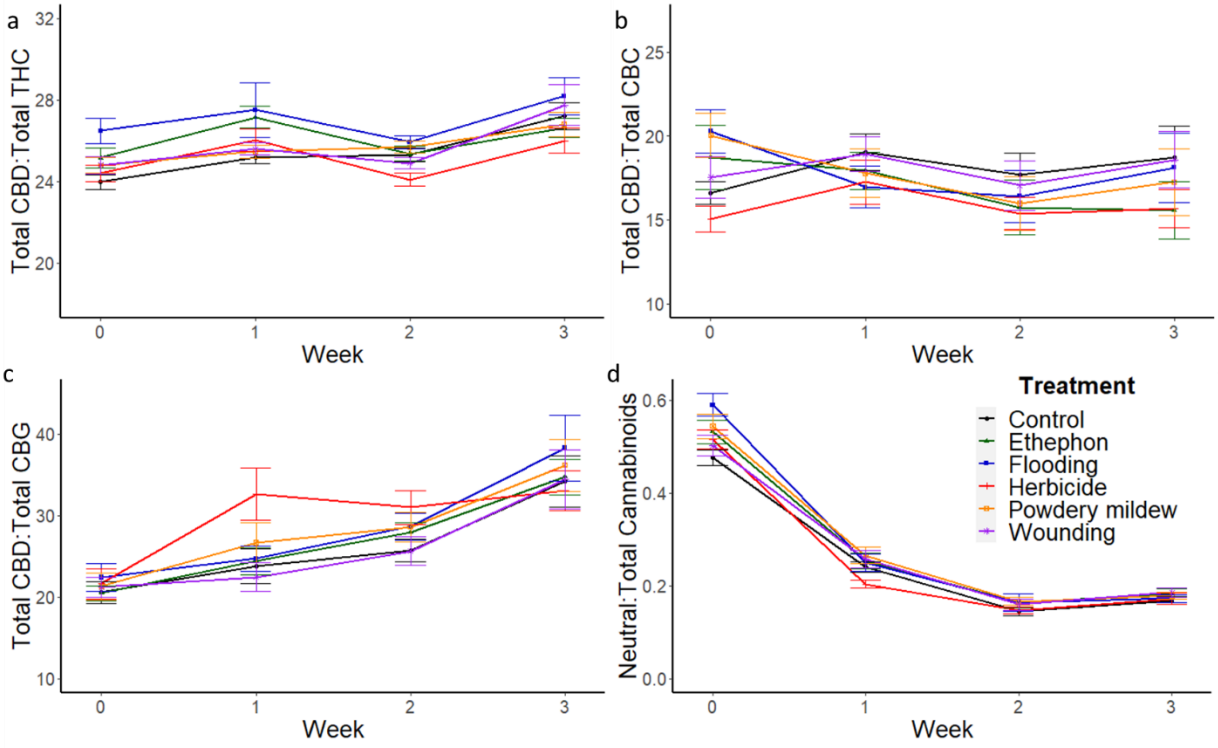

Figure 3.5. Cannabinoid ratios in response to stress treatment over three weeks. a) Total CBD:total THC, b) total CBD:total CBC, c) total CBD:total CBG, and d) neutral:total cannabinoids. Means represent all cultivars combined. Error bars are standard error (n=12).

Cannabinoid profiles at harvest

At the prospective harvest date, the only significant difference in any comparisons of THC levels were significantly lower levels in the herbicide treated plants (Figure 3.6a, Table 3.2). At this time point, there was no significant difference in CBD:THC ratios between plants exposed to any stress treatment (Figure 3.6b, Table 3.2). Stress treatment at harvest affected other measured cannabinoids (due to lowered production in the herbicide treated

plants) but not ratios of CBD:CBG or neutral:total cannabinoids (Figure 3.4, Figure 3.5, Table 3.2).

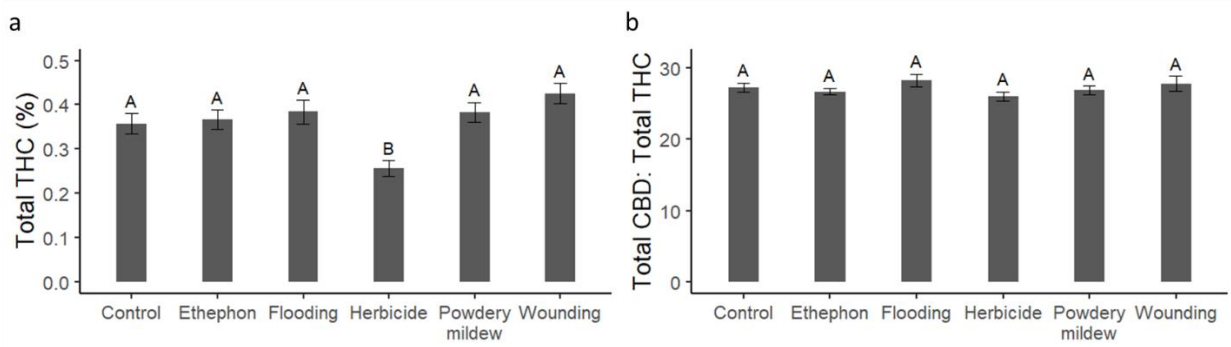

Figure 3.6. Bar chart of key measures at harvest. a) Total potential THC % and b) total potential CBD:THC ratio. Letters are Tukey HSD post-hoc levels ($p<0.05$).

Table 3.2. P-values from split-plot ANOVA results of traits at harvest.

	Treatment	Cultivar	Cultivar x Treatment	Rep
Total Potential CBD (%)	<0.001	0.0041	0.45	0.57
Total Potential THC (%)	<0.001	<0.001	0.12	0.45
Total Potential CBC (%)	0.025	<0.001	0.078	0.68
Total Potential CBG (%)	<0.001	<0.001	0.088	0.41
Total CBD:Total THC	0.44	0.29	0.74	0.94
Total CBD:Total CBC	0.023	<0.001	0.15	0.48
Total CBD:Total CBG	0.72	<0.001	0.94	0.43
Neutral:Total Cannabinoids	0.58	<0.001	0.029	0.22

3.6. Discussion

Cannabinoid accumulation and ratios

CBD, THC, and CBC accumulated during maturation of the inflorescence over the course the trial, as expected (Stack et al., 2021). Most plots sampled at week three (except herbicide treated plots) had a total THC concentration >0.3%. Considering all data in this

study, there is a strong linear relationship between total CBD and total THC (Pearson's r=0.98). Given this linear relationship, samples with > 8% total CBD would be expected to have >0.3% total THC. This level is slightly greater than the 6% CBD critical value reported in my previous study (Chapter 2, Toth et al., 2020), and may be due to differences in CBD:THC ratio in the cultivars tested or improvements in sample handling that resulted in reduced cannabinoid degradation. The critical value found here is in close agreement with Stack et al. (2021), which involved multiple cultivars in two different locations.

There was no effect of stress treatment on the total CBD to total THC ratio at harvest, supporting the hypothesis that this ratio is genetic and not strongly influenced by environmental stress. The variation in CBC in 'TJ's CBD' and 'T2' was also largely explained as a function of CBD, at a rate of about 19:1 CBD:CBC (Pearson's r=0.94). This corroborates data from heterologous expression of CBD synthase in yeast, where CBC was produced at a rate of about 5% of CBD (Zirpel et al. 2018). However, at harvest GVA-H-19-1039 had significantly greater total CBC than would be expected from this mechanism considering the CBD concentration. It is possible that GVA-H-19-1039 expresses an additional cannabinoid synthase enzyme, as other hemp plants have been noted to express additional cannabinoid synthase enzymes including a dedicated CBCA synthase (Kojoma et al., 2006; Laverty et al., 2019; Weiblen et al., 2015). It is unclear if this high CBC phenotype is the same "prolonged juvenile chemotype" leading to high CBC noted by de Meijer et al. (2009). In contrast to the high proportion of CBC observed at the beginning of flowering by de Meijer et al. (2009), GVA-H-19-1039 had a much lower proportion of CBC at the equivalent early time point. Further, de Meijer et al. (2009) reported a decrease in the proportion of CBC over time, whereas the proportion of CBC in GVA-H-19-1039

increased in successive weeks in this study. The minor effect of stress treatment noted in table 3.2 on total potential CBC and CBD:CBC ratio may be due to altered regulation of CBDAS and CBCAS.

While Yang et al. 2020 found that the CBD:THC ratio decreased throughout floral development with autoflowering cultivars experiencing a secondary increase, my data suggest a stable or slight increase in total CBD:total THC ratio over the course of floral development. This discrepancy may have been due to differences in cultivar or testing, or yet-unidentified environmental effects. Results from other field trials suggests there is a stable CBD:THC ratio throughout the life of the plant (De Backer et al., 2012; Pacifico et al., 2008; Stack et al., 2021)

Decarboxylation

There are limited data on cannabinoid decarboxylation (neutral:total cannabinoids) *in planta* (Chapter 2, Toth et al., 2020; Yang et al., 2020). Decarboxylation is largely thought to be non-enzymatic, and suggested to be promoted by age, heat, light, and small molecule catalysts such as formic acid and methanol, but repressed by antioxidants (Perrotin-Brunel et al., 2011; Singh et al., 2020). The broad trend of high percentage decarboxylation (mostly neutral forms) early in flowering followed by a rapid drop is consistent with a previous study (Yang et al., 2020). The high initial decarboxylation percentage in young flowers may be a result of different chemical environments promoting decarboxylation in young inflorescence tissue.

Further studies

There were several limitations to this study. First, only regulatory-type shoot tip cannabinoid testing was undertaken, and it is possible that these stresses affected total yield, but not shoot tip cannabinoid concentration. Second, three cultivars were chosen including seeded and clonal cultivars, but there is certainly a wide range of hemp genetic diversity that has yet to be studied. Third, the stress treatments examined here were chosen to be representative of growing conditions in a wet northeast US climate but did not include stresses that are typical of other growing areas, such as drought, extreme heat, or high salinity. The stresses were also only applied at a single intensity, which may have been insufficient to elicit a response. Lastly, these data also only reflect a single site in a single year, and it is possible that year-to-year variation or differences in site could lead to different results. Nevertheless, the evidence provided here supports the conclusion that THC accumulation is proportional to that of CBD and is not strongly affected by environmental stress.

3.7 Acknowledgements

Other individuals who made substantial intellectual contributions to this work include Lawrence B. Smart, Christine D. Smart, George M. Stack, Craig H. Carlson, Glenn Philippe, and Jocelyn K.C. Rose. I am grateful to the Smart and Rose research teams for their support, especially Deanna Gentner, Rebecca Wilk, Stephen Snyder, Savanna Shelnutt, Allison DeSario, Teagan Zingg, Lauren Carlson, Ben DeMoras, and Sarah Wright. This work was supported by New York State Department of Agriculture and Markets through grants AC477 and AC483 from Empire State Development Corporation and by a sponsored research agreement with Pyxus International.

3.8 References

Adesso, M., Laser, P., & Mills, A. (2019). An overview of industrial hemp law in the United States. *UDC/DCSL L. Rev.*, *22*, 85.

Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype× environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosystems, Geosciences & Environment, 2*(1).

Caplan, D., Dixon, M., & Zheng, Y. (2019). Increasing inflorescence dry weight and cannabinoid content in medical cannabis using controlled drought stress. *HortScience, 54*(5), 964-969.

Colmer, T., & Voesenek, L. (2009). Flooding tolerance: suites of plant traits in variable environments. *Functional Plant Biology, 36*(8), 665-681.

Davidson, M., Reed, S., Oosthuizen, J., O'Donnell, G., Gaur, P., Cross, M., & Dennis, G. (2018). Occupational health and safety in cannabis production: an Australian perspective. *International Journal of Occupational and Environmental Health, 24*(3-4), 75-85.

De Backer, B., Maebe, K., Verstraete, A. G., & Charlier, C. (2012). Evolution of the content of THC and other major cannabinoids in drug-type cannabis cuttings and seedlings during growth of plants. *Journal of Forensic Sciences, 57*(4), 918-922.

de Meijer, E. P., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. C., Ranalli, P., & Mandolino, G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics, 163*(1), 335-346.

de Mendiburu, F. (2014). Agricolae: statistical procedures for agricultural research. *R package version, 1*(1).

Ding, W., Reddy, K. N., Krutz, L. J., Thomson, S. J., Huang, Y., & Zablotowicz, R. M. (2011). Biological response of soybean and cotton to aerial glyphosate drift. *Journal of Crop Improvement, 25*(3), 291-302.

Duke, S. O., & Powles, S. B. (2008). Glyphosate: a once-in-a-century herbicide. *Pest Management Science: formerly Pesticide Science, 64*(4), 319-325.

Gorelick, J., & Bernstein, N. (2017). Chemical and physical elicitation for enhanced cannabinoid production in cannabis. In *Cannabis sativa L.-Botany and Biotechnology* (pp. 439-456): Springer.

Kojoma, M., Seki, H., Yoshida, S., & Muranaka, T. (2006). DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" *Cannabis sativa* L. *Forensic Science International, 159*(2-3), 132-140.

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. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Research, 29*(1), 146-156.

Lydon, J., Teramura, A. H., & Coffman, C. B. (1987). UV-B radiation effects on photosynthesis, growth and cannabinoid production of two *Cannabis sativa* chemotypes. *Photochemistry and Photobiology, 46*(2), 201-206.

Lyu, D., Backer, R. G., Robinson, W. G., & Smith, D. L. (2019). Plant-growth promoting rhizobacteria for cannabis production: Yield, cannabinoid profile and disease resistance. *Frontiers in Microbiology, 10*, 1761.

Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., & de Meijer, E. (2003). Qualitative and quantitative aspects of the inheritance of chemical phenotype in *Cannabis*. *Journal of Industrial Hemp, 8*(2), 51-72.

Mansouri, H., & Asrar, Z. (2012). Effects of abscisic acid on content and biosynthesis of terpenoids in *Cannabis sativa* at vegetative stage. *Biologia Plantarum, 56*(1), 153-156.

Mansouri, H., Asrar, Z., & Szopa, J. (2009). Effects of ABA on primary terpenoids and Δ 9-tetrahydrocannabinol in *Cannabis sativa* L. at flowering stage. *Plant Growth Regulation, 58*(3), 269-277.

Mansouri, H., Salari, F., & Asrar, Z. (2013). Ethephon application stimulats cannabinoids and plastidic terpenoids production in *Cannabis sativa* at flowering stage. *Industrial Crops and Products, 46*, 269-273.

Mansouri, H., Salari, F., Asrar, Z., & Nasibi, F. (2016). Effects of ethephon on terpenoids in *Cannabis sativa* L. in vegetative stage. *Journal of Essential Oil Bearing Plants, 19*(1), 94-102.

Nir, S. M. (2019). Hemp or Pot Farm? Police and Thieves Can't Always Tell. *The New York Times*. https://www.nytimes.com/2019/12/03/nyregion/hemp-marijuana-thc.html

Pacifico, D., Miselli, F., Carboni, A., Moschella, A., & Mandolino, G. (2008). Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica, 160*(2), 231-240.

Perrotin-Brunel, H., Buijs, W., Van Spronsen, J., Van Roosmalen, M. J., Peters, C. J., Verpoorte, R., & Witkamp, G.-J. (2011). Decarboxylation of Δ9-tetrahydrocannabinol: Kinetics and molecular modeling. *Journal of Molecular Structure, 987*(1-3), 67-73.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for
Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ram, H. M., & Jaiswal, V. S. (1970). Induction of female flowers on male plants of *Cannabis sativa* L. by 2-chloroethanephosphonic acid. *Experientia, 26*(2), 214-216

Salari, F., & Mansori, H. (2013). The effect of jasmonic acid on the terpenoid compounds in Cannabis sativa. *Journal of Plant Process and Function, 1*(2), 51-60.

Salentijn, E. M., Zhang, Q., Amaducci, S., Yang, M., & Trindade, L. M. (2015). New developments in fiber hemp (*Cannabis sativa* L.) breeding. *Industrial Crops and Products, 68*, 32-41.

Savatin, D. V., Gramegna, G., Modesti, V., & Cervone, F. (2014). Wounding in the plant tissue: the defense of a dangerous passage. *Frontiers in Plant Science, 5*, 470.

Singh, A. P., Fathordoobady, F., Guo, Y., Singh, A., & Kitts, D. D. (2020). Antioxidants help favorably regulate the kinetics of lipid peroxidation, polyunsaturated fatty acids degradation and acidic cannabinoids decarboxylation in hempseed oil. *Scientific Reports, 10*(1), 1-12.

Stack, G. M., Toth, J. A., Carlson, C. H., Cala, A. R., Marrero-González, M. I., Wilk, R. L., Gentner, D. R., Crawford, J.L, Philippe, G,. Rose, J. K., Viands, D.R., Smart,

C.D., & Smart, L.B. (2021). Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy, 13*(4), 546-561.

Szarka, D., Tymon, L., Amsden, B., Dixon, E., Judy, J., & Gauthier, N. (2019). First report of powdery mildew caused by *Golovinomyces spadiceus* on industrial hemp (*Cannabis sativa*) in Kentucky. *Plant Disease, 103*(7), 1773-1773.

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K. & Smart, L.B. (2020). Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy, 12*(3), 213-222.

Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., & Marks, M. D. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytologist, 208*(4), 1241-1250. doi:10.1111/nph.13562

Weldon, W. A., Ullrich, M. R., Smart, L. B., Smart, C. D., & Gadoury, D. M. (2020). Cross-infectivity of powdery mildew isolates originating from hemp (*Cannabis sativa*) and Japanese hop (*Humulus japonicus*) in New York. *Plant Health Progress, 21*(1), 47-53.

Yang, R., Berthold, E., McCurdy, C. R., da Silva Benevenute, S., Brym, Z. T., & Freeman, J. H. (2020). Development of cannabinoids in flowers of industrial hemp (*Cannabis sativa* L.)—a Pilot Study. *Journal of Agricultural and Food Chemistry*.

Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *Journal of Biotechnology, 284*, 17-26.

Chapter 4: Flowering time loci and photoperiod insensitivity in hemp[3]

4.1 Overview

Flowering time is an essential phenotype for nearly every market class of hemp, as it directly relates to maturity. As shown in Chapter 3, cannabinoids increase in the weeks following the initiation of flowering, so understanding the factors affecting flowering time is essential to predict the timing of regulatory sampling and harvest. Previous work (Stack et al., 2021) suggested that there are at least two major loci affecting flowering time, a locus responsible for the photoperiod-insensitivity trait known as "autoflower" and a locus segregating for early flowering time in 'Umpqua'. As discussed in Chapter 2, high-throughput molecular marker assays are effective in screening seedling populations for traits that are not easily assayed phenotypically when locus effect sizes are large. Flowering time is not always easy to assay phenotypically, and the two previously identified loci have large effects, making this a prime opportunity to develop and apply molecular markers. Also discussed in Chapter 2 is the large effect that altering flowering time can have on yield, with later flowering cultivars having more time to add biomass that supports the production of harvestable flower or seed product.

In Chapter 2, it was shown that many CBD cultivar populations in 2019 were segregating at the *B* locus, leading to inconsistent cannabinoid profiles and failed compliance tests.

[3] This work will be submitted for publication as Toth, J.A., Stack, G.M., Carlson, C. H., & Smart, L.B. Identification and mapping of flowering time loci in *Cannabis sativa* L.

Here, I show that many other cultivar populations grown in 2020-2021 are segregating for flowering time. The development of molecular marker assays in Chapter 2 provided a way to easily select for cannabinoid chemotype and produce consistent, stable cannabinoid profiles, and it is hoped that the assays presented here can have a similar effect on creating cultivars stable for flowering time.

4.2 Abstract

Hemp (*Cannabis sativa,* THC<0.3%) is a crop with multiple market classes and many end uses. Flowering time is an important trait for most market classes, affecting yields and harvestability of grain, fiber, and cannabinoids. Hemp is usually considered a short-day plant, flowering once night length reaches a certain threshold. However, this threshold may vary based on genetics. Additionally, some plants in the species are photoperiod insensitive and often referred to as "autoflowering". This trait has anecdotally been known as a simple recessive trait and has major impacts on phenology and yield. Variations in flowering time within and across cultivars in outdoor grown populations have also been noted, likely corresponding to differences in critical night length. In this work, molecular mapping of the locus responsible for the "autoflower" trait (*Autoflower1)* and subsequent marker development was performed, as well as mapping a major effect locus, *Early1,* controlling early flowering time segregating 1:1 in the cultivar 'Umpqua'. Also detailed are the results of growing diverse cultivars in continuous light and the result of crossing two photoperiod insensitive cultivars of differing breeding history.

4.3 Introduction

Hemp (*Cannabis sativa*, low THC) is a multi-use crop that is often described as a photoperiod sensitive, short-day plant. Flowering time is important for most market classes, and uniform flowering dates within a cultivar are essential for ease of harvest. Fiber hemp benefits from a long growing season, as harvest usually occurs around the flowering date, and early flowering results in less vegetative growth time to accumulate biomass. Grain hemp must flower early enough such that grain can be harvested before frost if growing in temperate latitudes, but precocious flowering can lead to severe yield penalties due to a lack of time to accumulate biomass that provides photosynthate for grain filling. This is especially an issue for subtropical and tropical latitudes where most days of the year have nights longer than the critical night threshold. For cannabinoid production, as with grain production, precocious flowering may result in low yield, while plants that do not flower by the end of the season will produce low cannabinoid yields. Additionally, cannabinoid profiles change throughout the maturation of the inflorescence, making initiation of flowering an important factor in timing compliance testing and harvest (Toth et al., 2021).

Previous work has outlined differences in photoperiod threshold across different cultivars, in greenhouse as well as field conditions (Stack et al., 2021; Zhang et al., 2021). It has also been well established that some plants are photoperiod insensitive (day neutral), a trait proposed to have been introgressed from northern populations, which have been classified as a putative species, *Cannabis ruderalis* (McPartland, 2018). Photoperiod insensitivity is sometimes referred to colloquially as “autoflower” (Gloss, 2015). This trait has been suggested to be inherited in a simple, recessive, Mendelian fashion, but there are limited

data on this in the peer-reviewed literature (Green, 2005). A patent covering molecular markers and biotechnological manipulation of genes responsible for “autoflower” is held by Phylos Biosciences (Phylos Bioscience, International Patent WO 2021/097496 A2). Several grain cultivars have also been proposed to be photoperiod insensitive, but it is not clear if the genetic mechanism for photoperiod insensitivity in high cannabinoid cultivars is the same as in grain cultivars (Zhang et al., 2021).

Cornell hemp research has previously identified several populations marketed as F_1 hybrids (‘Umpqua’ and ‘Deschutes’) that have individuals with two distinct flowering times, approximately a month apart (Stack et al., 2021). This segregation ratio would be expected if one parent was heterozygous for a major effect flowering time locus while the other parent was homozygous at that locus. If this were the case, such a significant locus would be well suited for development of high-throughput molecular markers such as PACE (PCR Allele Competitive Extension, 3CR). As differences in flowering are not obvious when the plant is in an early vegetative state, molecular markers for this trait for early screening could be very useful (Toth et al., 2018).

For essentially qualitative traits controlled by major effect flowering time loci as described here, bulk segregant analysis (BSA) has been successfully used to map genes and generate molecular markers and related assays (Song et al., 2017). Bulk segregant analysis is a technique that utilizes the sequencing of pooled DNA samples from individuals with the same phenotype in contrasting groups in a segregating population, and has been used effectively in a range of crops including *C. sativa* (Ban & Xu, 2020; Welling et al., 2020). Bulk segregant analysis usually involves short-read sequencing and subsequent alignment to a reference genome, but BSA involving long reads and reference-free techniques have

been developed (Nordström et al., 2013; Segawa et al., 2021). The number of individuals in the pools must be sufficiently large to randomize the association of all regions of the genome except the region or regions associated with the trait of interest. Compared with other methods of mapping, this technique has the advantage of obtaining whole genome sequences of the region of interest, alleviating the issue of ascertainment bias present in other methods of sequencing mapping populations such as single nucleotide polymorphism (SNP) chips or genotyping-by-sequencing (GBS). It is also cheaper and results in higher read depth than individually sequencing genomes, but multiple sequencing efforts would be required for mapping more than one trait. Bulk segregant analysis can also be conducted using pre-defined molecular markers instead of direct sequencing, but the decreasing cost of sequencing has made these approaches less common (Becker et al., 2011; Zhang et al., 2009).

Once sequencing data are obtained from contrasting pools, a comparison of regions that differ in allelic frequency can be performed (Magwene et al., 2011). In the case of a simple recessive trait in an F_2 population, one pool should be homozygous for a region containing the causative gene, while the other pool should have an alternative allelic frequency of ~33% in that region. In the case of a major gene in a backcross, one pool should be homozygous in a region and the other pool should be heterozygous. The difference between allele frequencies can be represented in a number of ways, including comparing the number of significantly different SNPs in a region determined through Fisher's Exact test, or a G-test statistic, or through the delta-SNP method, also known as the delta-allele method, which involves determining the difference in allele frequency directly (Zhang & Panthee, 2020).

Differences in flowering time have been noted across different hemp cultivars, both due to differing photoperiod threshold and photoperiod independent mechanisms (Zhang et al., 2021). It is a well-established phenomenon that there is significant population structure in *C. sativa*, associated at least in part with recent breeding history and geography (Carlson et al., 2021). While there is still debate on the specifics of the nature of this population structure (hindered in part because of the ease of intercrossing between subgroups of *C. sativa*), there is strong support for at least two subpopulations, which have been described as subspecies (McPartland, 2018). The two subspecies that have been described differ in end use and likely origin, with *C. sativa* ssp. *sativa* grown for grain and fiber originally in European northern latitudes and *C. sativa* ssp. *indica* grown for cannabinoid production originally in Southeast Asia, including India. Various other subpopulations have been described, including a distinct clade of *C. sativa* with geographic origins in China (Carlson et al., 2021; Ren et al., 2021). Different taxonomic classifications have been proposed, including *C. ruderalis*, which has been considered the source of the autoflower trait in all *C. sativa* populations. Several grain cultivars, such as 'FINOLA', have been referred to as autoflowering in the literature (Van Bakel et al., 2011), but have a distinct phenotype compared to photoperiod insensitive high-cannabinoid cultivars in that the height of mature 'FINOLA' depends greatly on latitude (being shorter at lower latitudes), while autoflowering high-cannabinoid cultivars do not appear to exhibit this phenomenon (Callaway, 2002; Stack et al., 2021; Yang et al., 2020). It is unclear if the gene(s) controlling photoperiod insensitivity in these populations is the same with differing effects of genetic background, or if there are multiple loci controlling photoperiod insensitivity in

C. sativa. To date there have been limited studies that can resolve the genetic basis of photoperiod insensitivity across diverse germplasm.

Genetic pathways for the induction of flowering are fairly well conserved across dicot plants, with major photosensory, thermosensory, and age-related pathways converging on major floral integrator genes, including transcription factors that result in the expression of floral meristem identity genes (Jung & Müller, 2009). A recent genome-wide association study (GWAS) in fiber hemp implicated genes in all of these pathways as well as a host of transcription factors in time to flower (Petit et al., 2020). More study on the genetics of flowering time control in hemp is required for predictive breeding efforts.

4.4 Materials and methods

Field and greenhouse trials of populations segregating for flowering time

An F_2 population segregating for photoperiod insensitivity was developed by first crossing a female autoflower plant from a feminized seed lot numbered KG9202 (generously provided by Kayagene, Hollister, CA) with a late flowering, photoperiod sensitive 'Otto II' plant (generously provided by Edgar Winters, WinterFox Farms, Klamath Falls, OR) determined to be male and cannabinoid chemotype III using molecular markers (Toth et al., 2020) to produce F_1 family GVA-H-19-1148. These cultivars were previously trialed in the 2019 Cornell high-cannabinoid hemp field trial (Stack et al., 2021). One selected photoperiod sensitive female F_1 plant (GVA-H-19-1148-002) was multiplied by rooting stem cuttings, then one ramet was treated with STS to induce male flowers that self-pollinated multiple female plants to generate F_2 seed labelled GVA-H-20-1080 (Carlson et al., 2021).

A second population segregating for photoperiod insensitivity was ‘TJ’s CBG’ (generously provided by Stem Holdings Agri, Eugene, OR), which was evaluated in the 2020 Cornell CBG hemp field trials and displayed a CBG-dominant chemotype (chemotype IV).

For initial assessment of photoperiod insensitivity in the segregating populations, seeds of each were sown in potting mix in 50-cell SureRoot trays on December 16, 2019 and grown in a greenhouse with a 16:8 light:dark schedule. Eighty-eight healthy plants in each population were transplanted to one-gallon pots on February 3, 2020. While flowering was evident on some plants at this point, rating for terminal flowering as previously defined (Stack et al., 2021) was completed on March 23, 2020.

The high-CBD cultivar ‘Umpqua’ (generously provided originally by Industrial Seed Innovations) was grown in the Cornell high-cannabinoid hemp field trials in 2019 and 2020 and flowering time was carefully assessed. Details about the 2019 trial are available in (Stack et al., 2021). The 2020 trial was executed using similar protocols, but with a different seed lot of ‘Umpqua’ generously provided by Arcadia Biosciences (Davis, CA). An additional 100 plants taken equally from both seed lots were planted on July 22, 2021 in a Cornell field trial in Geneva NY using similar protocols, in a dedicated flowering time field trial.

The 2021 flowering time field trial also included 96 individuals from the KG9202 × ‘Otto II’ F_2 population GVA-H-20-1080 and 26 plants of ‘Hempress’ (generously provided by Point3 Farma, Center, CO). Flowering time was noted as described in (Stack et al., 2021). Wet biomass was weighed and recorded for each plant in population GVA-H-20-1080. Additional populations segregating for photoperiod insensitivity were identified in the 2020 Cornell CBG hemp field trial.

To evaluate photoperiod insensitivity across diverse germplasm, 50 seeds of each population were sown in potting mix in 50-cell SureRoot trays on April 20, 2021 unless otherwise noted and grown in a greenhouse under continuous supplemental lighting from high pressure sodium lamps. Flowering was noted weekly. Male flowering was considered to have started when the length of internodes at the apex of the plant shortened and male buds were clearly visible at the growing tip.

A complementation cross was completed between two photoperiod-insensitive plants: male 'Picolo' (generously provided by Hemp Genetics International, Saskatoon, SK), and a female plant homozygous for *Autoflower1* from the 'La Crème' cultivar population (generously provided by Ventura Seed Company, Camarillo, CA). The F_1 plants from this cross were grown under 16L:8D in 50-cell SureRoot trays alongside known photoperiod sensitive and insensitive cultivars. Ten plants from each population were established on January 5, 2022. 'Auto CBD' and 'La Crème' are feminized populations with no males. 'RN16' is a dioecious long-day high CBD hemp cultivar described in Stack et al., 2021.

Bulk segregant analysis sequencing

DNA was extracted using a Qiagen DNeasy 96-well kit from young leaf tissue collected from plants in population GVA-H-20-1080 and dried on silica gel. Two pools were created by combining equal amounts of DNA from 28 flowering, photoperiod insensitive plants and 25 non-flowering, photoperiod sensitive plants. Illumina TruSeq libraries with an insert size of ~500 bp were constructed for each pool by the Cornell Institute of Biotechnology then paired end 151 bp sequencing was performed on the Illumina NextSeq 2000 platform with coverage of ~35X.

DNA was extracted from dried, milled floral biomass samples of ‘Umpqua’ as previously described (Toth et al., 2020) for 15 early-flowering and 15 late-flowering plants from the 2019 and 2020 trials and from 15 early-flowering and 15 late-flowering samples from the 2021 flowering time trial. Illumina TruSeq libraries were constructed for each phenological pool and sequenced on the Illumina NextSeq 2000 platform, as described above.

Reads were aligned to the CBDRx-CS10 (GCF_900626175.2) genome assembly (Grassa et al., 2021) using Geneious Prime software (Biomatters, Inc., San Diego, CA) using the Geneious mapper at the fastest speed with three iterations. Variants were also called in the Geneious Prime environment, with a minimum coverage of 3 and a minimum variant frequency of 0.05. Variant calls were exported as csv files and modified using custom Python script to be compatible with PyBSASeq (Zhang & Panthee, 2020). PyBSASeq was run for each chromosome as well as all chromosomes using the “BulksOnly” protocol, assuming an F_2 population structure for GVA-H-20-1080 and a backcross population structure for ‘Umpqua’ bulks.

PACE genotyping assays

PACE assays were designed manually in the Geneious Prime environment. PACE reactions were run according to the product manual (3CR). Polymorphic SNP in the *Autoflower1* region identified as perfectly associated with photoperiod phenotype in GVA-H-20-1080 pools were converted to PACE markers and assayed across multiple populations including the individual plants that formed the pool, a field grown population of GVA-H-20-1080, cultivar populations segregating 3:1 late:early grown under field conditions, diverse cultivars including stable autoflower cultivars, and representatives from several clades of high-cannabinoid cultivars.

Table 4.1. PACE primers designed for the *Autoflower1* (AUTO) and *Early1* (EARLY) loci.

Name	Sequence	Group/ SNP Location
AUTO-1-FAM	GAAGGTGACCAAGTTCATGCTATCC AGGGTCTGGCTTTAAAAA	WT
AUTO -1-HEX	GAAGGTCGGAGTCAACGGATTATCC AGGGTCTGGCTTTAAAAT	*Autoflower1*
AUTO-1-REV	CCATAAAATGATAAGTACACTCTAC	18464905
AUTO-2-FAM	GAAGGTGACCAAGTTCATGCTTTGG ACTTCACCAAATGAGCCC	WT
AUTO-2-HEX	GAAGGTCGGAGTCAACGGATTTTGG ACTTCACCAAATGAGCCT	*Autoflower1*
AUTO-2-REV	CTTCTAACCCTTTGCATGAATG	19701425
AUTO-3-FAM	GAAGGTGACCAAGTTCATGCTCACA AGAATAATGCCCAAGAT	*Autoflower1*
AUTO-3-HEX	GAAGGTCGGAGTCAACGGATTCACA AGAATAATGCCCAAGAC	WT
AUTO-3-REV	CCTAGGTTGACATAGCCACCA	19731625
AUTO-4-FAM	GAAGGTGACCAAGTTCATGCTTCTC ACTTTCTGTCTTTTTCCCT	*Autoflower1*
AUTO-4-HEX	GAAGGTCGGAGTCAACGGATTTCTC ACTTTCTGTCTTTTTCCCC	WT
AUTO-4-REV	TCACAGTCTCAACAGGAGTGG	19991224
AUTO-5-FAM	GAAGGTGACCAAGTTCATGCTTTTT CATTTTCGGTGGGGTTTC	*Autoflower1*
AUTO-5-HEX	GAAGGTCGGAGTCAACGGATTTTTT CATTTTCGGTGGGGTTTT	WT
AUTO-5-REV	GGTTGGATGTTTCAGCTGAAG	21536161
EARLY-1-FAM	GAAGGTGACCAAGTTCATGCTGGAT ACTAGCCACTAGAAAGGTTT	*Early1*
EARLY-1-HEX	GAAGGTCGGAGTCAACGGATTGGATA CTAGCCACTAGAAAGGTTG	WT
EARLY-1-REV	CGAAGGAGATAAAGACTGTGAG	41445929
EARLY-2-FAM	GAAGGTGACCAAGTTCATGCT GGAAGCGATGAGTGAGTTCT	*Early1*
EARLY-2-HEX	GAAGGTCGGAGTCAACGGATTGGA AGCGATGAGTGAGTTCA	WT
EARLY-2-REV	CTAGATCTTGGTTTGGTATCTCC	46288769

4.5 Results

Autoflower1 photoperiod insensitivity is a recessive Mendelian trait

Two populations segregating for photoperiod insensitivity (GVA-H-20-1080 and 'TJ's CBG') were planted under non-inductive, long day conditions (16H light:8H dark). In the GVA-H-20-1080 population, 28/88 plants flowered, and in the TJ's CBG population, 24/88 plants flowered. These data are not significantly different from 25% of the plants flowering (Chi-square $p > 0.05$), consistent with a recessive allele at a single gene I am designating *Autoflower1* that was homozygous in KG9202, heterozygous in the photoperiod sensitive F_1 progeny of KG9202 × 'Otto II', and segregating 1:2:1 in the GVA-H-20-1080 F_2 population. This also suggests that 'TJ's CBG' was produced by a cross between two parents heterozygous for *Autoflower1,* possibly the self-pollination of a plant heterozygous for *Autoflower1*. With respect to photoperiod insensitivity, *Autoflower1* is inherited in a simple, recessive fashion.

Mapping of the *Autoflower1* locus

Bulk segregant analysis of Illumina sequence pools of photoperiod sensitive and insensitive individuals showed clear statistical significance for the G-test statistic in a region of chromosome 1 (NC_044371.1) for population GVA-H-20-1080 (Figure 4.2). No other chromosome reached significance by this metric (Figure 4.2). The significant region associated with *Autoflower1* spanned chromosome 1 bases 17,740,001-22,940,001, with a highly significant peak centered around 18,590,001-19,690,001.

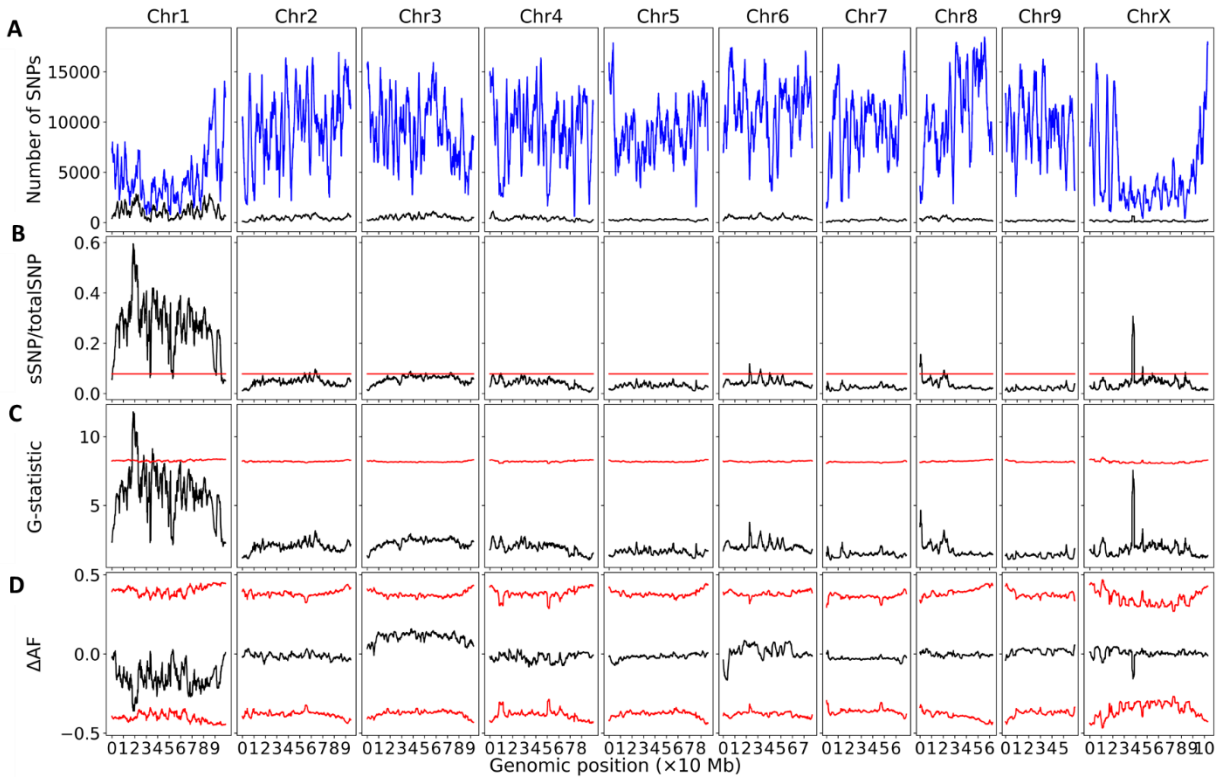

Figure 4.1. Bulk segregant analysis examining pools of photoperiod insensitive and photoperiod sensitive plants from GVA-H-20-1080. Window size is 2 Mb and step size is 10,000 bp. A) Number of SNP called per region. B) Ratio of significant SNP:Total SNP. C) G-statistic. D) Delta-allele (delta-SNP) frequency value. Red lines represent significance thresholds.

Autoflower1 candidate gene analysis

Within the G-statistic significant region of *Autoflower1* on chromosome 1 defined by the BSA of GVA-H-20-1080, 237 annotated genes were identified using the NCBI Genome Data Viewer. Of these, 75 were uncharacterized. Candidate genes potentially involved in controlling flowering time based on molecular function include: DOF zinc finger nucleases (LOC115704700, LOC115704742), nuclear transcription factor Y subunit B-1 (NFYB1, LOC115706176), floral homeotic protein APETALA 2 (AP2, LOC115708151), regulator of nonsense transcripts UPF2 (LOC115706264), zinc finger CCCH domain-containing protein 11 (LOC115706080), two-component response regulator-like PRR37

(LOC115705128), protein FAR1-RELATED SEQUENCE 5-like (LOC115703878, LOC115703890), and protein LONG AFTER FAR RED 3 (LOC115705698). The 237 genes within the significant region are detailed in Table 4.2.

Table 4.2. Annotated genes within the G-statistic significant region on Chromosome 1 for *Autoflower1* defined by bulk segregant analysis of GVA-H-20-1080.

Chromosome	Start	Stop	Gene symbol	Gene Name
NC_044371.1	17757915	17761814	LOC115705412	26S proteasome non-ATPase regulatory subunit 7 homolog A
NC_044371.1	17761898	17763259	LOC115705413	uncharacterized LOC115705413
NC_044371.1	17787969	17788075	LOC115708410	small nucleolar RNA R71
NC_044371.1	17809134	17811013	LOC115704794	aquaporin PIP1-2-like
NC_044371.1	17810460	17810532	TRNAK-CUU	
NC_044371.1	17860930	17861292	LOC115703844	uncharacterized LOC115703844
NC_044371.1	17861734	17870351	LOC115705898	uncharacterized LOC115705898
NC_044371.1	17904682	17904753	TRNAD-GUC	
NC_044371.1	17919663	17921702	LOC115706071	general transcriptional corepressor CYC8-like
NC_044371.1	17976187	17979044	LOC115708269	protein yippee-like
NC_044371.1	17994036	18001546	LOC115705688	bifunctional purine biosynthesis protein PurH
NC_044371.1	18004646	18014013	LOC115703846	uncharacterized LOC115703846
NC_044371.1	18013921	18018088	LOC115706360	proteoglycan 4
NC_044371.1	18018874	18021131	LOC115706362	clathrin light chain 2
NC_044371.1	18086527	18088357	LOC115706923	uncharacterized LOC115706923
NC_044371.1	18111469	18116010	LOC115703847	glucan endo-1,3-beta-glucosidase-like
NC_044371.1	18133132	18135405	LOC115707100	uncharacterized LOC115707100
NC_044371.1	18138092	18138795	LOC115703848	uncharacterized LOC115703848
NC_044371.1	18174222	18174294	TRNAK-CUU	
NC_044371.1	18174349	18177391	LOC115708200	dolichyl-phosphate beta-glucosyltransferase-like
NC_044371.1	18186228	18187327	LOC115703849	epsin-3-like
NC_044371.1	18187626	18190236	LOC115705535	actin
NC_044371.1	18217309	18219776	LOC115703850	CASP-like protein 4D2
NC_044371.1	18281470	18283194	LOC115703851	uncharacterized LOC115703851
NC_044371.1	18291896	18292757	LOC115703853	CASP-like protein 4D1
NC_044371.1	18296987	18297774	LOC115707613	CASP-like protein 4D1
NC_044371.1	18319699	18321222	LOC115704700	dof zinc finger protein DOF3.6-like
NC_044371.1	18381741	18387836	LOC115707297	uncharacterized LOC115707297
NC_044371.1	18393658	18394370	LOC115704699	uncharacterized LOC115704699

NC_044371.1	18397696	18399429	LOC115707222	uncharacterized LOC115707222
NC_044371.1	18400528	18402632	LOC115703854	adenosine deaminase-like protein
NC_044371.1	18431424	18433582	LOC115704742	dof zinc finger protein DOF3.6
NC_044371.1	18455778	18456591	LOC115707796	CASP-like protein 4D1
NC_044371.1	18463090	18463934	LOC115703855	CASP-like protein 4D1
NC_044371.1	18467789	18469378	LOC115703856	CASP-like protein 4D2
NC_044371.1	18481235	18483648	LOC115704910	actin
NC_044371.1	18483833	18485141	LOC115704911	epsin-3-like
NC_044371.1	18486944	18488635	LOC115703857	uncharacterized LOC115703857
NC_044371.1	18490400	18493444	LOC115707581	dolichyl-phosphate beta-glucosyltransferase
NC_044371.1	18493497	18493569	TRNAK-CUU	
NC_044371.1	18529902	18532213	LOC115706390	uncharacterized LOC115706390
NC_044371.1	18550573	18555132	LOC115705614	glucan endo-1,3-beta-glucosidase
NC_044371.1	18590256	18592335	LOC115706272	uncharacterized LOC115706272
NC_044371.1	18900705	18901809	LOC115704992	zinc finger protein ZAT12
NC_044371.1	18915701	18917143	LOC115703858	uncharacterized LOC115703858
NC_044371.1	18921169	18921275	LOC115708554	small nucleolar RNA R71
NC_044371.1	18940084	18942251	LOC115705890	WAT1-related protein At2g39510-like
NC_044371.1	18974401	18976327	LOC115705892	WAT1-related protein At2g39510
NC_044371.1	19004801	19004904	LOC115708453	small nucleolar RNA R71
NC_044371.1	19033360	19035330	LOC115705891	WAT1-related protein At2g39510
NC_044371.1	19068823	19070752	LOC115708167	WAT1-related protein At2g39510-like
NC_044371.1	19128938	19130917	LOC115704023	transcription factor bHLH114-like
NC_044371.1	19131885	19133568	LOC115704024	actin-depolymerizing factor 2
NC_044371.1	19134177	19137480	LOC115704021	protein ALP1-like
NC_044371.1	19143232	19144512	LOC115707008	protein UPSTREAM OF FLC
NC_044371.1	19145357	19148341	LOC115707046	nucleolin 1
NC_044371.1	19173526	19176502	LOC115704213	receptor-like protein kinase THESEUS 1
NC_044371.1	19176467	19178091	LOC115704214	F-box protein At2g39490
NC_044371.1	19193319	19195723	LOC115706369	photosynthetic NDH subunit of lumenal location 1, chloroplastic
NC_044371.1	19195927	19203939	LOC115706368	ABC transporter B family member 6
NC_044371.1	19214594	19216857	LOC115703860	uncharacterized LOC115703860
NC_044371.1	19217473	19223517	LOC115707202	transcription initiation factor TFIID subunit 6
NC_044371.1	19229654	19234420	LOC115707264	spliceosome-associated protein 130 A
NC_044371.1	19250006	19253484	LOC115706831	caffeoylshikimate esterase
NC_044371.1	19260772	19265413	LOC115706075	protein-tyrosine-phosphatase MKP1-like
NC_044371.1	19269185	19276537	LOC115706691	beta-hexosaminidase 1-like
NC_044371.1	19279579	19291479	LOC115705550	protein GRIP
NC_044371.1	19302758	19304050	LOC115705052	probable membrane-associated kinase regulator 4

NC_044371.1	19314102	19320767	LOC115703861	ankyrin repeat-containing protein ITN1-like
NC_044371.1	19342709	19347249	LOC115707983	protein-tyrosine-phosphatase MKP1
NC_044371.1	19347725	19349376	LOC115707988	uncharacterized LOC115707988
NC_044371.1	19354466	19362100	LOC115707986	beta-hexosaminidase 1
NC_044371.1	19368217	19380104	LOC115707984	probable DNA double-strand break repair Rad50 ATPase
NC_044371.1	19381034	19403194	LOC115707987	probable membrane-associated kinase regulator 4
NC_044371.1	19411191	19415240	LOC115707985	ankyrin repeat-containing protein ITN1
NC_044371.1	19447336	19447440	LOC115708435	small nucleolar RNA R71
NC_044371.1	19453408	19453514	LOC115708463	small nucleolar RNA R71
NC_044371.1	19526596	19531022	LOC115706635	uncharacterized LOC115706635
NC_044371.1	19586800	19591181	LOC115706681	uncharacterized LOC115706681
NC_044371.1	19597631	19597737	LOC115708399	small nucleolar RNA R71
NC_044371.1	19623001	19626945	LOC115708189	protein NRT1/ PTR FAMILY 2.7
NC_044371.1	19670607	19672347	LOC115703863	uncharacterized LOC115703863
NC_044371.1	19675794	19679721	LOC115706683	protein NRT1/ PTR FAMILY 2.7-like
NC_044371.1	19691506	19696923	LOC115706176	nuclear transcription factor Y subunit B-1
NC_044371.1	19712612	19715469	LOC115704691	probable RNA-binding protein ARP1
NC_044371.1	19726723	19728921	LOC115708151	floral homeotic protein APETALA 2
NC_044371.1	19778639	19780198	LOC115703865	uncharacterized LOC115703865
NC_044371.1	19782063	19783840	LOC115703866	uncharacterized LOC115703866
NC_044371.1	19802609	19815150	LOC115706264	regulator of nonsense transcripts UPF2
NC_044371.1	19822088	19823007	LOC115703868	uncharacterized LOC115703868
NC_044371.1	19826131	19827204	LOC115703869	uncharacterized LOC115703869
NC_044371.1	19843513	19847204	LOC115706080	zinc finger CCCH domain-containing protein 11
NC_044371.1	19849983	19850489	LOC115703870	uncharacterized LOC115703870
NC_044371.1	19860264	19863668	LOC115703871	protein TONNEAU 1a-like
NC_044371.1	19985933	19992033	LOC115705128	two-component response regulator-like PRR37
NC_044371.1	19988482	19992665	LOC115705129	uncharacterized LOC115705129
NC_044371.1	20010950	20018438	LOC115704703	TBC1 domain family member 8B
NC_044371.1	20032520	20036951	LOC115705441	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase 2
NC_044371.1	20574051	20576803	LOC115705487	uncharacterized LOC115705487
NC_044371.1	20582024	20583639	LOC115704827	uncharacterized LOC115704827
NC_044371.1	20588649	20591700	LOC115706793	uncharacterized LOC115706793
NC_044371.1	20595436	20599191	LOC115703873	uncharacterized LOC115703873
NC_044371.1	20601551	20602941	LOC115708318	uncharacterized LOC115708318
NC_044371.1	20603094	20604601	LOC115708319	uncharacterized LOC115708319
NC_044371.1	20609630	20610957	LOC115704823	uncharacterized LOC115704823
NC_044371.1	20612149	20612975	LOC115708152	uncharacterized LOC115708152

NC_044371.1	20615998	20619859	LOC115708215	WD repeat-containing protein WRAP73
NC_044371.1	20624018	20631775	LOC115706210	nucleolar complex protein 2 homolog
NC_044371.1	20640845	20644771	LOC115706652	protein IQ-DOMAIN 1-like
NC_044371.1	20653407	20659939	LOC115705663	calcium-binding mitochondrial carrier protein SCaMC-1-like
NC_044371.1	20664332	20664739	LOC115707338	low temperature-induced protein lt101.2
NC_044371.1	20667500	20669307	LOC115704698	LOB domain-containing protein 1
NC_044371.1	20696892	20698904	LOC115708282	uncharacterized LOC115708282
NC_044371.1	20713556	20727975	LOC115705207	Golgi to ER traffic protein 4 homolog
NC_044371.1	20732258	20733326	LOC115705208	uncharacterized LOC115705208
NC_044371.1	20734426	20735177	LOC115703874	uncharacterized LOC115703874
NC_044371.1	20735420	20738200	LOC115703875	uncharacterized LOC115703875
NC_044371.1	20741920	20742444	LOC115707794	uncharacterized LOC115707794
NC_044371.1	20760091	20762582	LOC115703876	uncharacterized LOC115707794
NC_044371.1	20775753	20778199	LOC115703877	uncharacterized LOC115703877
NC_044371.1	20778199	20781809	LOC115706729	uncharacterized LOC115706729
NC_044371.1	20790932	20795500	LOC115706745	uncharacterized LOC115706745
NC_044371.1	20801164	20801270	LOC115708543	small nucleolar RNA R71
NC_044371.1	20816258	20818673	LOC115703878	protein FAR1-RELATED SEQUENCE 5-like
NC_044371.1	20824032	20824510	LOC115706783	uncharacterized LOC115706783
NC_044371.1	20830310	20833207	LOC115703879	uncharacterized LOC115703879
NC_044371.1	20833217	20836949	LOC115706770	uncharacterized LOC115706770
NC_044371.1	20848052	20848331	LOC115706785	uncharacterized LOC115706785
NC_044371.1	20852425	20858895	LOC115706767	pre-rRNA-processing protein TSR1 homolog
NC_044371.1	20861533	20868270	LOC115706769	phosphoglucomutase
NC_044371.1	20868998	20871337	LOC115706752	uncharacterized LOC115706752
NC_044371.1	20874609	20881142	LOC115706728	endoplasmic reticulum metallopeptidase 1-like
NC_044371.1	20885271	20889037	LOC115706773	mRNA-decapping enzyme subunit 2-like
NC_044371.1	20892287	20897961	LOC115706762	DNA polymerase epsilon subunit 3-like
NC_044371.1	20898688	20900527	LOC115703880	uncharacterized LOC115703880
NC_044371.1	20901023	20905614	LOC115706743	3-hydroxyisobutyryl-CoA hydrolase-like protein 2, mitochondrial
NC_044371.1	20944539	20947901	LOC115706755	aquaporin PIP2-2-like
NC_044371.1	20957532	20960672	LOC115703881	bifunctional dihydrofolate reductase-thymidylate synthase 1-like
NC_044371.1	20962955	20970736	LOC115706734	diaminopimelate decarboxylase 2, chloroplastic
NC_044371.1	20977107	20983589	LOC115706768	pre-rRNA-processing protein TSR1 homolog
NC_044371.1	20986209	20992938	LOC115706771	phosphomannomutase/phosphoglucomutase-like
NC_044371.1	20996324	20998378	LOC115703882	uncharacterized LOC115703882

NC_044371.1	20998925	20999638	LOC115706761	protein PXR1-like
NC_044371.1	21001989	21004329	LOC115706753	uncharacterized LOC115706753
NC_044371.1	21007607	21014141	LOC115706766	endoplasmic reticulum metallopeptidase 1-like
NC_044371.1	21021481	21025532	LOC115706748	mRNA-decapping enzyme subunit 2
NC_044371.1	21030259	21033631	LOC115706763	DNA polymerase epsilon subunit 3
NC_044371.1	21044054	21048463	LOC115706744	3-hydroxyisobutyryl-CoA hydrolase-like protein 2, mitochondrial
NC_044371.1	21082797	21086224	LOC115706754	aquaporin PIP2-2
NC_044371.1	21099518	21104416	LOC115706733	bifunctional dihydrofolate reductase-thymidylate synthase
NC_044371.1	21105580	21109352	LOC115706735	diaminopimelate decarboxylase 2, chloroplastic-like
NC_044371.1	21134331	21139980	LOC115703883	phosphatidylinositol/phosphatidylcholine transfer protein SFH3-like
NC_044371.1	21142406	21146635	LOC115706760	trafficking protein particle complex subunit 1
NC_044371.1	21147123	21147770	LOC115703884	uncharacterized LOC115703884
NC_044371.1	21152489	21155502	LOC115706764	uncharacterized LOC115706764
NC_044371.1	21155973	21157289	LOC115706749	caffeoylshikimate esterase
NC_044371.1	21157426	21161133	LOC115706727	WPP domain-associated protein
NC_044371.1	21165867	21168970	LOC115706732	asparagine--tRNA ligase, cytoplasmic 1
NC_044371.1	21171737	21172419	LOC115703886	sulfated surface glycoprotein 185
NC_044371.1	21178192	21184371	LOC115706736	patatin-like protein 6
NC_044371.1	21198455	21204613	LOC115706741	chorismate synthase, chloroplastic
NC_044371.1	21270041	21271053	LOC115703887	uncharacterized LOC115703887
NC_044371.1	21328132	21332291	LOC115706740	protein IQ-DOMAIN 1
NC_044371.1	21354815	21362577	LOC115706730	nucleolar complex protein 2 homolog
NC_044371.1	21371455	21375371	LOC115706772	WD repeat-containing protein WRAP73-like
NC_044371.1	21378084	21381142	LOC115706777	uncharacterized LOC115706777
NC_044371.1	21381497	21382484	LOC115703888	uncharacterized LOC115703888
NC_044371.1	21386169	21387515	LOC115706778	uncharacterized LOC115706778
NC_044371.1	21401203	21402937	LOC115706774	uncharacterized LOC115706774
NC_044371.1	21409533	21410611	LOC115706779	uncharacterized LOC115706779
NC_044371.1	21410796	21412220	LOC115706784	uncharacterized LOC115706784
NC_044371.1	21416708	21419512	LOC115706747	uncharacterized LOC115706747
NC_044371.1	21433547	21437041	LOC115706751	18S rRNA (guanine-N(7))-methyltransferase RID2
NC_044371.1	21437550	21440586	LOC115706756	general transcription factor IIF subunit 2
NC_044371.1	21447348	21462402	LOC115706737	beta-taxilin
NC_044371.1	21474635	21477538	LOC115706739	elongation factor 1-alpha
NC_044371.1	21477812	21479214	LOC115706758	uncharacterized LOC115706758
NC_044371.1	21479260	21479331	TRNAP-AGG	
NC_044371.1	21483096	21486104	LOC115706731	heat shock protein 83

NC_044371.1	21487600	21512554	LOC115706746	RNA pseudouridine synthase 5
NC_044371.1	21512850	21522694	LOC115706726	uncharacterized LOC115706726
NC_044371.1	21535217	21536302	LOC115703889	uncharacterized LOC115703889
NC_044371.1	21542941	21546746	LOC115706742	mediator of RNA polymerase II transcription subunit 4
NC_044371.1	21548540	21548654	LOC115708485	U5 spliceosomal RNA
NC_044371.1	21548989	21562933	LOC115705698	protein LONG AFTER FAR-RED 3
NC_044371.1	21576898	21580894	LOC115705816	serine carboxypeptidase-like 27
NC_044371.1	21587198	21589633	LOC115708138	tetraspanin-6
NC_044371.1	21631321	21636641	LOC115705883	MLO-like protein 12
NC_044371.1	21641087	21642766	LOC115707955	probable ubiquitin-conjugating enzyme E2 C
NC_044371.1	21646660	21649551	LOC115706419	rhomboid-like protein 19
NC_044371.1	21651759	21659430	LOC115705459	uncharacterized protein slr1919
NC_044371.1	21678269	21680923	LOC115704817	serine/threonine-protein kinase-like protein CCR2
NC_044371.1	21707280	21711213	LOC115705568	UDP-glucuronic acid decarboxylase 6
NC_044371.1	21744345	21748626	LOC115705371	UDP-glucuronic acid decarboxylase 6
NC_044371.1	21762341	21764108	LOC115706295	60S ribosomal protein L35
NC_044371.1	21783471	21786230	LOC115705010	psbP-like protein 1, chloroplastic
NC_044371.1	21786352	21788085	LOC115705011	60S ribosomal protein L23a
NC_044371.1	21798933	21806810	LOC115705469	sulfate transporter 4.1, chloroplastic
NC_044371.1	21844601	21844707	LOC115708356	small nucleolar RNA R71
NC_044371.1	21862431	21868753	LOC115707642	methionine--tRNA ligase, chloroplastic/mitochondrial
NC_044371.1	21868959	21873269	LOC115707643	UDP-glucuronic acid decarboxylase 6
NC_044371.1	21922249	21922869	LOC115707900	uncharacterized LOC115707900
NC_044371.1	22181332	22181435	LOC115708437	small nucleolar RNA R71
NC_044371.1	22253112	22253711	LOC115703890	protein FAR1-RELATED SEQUENCE 5-like
NC_044371.1	22256688	22258022	LOC115703891	putative nuclease HARBI1
NC_044371.1	22258693	22260784	LOC115703892	L10-interacting MYB domain-containing protein-like
NC_044371.1	22262273	22267661	LOC115705846	cell division control protein 48 homolog C-like
NC_044371.1	22293072	22298256	LOC115706223	cell division control protein 48 homolog C-like
NC_044371.1	22301422	22301994	LOC115703893	uncharacterized mitochondrial protein AtMg00810-like
NC_044371.1	22368938	22369380	LOC115703894	uncharacterized LOC115703894
NC_044371.1	22370786	22372795	LOC115703895	uncharacterized LOC115703895
NC_044371.1	22375502	22376137	LOC115704323	uncharacterized LOC115704323
NC_044371.1	22384916	22388552	LOC115705857	ATP-dependent Clp protease proteolytic subunit 3, chloroplastic
NC_044371.1	22398199	22399764	LOC115703896	uncharacterized LOC115703896
NC_044371.1	22449283	22450610	LOC115705147	cysteine proteinase inhibitor 12

NC_044371.1	22450959	22455201	LOC115705148	dr1-associated corepressor-like
NC_044371.1	22461826	22461932	LOC115708407	small nucleolar RNA R71
NC_044371.1	22464826	22470153	LOC115705221	KHG/KDPG aldolase
NC_044371.1	22470184	22471754	LOC115705220	protein NDH-DEPENDENT CYCLIC ELECTRON FLOW 5
NC_044371.1	22471722	22474796	LOC115705219	protein SINE1
NC_044371.1	22533037	22537761	LOC115707349	dr1-associated corepressor
NC_044371.1	22537889	22539186	LOC115707350	cysteine proteinase inhibitor 12
NC_044371.1	22563827	22567881	LOC115703897	uncharacterized LOC115703897
NC_044371.1	22596604	22599884	LOC115703899	ATP-dependent Clp protease proteolytic subunit 3, chloroplastic-like
NC_044371.1	22613465	22615178	LOC115708244	allene oxide synthase 3
NC_044371.1	22629036	22632528	LOC115706159	WAT1-related protein At5g07050
NC_044371.1	22672126	22673222	LOC115703900	uncharacterized LOC115703900
NC_044371.1	22674151	22693965	LOC115705549	importin-11
NC_044371.1	22715424	22720089	LOC115708272	uncharacterized LOC115708272
NC_044371.1	22743955	22750802	LOC115705717	3-ketoacyl-CoA synthase 19
NC_044371.1	22783056	22783160	LOC115708458	small nucleolar RNA R71
NC_044371.1	22789195	22789301	LOC115708379	small nucleolar RNA R71
NC_044371.1	22836060	22837352	LOC115704792	putative mitochondrial carrier protein PET8
NC_044371.1	22841731	22842155	LOC115707033	low temperature-induced protein lt101.2
NC_044371.1	22844971	22845690	LOC115703903	LOB domain-containing protein 1-like
NC_044371.1	22848599	22849205	LOC115703753	plant UBX domain-containing protein 10-like
NC_044371.1	22870648	22872672	LOC115708325	uncharacterized LOC115708325

Germplasm screening with *Autoflower1* molecular assays

Within the significant region, some SNP alleles that were homozygous in the photoperiod insensitive bulk and had an allele frequency of ~33% in the photoperiod sensitive bulk were converted to PACE assays and screened on diverse germplasm. A summary of the genotype group calls is presented in Table 4.3. Within segregating populations, the association of the *Autoflower1* marker assay with the photoperiod phenotype was considered to be perfect if the homozygous allelic group associated with photoperiod insensitive plants from the bulk was associated with photoperiod insensitive plants only, while photoperiod sensitive plants were either heterozygous or in the other homozygous allelic group.

Table 4.3. Genotype group calls by population. Seg*: segregating imperfectly. Seg†: Segregating perfectly.

				Primer set	AUTO-1	AUTO-2	AUTO-3	AUTO-4	AUTO-5
				SNP location	18464905	19701425	19731625	19991224	21536161
				Nearest Gene	mRNA-CASP-like protein 4D1	*NFYB1*	*AP2*	*PRR37*	Uncharacterized
Cultivar / Population	Source	*Autoflower1*	# Tested						
'Anka'	UniSeeds	No	4		A/A	C/C	C/C	T/T	T/T
'Bish Feral'	Bish Enterprises	?	8		A/A	C/C	C/C	C/C	T/T
'C16'	Arcadia	No	2		A/A	C/C	C/C	T/T	T/T
'CFX-2'	Hemp Genetics International	?	8		A/A	C/C	C/C	C/C	T/T
'Henola'	International Hemp	?	8		A/A	C/C	C/C	T/T	T/T
'Picolo'	Hemp Genetics International	?	16		A/A	C/C	C/C	Seg*	T/T
'Puma'	CN Kenaf and Hemp	No	4		A/A	C/C	C/C	Seg*	T/T
'RN13A'	Paul Smith Denver Co.	No	4		A/A	C/C	C/C	T/T	T/T
'RN17'	Paul Smith Denver Co.	No	8		A/A	C/C	C/C	Seg*	T/T
'Si-1'	CN Kenaf and Hemp	No	19		A/A	C/C	C/C	Seg*	T/T
'Canda'	Parkland Industrial Hemp Growers	?	8		A/A	C/C	C/C	C/C	T/T
Missouri Feral	John Fike (40.228, -94.56)	?	8		A/A	C/C	C/C	C/C	T/T
'Nebraska'	Winter Fox Farms	?	8		A/A	C/C	C/C	Seg*	T/T
'NWG-Elite'	New West Genetics	?	8		A/A	C/C	C/C	Seg*	T/T
'T2'	Boring Hemp Company	No	8		A/A	C/C	C/C	Seg*	T/T
'USO-31'	UniSeeds	?	8		A/A	C/C	C/C	C/C	T/T
'CBG Delight'	Flura	1/4	32		Seg†	Seg†	Seg†	Seg†	Seg*

‘H5’	American Hemp Co.	1/4	32	Seg*	Seg†	Seg†	Seg*	Seg*
‘Hempress’	Point3 Farma	1/4	24	Seg†	Seg†	Seg†	Seg*	Seg†
‘La Crème’	Ventura Seed Company	1/4	44	ND	Seg†	Seg†	ND	ND
GVA-H-20-1080	Cornell Hemp	1/4	184	Seg*	Seg†	Seg†	Seg†	Seg*
‘TJ's CBG’	Stem Holdings Agri	1/4	88	Seg†	Seg†	Seg†	Seg†	Seg*
‘Suver Haze’	Oregon CBD	Heterozygous	8	T/C	T/C	T/C	C/C	C/T
‘Umpqua’	Industrial Seed Innovations	Heterozygous	4	T/C	T/C	T/C	T/T	T/C
‘AD1010’	Phylos	Yes	4	T/T	T/T	T/T	T/T	Seg*
‘Alpha Explorer’	Phylos	Yes	4	T/T	T/T	T/T	T/T	Seg*’
‘Alpha Nebula’	Phylos	Yes	4	T/T	T/T	T/T	T/T	Seg*
‘Auto CBD’	Phylos	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Auto CBG’	Oregon CBD	Yes	4	T/T	T/T	T/T	T/T	Seg*
‘DNCBD’	Arcadia	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Dr. Chunk’	Kayagene	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Maverick’	Kayagene	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Purple Star’	Atlas Seeds	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Rincon’	Kayagene	Yes	4	T/T	T/T	T/T	T/T	Seg*
‘Sour Citron’	Kayagene	Yes	4	T/T	T/T	T/T	T/T	C/C
‘Sour RNA Seedless’ (Triploid)	Oregon CBD	Yes	4	T/T	T/T	T/T	T/T	C/C

Effect of *Autoflower1* genotype on agronomic performance

Ninety-six individuals of GVA-H-20-1080 grown in the 2021 flowering time field trial were genotyped at *Autoflower1* using AUTO-2 to determine the effect of heterozygosity grown under field conditions. There was a significant effect of the allelic group on flowering date, height, and biomass, with heterozygotes being intermediate with respect to flowering date, height, and wet biomass (Figure 4.2).

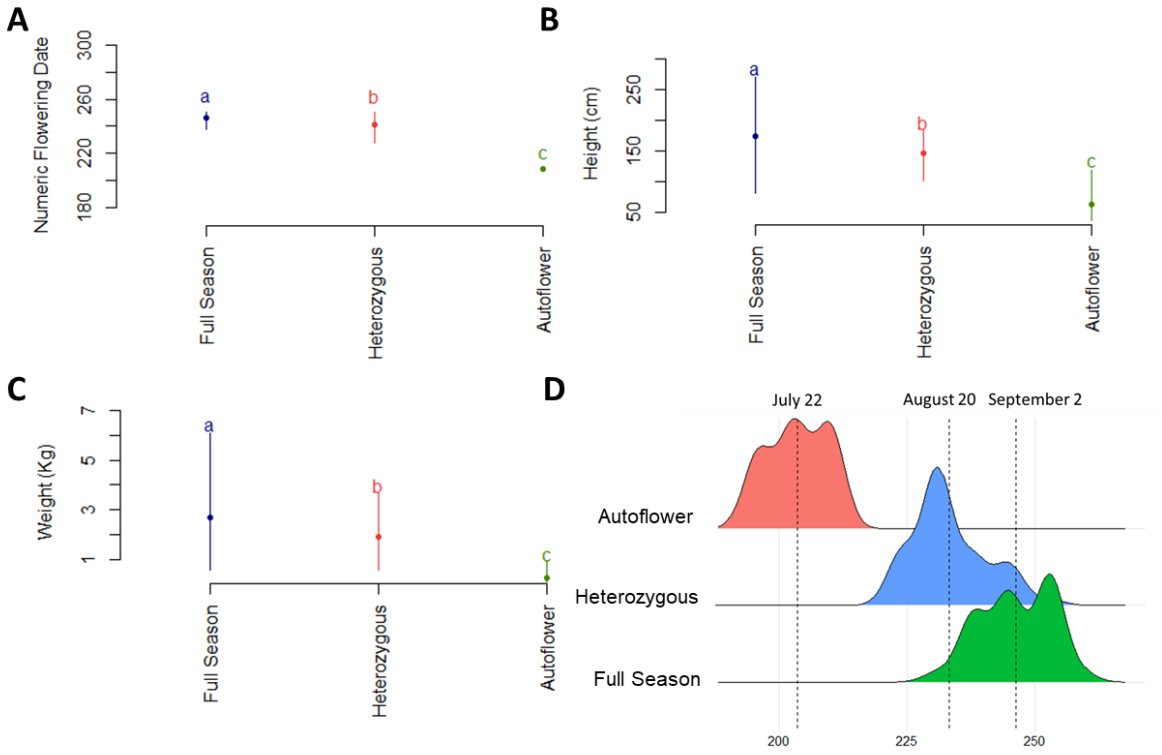

Figure 4.2. Effect of genotype at *Autoflower1* on agronomic traits. A) Numeric (ordinal) flowering day. B) Height, measured from base to tip at end of season. C) Wet biomass. D) Density ridge plot of flowering times within groups. Letters are groups determined through the Tukey post-hoc test.

Flowering of diverse germplasm under continuous light

The behavior of diverse germplasm in continuous light is poorly understood. Several populations were grown under continuous light to determine if non-*Autoflower1* cultivars could be induced to flower. The results, summarized in Table 4.4, show distinct behavior between and within populations. Most cultivars grown primarily for CBD did not flower under continuous light, except for the cultivars homozygous for *Autoflower1*. Notably, plants heterozygous at the *Autoflower1* locus did not flower under continuous light. Fiber and Chinese cultivars tended not to flower, although male plants in closely related feral populations did flower. Some grain cultivars from Canada ('Picolo' and 'CFX-1', Hemp Genetics International, Saskatoon, SK) flowered under continuous light.

Table 4.4. Time to flower under continuous light. * Denotes planted May 31 2021. Taxonomical group data described in Carlson et al., 2021.

	Males	Taxonomic Group	Source	Flowers May 10	Flowers May 17	Flowers May 24	Flowers May 31	Flowers June 28
'Carmagnola'	Yes	Fiber/Feral	Schiavi Seed	No	No	No	No	No
GVA-H-20-1179-156	Yes	Chinese	Cornell Hemp	No	No	No	No	No
'PuMa'	Yes	Chinese	CN Kenaf and Hemp Seed Farm	No	No	No	No	No
'Han-Cold'	Yes	Chinese	CN Kenaf and Hemp Seed Farm	No	No	No	No	No
GVA-H-19-1052	No	West Coast	Cornell Hemp	No	No	No	No	No
'RN16'	Yes	T1/R4	Paul Smith Denver Co.					
'Umpqua'	No	West Coast	Arcadia Bioscience	No	No	No	No	No
'NS52'	No	Not tested	Phytonyx	No	No	No	No	No
'Fedora 17'	No	Grain/Dual	UniSeeds, Inc.	No	No	No	No	Axial
'A2R4'	Yes	Fiber/Feral	WinterFox Farms	No	No	No	No	Axial
'Baox'	Yes	BaOx/Otto II	Ryes Creek	No	No	No	No	Males, axial females
'Nebraska'	Yes	Fiber/Feral	WinterFox Farms	No	No	No	No	Males, axial females
Missouri Feral	Yes	Fiber/Feral	John Fike, (Feral, 40.228, -94.56)	No	No	No	Males	Males, axial females
Wisconsin Feral	Yes	Fiber/Feral	(Feral, 42.67, -88.934)	No	No	Males	Males	Males, axial females
'Victoria'	Yes	Fiber/Feral	Hiliard	No	No	No	Males	Males, axial females
GVA-H-20-1080	No	Intercross	Cornell Hemp	No	1/4	1/4	1/4	1/4

‘Auto CBD’	No	Not tested	Phylos Bioscience	No	Yes	Yes	Yes	Yes
‘Auto CBG’	No	Not tested	Oregon CBD	No	Yes	Tes	Yes	Yes
‘Socati Auto’	No	Not tested	Boring Hemp Co.	No	Yes	Yes	Yes	Yes
KG9202	No	West Coast	Kayagene	No	Yes	Yes	Yes	Yes
‘Dr. Chunk’	No	Not tested	Kayagene	No	Yes	Yes	Yes	Yes
‘Anka’	Yes	Grain/Dual	UniSeeds Inc.	No	Males	Males	Males, axial females	Males, axial females
‘Henola’	No	Grain/Dual	Bija Hemp	No	Some	Yes	Yes	Yes
‘CFX-1’	Yes	Grain/Dual	Hemp Genetics International	Males	Yes	Yes	Yes	Yes
‘Picolo’	Yes	Grain/Dual	Hemp Genetics International	Males	Yes	Yes	Yes	Yes
‘Finola’ (Oregon)	Yes	Not tested	Calderone	Some males (6/50 plants)	Some males (6/50 plants)	Some males (6/50 plants)	Some males (10/50)	20 Flowering (10 males, 5 females)
Lithuanian	Yes	Grain/Dual	Endoca	*	*	*	*	Yes
‘X-59’	Yes	Grain/Dual	Legacy Hemp	*	*	*	*	Yes

Complementation test of photoperiod insensitive cultivars

As ‘Picolo’ and individuals within the ‘La Crème’ population were both induced to flower under continuous light but appeared to differ at the *Autoflower1* locus, a complementation test was performed to determine if there were distinct genes underlying their respective photoperiod insensitivity. All F_1 plants from this cross were induced to flower, although timing was consistently different compared to the parents (Table 4.5). Notably, female plants known to be homozygous for the *Autoflower1* locus flowered 3 weeks earlier than female ‘Picolo’ and female F_1 plants, and male ‘Picolo’ plants flowered two weeks earlier than the male F_1 plants. Female ‘Picolo’ and F_1 plants were morphologically similar, while ‘La Crème’ *Autoflower1* plants were distinct (Figure 4.3).

Table 4.5. Time to flower under long day (16L:8D) lighting.

Cultivar or Pedigree	Male Flowers	Female Flowers
‘AutoCBD’, ‘La Crème’ *Autoflower1*	NA	4 weeks
‘Picolo’	4 weeks	7 weeks
‘La Crème’ *Autoflower1* × ‘Picolo’ (F_1)	6 weeks	7 weeks
‘RN16’	None	None

Figure 4.3. Photoperiod insensitive hemp grown under long days (16L:8D) and photographed 85 days after planting. A) Representative 'Picolo' females B) Representative 'La Crème' *Autoflower1* plants C) Representative plants of 'La Crème' *Autoflower1* × 'Picolo' F_1 females. Receptive white pistils were present at the apex of plants in A and C, but only dried brown pistils were present on plants in B.

Segregation for flowering time in 'Umpqua'

The cultivar 'Umpqua' has been grown in Cornell field trials in 2019, 2020, and 2021. In each year, two distinct flowering times were noted (Figure 4.5). Over the course of three years, clear grouping was apparent, with 78 plants total in the early flowering group and 97 plants total in the later flowering group. This data is consistent with a 1:1 segregation of early and late phenotypes (Chi-square p=0.31), characteristic of a backcross involving a major effect gene (here designated *Early1*) that is heterozygous in one parent and homozygous in the other. Neither phenotype was induced to flower under continuous light (Table 4.4).

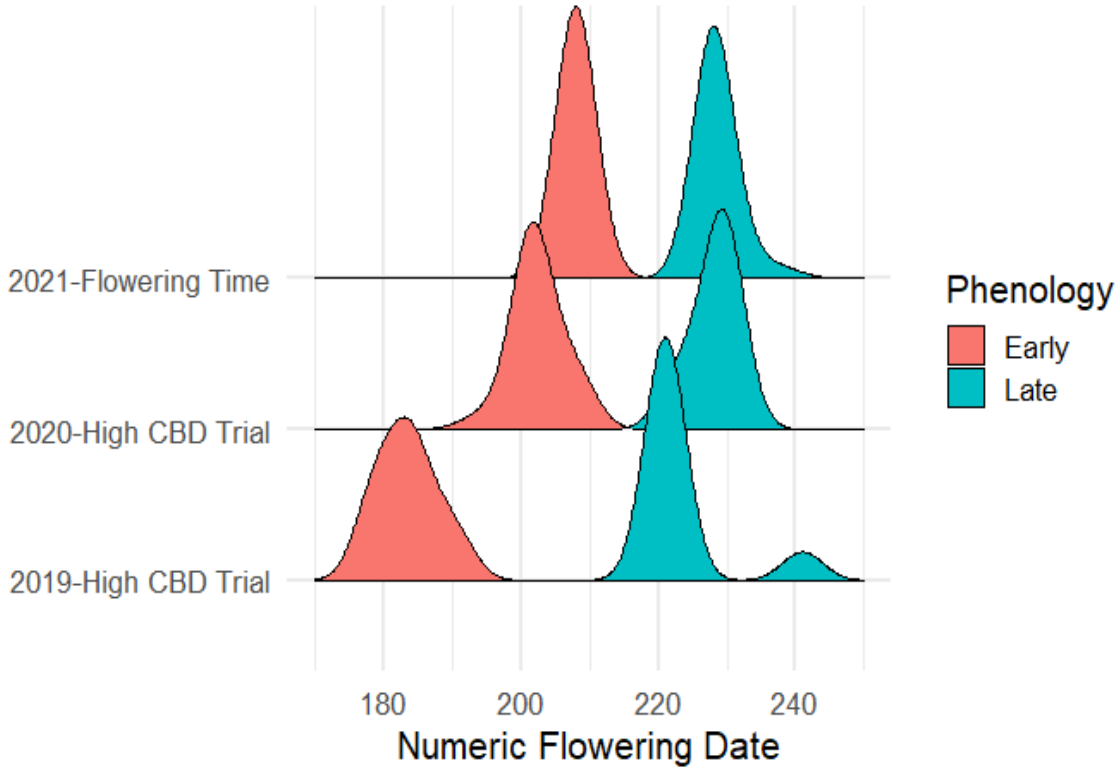

Figure 4.4. Density ridge plot of 'Umpqua' flowering time over three field trials in Geneva, NY.

Mapping of *Early1* in 'Umpqua'

Bulk segregant analysis showed clear statistical significance for the *Early1* locus on Chromosome 1 (NC_044371.1), reaching significance using the delta-SNP frequency approach (Figure 4.5). Visual examination of the significant SNP data showed that the early flowering 'Umpqua' group was heterozygous at *Early1* while late flowering 'Umpqua' group was not.

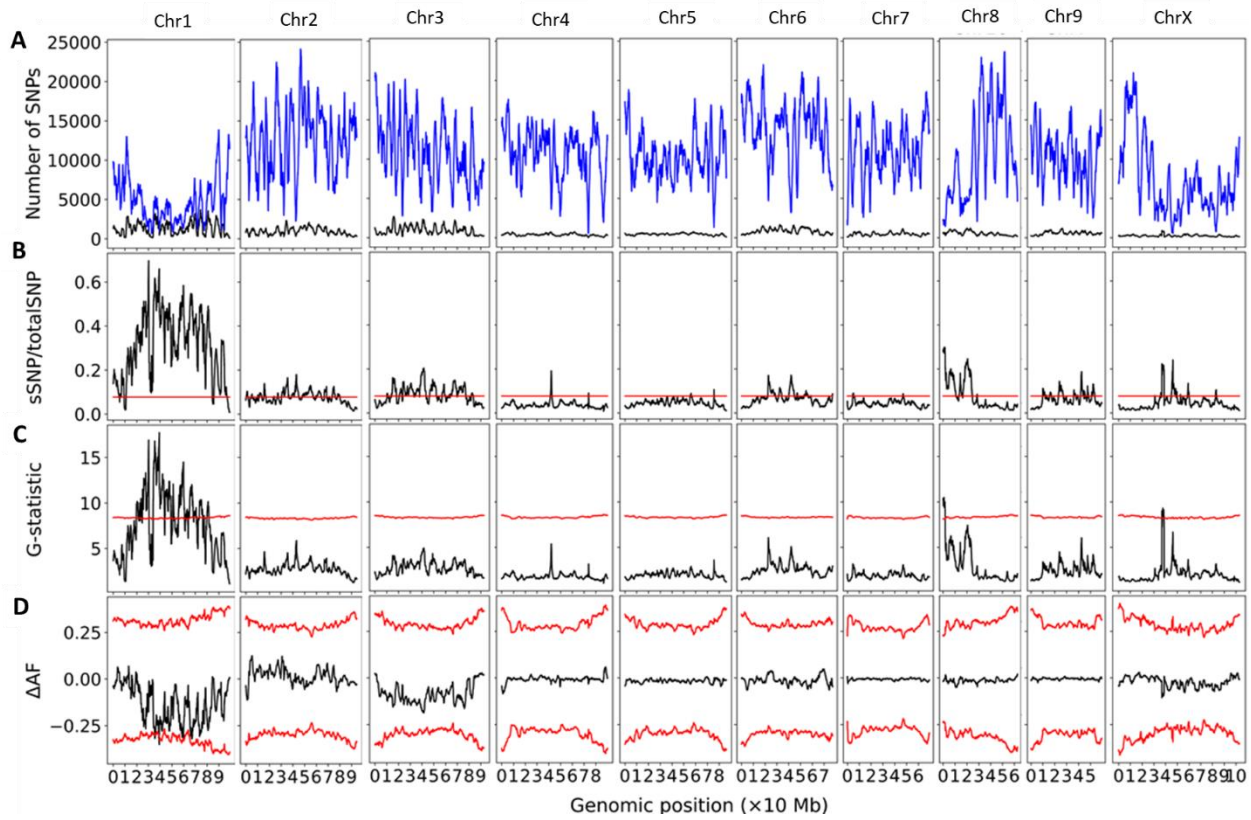

Figure 4.5. Bulk segregant analysis of Chromosome 1 for pools of early- and late-flowering plants from 'Umpqua'. Window size is 2 Mb and step size is 10,000 bp. A) Number of SNP called per region. B) Ratio of significant SNP:Total SNP. C) G-statistic. D) Delta-allele (Delta-SNP) frequency value. Red lines represent significance thresholds.

Early1 candidate gene analysis

For the two 'Umpqua' pools, there were several small peaks that reached significance by the delta-SNP metric on Chromosome 1, located at 35,260,001-36,230,001; 38,670,001-39,360,001; and 59,800,001-59,900,001. Within these regions there are 45 genes, of which the strongest candidate gene for *Early1* based on molecular function is LOC115705415 (annotated to encode Casein kinase 1-like protein 1), located at 39,265,477-39,269,512. Polymorphic SNP that were heterozygous for *Early1* in the early flowering pool were developed into high-throughput PACE assays (named EARLY, Table 4.1) and reported genotypes were found to correlate perfectly with the early- and late-flowering phenotypes of 'Umpqua' across all tested plants (N=175). The 45 genes within the significant region are detailed in Table 4.6.

Table 4.6. Annotated genes within the delta-SNP significant region on Chromosome 1 for *Early1* defined by bulk segregant analysis of 'Umpqua'.

Chromosome	Start	Stop	Gene symbol	Gene Name
NC_044371.1	38708527	38713025	LOC115705813	uncharacterized LOC115705813
NC_044371.1	38724607	38726891	LOC115707485	syntaxin-22
NC_044371.1	38729916	38733199	LOC115706081	transcription factor IIIA
NC_044371.1	38735468	38738063	LOC115705986	histone acetyltransferase MCC1
NC_044371.1	38740394	38741329	LOC115704029	protein EXORDIUM-like 6
NC_044371.1	38742998	38744151	LOC115706076	THO complex subunit 7B-like
NC_044371.1	38745459	38746158	LOC115706077	S-norcoclaurine synthase 2
NC_044371.1	38745617	38748388	LOC115706078	uncharacterized LOC115706078
NC_044371.1	38755581	38756697	LOC115705063	S-norcoclaurine synthase 2
NC_044371.1	38781233	38781825	LOC115704031	S-norcoclaurine synthase 2-like
NC_044371.1	38785491	38786161	LOC115704032	S-norcoclaurine synthase 2-like
NC_044371.1	38794465	38796045	LOC115703764	uncharacterized LOC115703764
NC_044371.1	38801599	38802357	LOC115705164	uncharacterized LOC115705164
NC_044371.1	38803330	38804310	LOC115707475	S-norcoclaurine synthase 2
NC_044371.1	38804614	38806086	LOC115706447	S-norcoclaurine synthase 2
NC_044371.1	38936030	38936745	LOC115704033	uncharacterized LOC115704033
NC_044371.1	39032675	39034746	LOC115704231	protein DETOXIFICATION 51
NC_044371.1	39059472	39060592	LOC115707502	uncharacterized LOC115707502

NC_044371.1	39238848	39240715	LOC115705112	uncharacterized LOC115705112
NC_044371.1	39265477	39269512	LOC115705415	casein kinase 1-like protein 1
NC_044371.1	39272873	39273746	LOC115704841	uncharacterized LOC115704841
NC_044371.1	39274362	39279949	LOC115704840	diacylglycerol O-acyltransferase 1B
NC_044371.1	39292464	39316101	LOC115705447	eukaryotic initiation factor 4A-10
NC_044371.1	39309335	39315575	LOC115705449	THO complex subunit 6-like
NC_044371.1	39353098	39353685	LOC115705450	uncharacterized LOC115705450
NC_044371.1	39353865	39355789	LOC115704035	THO complex subunit 6-like
NC_044371.1	35279033	35283655	LOC115706124	beta-galactosidase 1-like
NC_044371.1	35284120	35287188	LOC115706126	uncharacterized LOC115706126
NC_044371.1	35287250	35292827	LOC115706125	beta-galactosidase
NC_044371.1	35289677	35290530	LOC115706137	uncharacterized LOC115706137
NC_044371.1	35421311	35422996	LOC115706129	GATA transcription factor 21-like
NC_044371.1	35944099	35946237	LOC115706127	myb family transcription factor APL-like
NC_044371.1	35950263	35952789	LOC115706135	uncharacterized LOC115706135
NC_044371.1	35977958	35978491	LOC115704001	uncharacterized LOC115704001
NC_044371.1	36069682	36070734	LOC115705480	wound-induced protein 1
NC_044371.1	36147501	36150228	LOC115703563	uncharacterized LOC115703563
NC_044371.1	36151362	36155217	LOC115705850	alpha-soluble NSF attachment protein 2
NC_044371.1	36160041	36163153	LOC115708134	amino acid transporter AVT6C
NC_044371.1	36196345	36198771	LOC115707593	amino acid transporter AVT6C
NC_044371.1	59843988	59844515	LOC115704180	uncharacterized LOC115704180
NC_044371.1	59864864	59867947	LOC115706631	eukaryotic translation initiation factor 2 subunit alpha homolog
NC_044371.1	59868903	59871189	LOC115704181	probable methyltransferase PMT23
NC_044371.1	59871724	59872621	LOC115704182	26S proteasome non-ATPase regulatory subunit 10-like
NC_044371.1	59896681	59899567	LOC115704768	auxin-responsive protein IAA27

4.6 Discussion

Autoflower1

Autoflower1 behaves as a simple, recessive, Mendelian locus with respect to photoperiod insensitivity. Here, I mapped the *Autoflower1* locus derived from KG9202 controlling the photoperiod insensitive phenotype in the GVA-H-20-1080 population to a relatively small region on Chromosome 1 using bulk segregant analysis. I used SNP polymorphisms from the Illumina data to develop *Autoflower1* molecular assays that correctly predicted the photoperiod sensitivity phenotype of individuals based on genotype from 10 sources, with no false positives across diverse germplasm.

While *Autoflower1* is recessive with respect to photoperiod insensitivity, under field conditions plants that were heterozygous for *Autoflower1* flowered about two weeks earlier than plants that were homozygous for *Autoflower1*, and this earlier flowering resulted in smaller plants with less total biomass. This earlier flowering may be useful for higher latitudes, and as previously shown (Stack et al., 2021), cultivars that are heterozygous for *Autoflower1* can produce very high yields. Many available cultivars are heterozygous for *Autoflower1*, which may present an effective breeding strategy for IP protection, although the prevalence of segregating populations suggests that some (perhaps unscrupulous) breeders used parents that were heterozygous at *Autoflower1* leading to ¼ photoperiod insensitive plants in the seed population and a potential major loss for growers. As detailed in Table 4.4, multiple populations from multiple sources were segregating for *Autoflower1*.

Further work to identify the taxonomic source of *Autoflower1* is also pertinent. *Autoflower1* is often ascribed in the grey literature as derived from *C. ruderalis*, but the

most recent and in-depth genomic studies do not support the existence of this group (Carlson et al., 2021; Green, 2005; Ren et al., 2021). *Autoflower1* would be expected to have evolved either at very high or very low latitudes, where daylength variation would not be a reliable method of determining growing season and photoperiod insensitivity could be advantageous. Further plant collecting expeditions and genomic analysis may help resolve this in the future.

Future work to determine the causative gene at *Autoflower1* will allow biotechnological manipulation of the photoperiod sensitivity phenotype and more facile conversion of elite cultivars to and from photoperiod insensitivity. There were a number of strong candidate genes for *Autoflower1* based on annotated predicted molecular function in the significant QTL interval identified for the GVA-H-20-1080 pools. Notably, SNP near the gene for nuclear transcription factor Y subunit B-1 (*NFYB1*, LOC115706176) and for floral homeotic gene *APETALA 2* (*AP2*, LOC115708151) were in linkage disequilibrium and were perfectly associated with predicted trait phenotype across all accessions tested. These genes have the potential to be causative for the trait, as a *Nuclear factor Y* gene (*DTH8*) plays an important repressive role related to photoperiod in rice (Wei et al., 2010) while *AP2* homologs are also important flowering time repressors in pepper (Yuan et al., 2021) and Arabidopsis (Yant et al., 2010). Future gene silencing or knockouts of these and other potential candidate genes may lead to identification of the true gene or set of genes responsible for this trait, although a patent is already held covering biotechnological manipulation of genes within this genetic interval (Phylos Bioscience, International Patent WO 2021/097496 A2).

Continuous light

Diverse germplasm responded differently to continuous light. Some cultivars, notably high-cannabinoid cultivars, display quantitative daylength sensitivity and will not flower under non-inductive conditions. Others, such as the Canadian grain cultivars tested here, flowered under continuous light, but may have a different genetic mechanism determining photoperiod insensitivity than the *Autoflower1* high-cannabinoid cultivars. Fiber cultivars in the European and Chinese clades did not flower under continuous light. Fiber cultivars have been selected for their ability to continue to grow vegetatively until late in the season, which maximizes stem biomass yield. Interestingly, some feral populations, which are closely related to European fiber cultivars (Carlson et al., 2021), displayed male flowering, but not female flowering, perhaps indicating some selective advantage to early male flowering in the natural environment. This may also reflect the ancestral genetics of the progenitors of these feral populations, but it is difficult to know the original provenance of those progenitors.

Interestingly, despite not being reported by the *Autoflower1* markers, Canadian grain cultivars 'Picolo' and 'CFX-1' flowered readily under continuous light conditions. This could be due to the molecular markers not being polymorphic or effective in these populations, or due to a different genetic basis for photoperiod insensitivity. Different genetic bases may be resolved with a complementation test. If the same gene was responsible for photoperiod sensitivity in 'Picolo' and *Autoflower1* 'La Crème', F_1 progeny from an intercross should be uniformly photoperiod insensitive. Otherwise, other genes, dominance, or epistatis may be involved. The results (Table 4.5) were somewhat inconclusive, as all plants flowered under long days, but the timing and architecture of

flowering (Figure 4.4) suggests more complex genetic regulation in photoperiod insensitive plants across broad germplasm.

Early1

Beyond segregation for *Autoflower1*, several elite populations marketed as cultivars have been demonstrated as segregating 1:1 for a major effect early flowering time phenotype (Stack et al., 2021). The *Early1* locus, which confers an apparent effect size of 2-4 weeks earlier flowering in 'Umpqua', was also mapped to Chromosome 1 using BSA, but to a different location than *Autoflower1*. Molecular markers for *Early1* were identified and high-throughput assays developed for this locus, which could further aid in development of cultivars with uniform flowering time.

In searching for candidate genes in the confidence interval for the *Early1* locus identified in 'Umpqua' populations, only a small portion of Chromosome 1 was found to be significant by the delta-SNP method. Within this small significant peak was one possible candidate gene for early flowering based on annotation, predicted to encode a Casein kinase 1-like protein 1 (LOC115705415). This gene is homologous to the major flowering time gene *Early flowering 1*/*Heading date 16* in rice, another short day plant (Hori et al., 2013). Future validation work could involve genetic engineering or genome editing to accomplish gene knockout or gene knock-in to confirm loss or gain of function. The molecular markers and assays for *Early1* presented in this work will be helpful in future breeding. Studies to further explore the interactions between these two flowering time loci, *Autoflower1* and *Early1*, will likely lead to a better understanding of the genetics of flowering time and development of stable cultivars with unique flowering times. As early-flowering 'Umpqua' plants were heterozygous for both traits, the progeny of an inbred

population would be expected to form nine genotypic groups, whose phenotypes would reveal the role of epistasis between these loci. I have conducted this cross to address this question, and intend to grow the population in the summer of 2022.

There were some differences in the statistical outcomes of the BSA for *Autoflower1* in comparison to *Early1* in 'Umpqua'. The statistically significant region of *Early1* in 'Umpqua' as determined by the G statistic was much larger and broader than that of *Autoflower1*. This is not surprising if this segregation truly is the result of a simple backcross, as recombination occurs only in one parent, rather than in both parents. This reduces the number of crossover events and therefore increases the apparent QTL size. However, analysis using the delta-SNP method resulted in a small peak and reliable diagnostic molecular assays were developed for the *Early1* locus.

Many hemp cultivars produced during the rapid expansion of the CBD industry in the US were segregating for flowering time or photoperiod insensitivity. There is a critical need in the industry to develop uniform and stable cultivars that represent a range of critical photoperiod lengths that can be matched with the latitude of agricultural regions. A better understanding of the genetics of flowering time, coupled with molecular tools to accelerate breeding and selection, will enable the development of new uniform cultivars to meet this need.

4.7 Acknowledgements

Other individuals who made substantial intellectual contributions to this work include George Stack, Craig H. Carlson, and Lawrence B. Smart. I am grateful to the technical staff of the Smart lab, especially Rebecca Wilk, Teagan Zingg, Allison DeSario, Deanna Gentner, Lauren Carlson, Brian Nardone, Michael Quade, Alexander Wares, and McKenzie Schessl. This work was partially supported by New York State Department of Agriculture and Markets through grants AC477 and AC483 from Empire State Development Corporation, as well as the Scotts Corporation through a grant from the Foundation for Food and Agriculture Research Hemp Research Consortium, and a sponsored research agreement with Pyxus International.

4.8 References

Ban, S., & Xu, K. (2020). Identification of two QTLs associated with high fruit acidity in apple using pooled genome sequencing analysis. *Horticulture Research, 7*.

Becker, A., Chao, D.-Y., Zhang, X., Salt, D. E., & Baxter, I. (2011). Bulk segregant analysis using single nucleotide polymorphism microarrays. *PLoS One, 6*(1), e15993.

Callaway, J. (2002). Hemp as food at high latitudes. *Journal of Industrial Hemp, 7*(1), 105-117.

Carlson, C. H., Stack, G. M., Jiang, Y., Taşkıran, B., Cala, A. R., Toth, J. A., Philippe, G., Rose, J.K., Smart, C.D., & Smart, L. B. (2021). Morphometric relationships and their contribution to biomass and cannabinoid yield in hybrids of hemp (*Cannabis sativa*). *Journal of Experimental Botany, 72*(22), 7694-7709.

Gloss, D. (2015). An overview of products and bias in research. *Neurotherapeutics, 12*(4), 731-734.

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

Green, G. (2005). *The Cannabis Breeder's Bible.* San Fransisco, USA: Green Candy Press.

Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., & Yano, M. (2013). H d16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *The Plant Journal, 76*(1), 36-46.

Jung, C., & Müller, A. E. (2009). Flowering time control and applications in plant breeding. *Trends in Plant Science, 14*(10), 563-573.

Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Computational Biology, 7*(11), e1002255.

McPartland, J. M. (2018). *Cannabis* systematics at the levels of family, genus, and species. *Cannabis and Cannabinoid Research, 3*(1), 203-212.

Nordström, K. J., Albani, M. C., James, G. V., Gutjahr, C., Hartwig, B., Turck, F., Paszkowski, U., Coupland, G., & Schneeberger, K. (2013). Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nature Biotechnology, 31*(4), 325-330.

Petit, J., Salentijn, E. M., Paulo, M.-J., Denneboom, C., & Trindade, L. M. (2020). Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa* L.): A genome-wide association study. *Frontiers in Plant Science, 11*, 1704.

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.

Segawa, T., Nishiyama, C., Tamiru-Oli, M., Sugihara, Y., Abe, A., Sone, H., Itoh, N., Asukai, M., Uemura, A., Oikawa, K., & Utsushi, H. (2021). Sat-BSA: an NGS-based method using local de novo assembly of long reads for rapid identification of

genomic structural variations associated with agronomic traits. *Breeding Science*, 20148.

Song, J., Li, Z., Liu, Z., Guo, Y., & Qiu, L.-J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Frontiers in Plant Science, 8*, 919.

Stack, G. M., Toth, J. A., Carlson, C. H., Cala, A. R., Marrero-González, M. I., Wilk, R. L., Gentner, D. R., Crawford, J.L, Philippe, G,. Rose, J. K., Viands, D.R., Smart, C.D., & Smart, L.B. (2021). Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy, 13*(4), 546-561.

Toth, J., Pandurangan, S., Burt, A., Mitchell Fetch, J., & Kumar, S. (2018). Marker-assisted breeding of hexaploid spring wheat in the Canadian prairies. *Canadian Journal of Plant Science, 99*(2), 111-127.

Toth, J. A., Smart, L. B., Smart, C. D., Stack, G. M., Carlson, C. H., Philippe, G., & Rose, J. K. (2021). Limited effect of environmental stress on cannabinoid profiles in high-cannabidiol hemp (*Cannabis sativa* L.). *GCB Bioenergy, 13*(10), 1666-1674.

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K. & Smart, L.B. (2020). Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy, 12*(3), 213-222.

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.

Wei, X., Xu, J., Guo, H., Jiang, L., Chen, S., Yu, C., Zhou, Z., Hu, P., Zhai, H., & Wan, J. (2010). DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology, 153*(4), 1747-1758.

Welling, M. T., Liu, L., Kretzschmar, T., Mauleon, R., Ansari, O., & King, G. J. (2020). An extreme-phenotype genome-wide association study identifies candidate cannabinoid pathway genes in *Cannabis*. *Scientific Reports, 10*(1), 1-14.

Yang, R., Berthold, E., McCurdy, C. R., da Silva Benevenute, S., Brym, Z. T., & Freeman, J. H. (2020). Development of Cannabinoids in Flowers of Industrial Hemp (*Cannabis sativa* L.)—a Pilot Study. *Journal of Agricultural and Food Chemistry*.

Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., Chen, X., & Schmid, M. (2010). Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor APETALA2. *The Plant Cell, 22*(7), 2156-2170.

Yuan, X., Fang, R., Zhou, K., Huang, Y., Lei, G., Wang, X., & Chen, X. (2021). The APETALA2 homolog CaFFN regulates flowering time in pepper. *Horticulture Research, 8*.

Zhang, G.-L., Chen, L.-Y., Xiao, G.-Y., Xiao, Y.-H., Chen, X.-B., & Zhang, S.-T. (2009). Bulked segregant analysis to detect QTL related to heat tolerance in rice (*Oryza sativa* L.) using SSR markers. *Agricultural Sciences in China, 8*(4), 482-487.

Zhang, J., & Panthee, D. R. (2020). PyBSASeq: a simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data. *BMC Bioinformatics, 21*(1), 1-10.

Zhang, M., Anderson, S. L., Brym, Z. T., & Pearson, B. J. (2021). Photoperiodic Flowering Response of Essential Oil, Grain, and Fiber Hemp (*Cannabis sativa* L.) Cultivars. *Frontiers in Plant Science*, 1498.

Chapter 5: Conclusions and future prospects

5.1 Chapter conclusions

The results of testing the hypotheses presented in Chapter 1 and relevant consequences are detailed below, separated by chapter. Important tools and germplasm derived from this work include the following:

1) High-throughput molecular markers for sex, cannabinoid chemotype, and major-effect flowering time loci.
2) GVA-H-21-1135 and GVA-H-22-1061, high-yielding THC-compliant dual purpose and fiber cultivars.
3) GVA-H-19-1039, a breeding line with a novel mechanism of CBC accumulation.

Chapter 2

Sex (male vs female) and cannabinoid chemotype are qualitative traits in hemp with simple genetic architecture. There is segregation for active THCAS within CBD, grain, and fiber cultivars. There is sex chromosome segregation distortion in seeds derived from XY plants, and the existence of supermales seems unlikely.

Chapter 3

f

The CBD:THC ratio is fixed in high-cannabinoid chemotype III plants even under biotic and abiotic stress, meaning the more productive a CBD plant, the more likely it is to test above 0.3% THC. Future exploration or biotechnological innovation will likely be required to improve this ratio.

Chapter 4

There are several major effect flowering time loci in *C. sativa*, including loci I am naming *Autoflower1* derived from KG9202, and *Early1* derived from 'Umpqua'. Both these loci are located on chromosome 1 (Grassa et al., 2021). Plants heterozygous for *Autoflower1* are earlier under field conditions than wild-type plants. There are likely additional loci that lead to photoperiod-insensitive flowering beyond *Autoflower1*.

There are several questions left unanswered by this work. Detailed below are extensions of the information presented in the chapters and brief experimental plans for addressing them.

5.2 Future prospects

Extensions of Chapter 2

1) The existence of supermale plants, while not demonstrated to be viable in the tested population, might be possible in other cultivars or with a larger population size. The low vigor of the ethephon-treated male plants and subsequent low seed yield makes the generation of a large population difficult, but this could be remedied by using a more vigorous cultivar. The hypothesis that haploid female germ cells without an X chromosome are less viable or inviable could also be tested by crossing XY plants treated with ethephon with XX-derived pollen, either from monoecious individuals or plants treated with STS. If Y haploid female germ cells are not viable, all the offspring of this XY by XX cross would be expected to be XX.
2) There are five accepted cannabinoid chemotypes (de Meijer & Hammond 2016), with cannabinoid chemotypes IV and V referring to plants producing dominantly

CBG and cannabinoid-free plants, respectively. Some preliminary evidence (not shown) suggests that there are different genetic bases among some chemotype IV cultivars, but unique molecular markers are not available for all of these types. Further work sequencing representative individuals and examining complementation and heterozygote effects will improve knowledge in this area.

3) The new cultivar detailed in Chapter 2 developed through marker assisted selection for cannabinoid chemotype and phenotypic selection for flowering time has significant potential for grain and possibly fiber production. Trials of this cultivar at different latitudes and growing conditions will inform optimal locations and methods for production.

Extensions of Chapter 3

1) As demonstrated in Chapter 3, ethephon, flooding, herbicide, powdery mildew, and wounding failed to alter the CBD:THC ratio of high-cannabinoid chemotype III plants of three cultivars at harvest. While these stresses are varied and representative of some issues facing New York growers, it is possible that other stresses such as heat, salinity, or drought could affect this ratio, perhaps through altering trichome pH. Heat and salinity may be better examined under greenhouse conditions, while drought could be tested in a field trial in an arid environment or in well-constructed rain-out shelters. There is also some evidence of some minor genotype-dependent variation in CBD:THC ratio (Stack et al., 2021), and these unique genotypes may behave under stress differently.
2) The work described in chapter 3 is focused solely on the content of cannabinoids in a regulatory-style shoot tip sample. While this is important information, it may be

more important to the producer of the effect of stress on whole plant biomass, or total cannabinoid yield from stressed plants. This could be addressed through more rigorous end-of season phenotyping of a similar trial.

3) Only chemotype III plants were used in this trial. This is useful information, as chemotype III plants produce predominantly CBD, a legal compound with recreational and medicinal merit. However, it would be very interesting to see the effect of stress on chemotype II plants, as they have active CBDAS and active THCAS, which may be under different transcriptional control. If stress altered this transcriptional control, it would have the potential to alter the CBD:THC ratio in chemotype II plants. While this study could follow a similar outline as described in Chapter 3, the plants would certainly produce >0.3% THC and could not be grown in a university setting without a Schedule 1 Drug Enforcement Agency registration.
4) It is very likely that the source of THC in chemotype III plants is due to product promiscuity of CBDAS, as discussed in chapter 2. Environmental stress does not appear to affect this ratio. However, it has been noted that various mutations can affect this ratio in an *in vitro* system (Zirpel et al., 2018). A more thorough analysis of mutations in this enzyme class, perhaps through a staggered extension PCR scheme (Aguinaldo and Arnold 2002) utilizing genes encoding THCAS and CBDAS, could lead to a better understanding of product promiscuity and engineering of more- or less-promiscuous enzymes.

Extensions of Chapter 4

1) Completely linked molecular markers and candidate genes at the *Autoflower1* and *Early1* flowering time loci are presented in Chapter 4. However, further work,

potentially involving transcriptome analysis, transgenics, and gene knockouts could lead to a mechanistic understanding of the cause of early flowering, rather than just an associative understanding.

2) The *Autoflower1* and *Early1* loci were found to be segregating in elite germplasm. There is also some evidence (not shown) that they interact epistatically, but a full population to study epistasis between these loci is lacking. Double heterozygotes already exist in the 'Umpqua' population, so an intercrossed population should show nine distinct genotypes, separable by molecular markers, which could be examined for phenotype. However, since these loci are located fairly close together (~20 MB), a large population to overcome linkage will be necessary.
3) The complementation test detailed in Chapter 4 between two potentially distinct sources of daylength insensitivity was somewhat inconclusive, with putative double heterozygotes flowering under long days, but with a unique flowering time. This could be due to complex epistasis or a dominant form of daylength insensitivity in 'Picolo'. Further work to determine the genetic basis for daylength insensitivity in 'Picolo' (which may not be expressed under all environments) will aid in understanding photoperiod insensitivity in *C. sativa*.
4) Chapter 4 describes mapping of the *Autoflower1* and *Early1* loci involved in early flowering. These are useful for breeding fast-maturing cultivars for high latitude locales and will give better control over uniformity of flowering under field conditions. However, there is great potential for growing hemp at subtropical and tropical latitudes, which will involve breeding late-maturing cultivars. Some late maturing cultivars have been described, such as 'Late Sue' in Chapter 4 and GVA-

H-21-1135 in Chapter 2, but the genetic basis of this late flowering is not well understood. Creating mapping populations with these cultivars will be useful for future low-latitude breeding.

Other extensions of this work

1) The cultivar GVA-H-21-1135, derived from GVA-H-20-1179 described in Chapter 2, has many unique traits, including high thousand kernel weight, high proportion varin cannabinoid production, and low protein content. Its protein also has a high albumin:edestin ratio, leading to high solubility and better functional properties (Liu et al., submitted 2022). A mapping population to understand the nature of these traits would be useful for future breeding. It could also serve as useful germplasm for incorporating these traits into new cultivars that are not as late maturing as GVA-H-21-1135.
2) The work in Chapter 2 and 3 relies heavily on HPLC analysis, which is relatively slow and expensive. Preliminary work has examined the capability of NIR analysis as an alternative to HPLC, saving time and money for routine testing (Callado et al., 2018). Further development of these NIR models will result in useful tools for cannabinoid sampling.
3) A third mapping population was developed to examine flowering time, but was abandoned due to ¼ of the seedlings dying at approximately 7-10 days after planting. While this reduced the population size to below what would be useful for mapping flowering time, mapping this seedling death phenotype lead to identification of a unique genomic confidence interval. Further study into the cause of this phenotype could lead to better understanding of *C. sativa* physiology,

perfectly linked molecular markers to select against plants heterozygous for this trait in elite populations, and potentially a new form of intellectual property control if heterozygotes can be bred.

The research detailed here is primarily concerned with qualitative traits for which the phenotypes of segregating populations can be easily partitioned into discrete groups. These traits can be the basis of unique market classes in plant breeding, and the methods described here may allow for more facile selection. However, plant breeding is often concerned with quantitative traits, controlled by many loci with small effect. Future work into better understanding these traits may lead to improved plant breeding of hemp. Relevant quantitative traits that might be the subject of future selection include cannabinoid content, quantitative photoperiod threshold for flowering, grain yield, grain quality characteristics such as protein and oil content, fiber yield, and fiber quality.

5.3 References

Aguinaldo, A. M., & Arnold, F. (2002). Staggered extension process (StEP) in vitro recombination. In *PCR Cloning Protocols* (pp. 235-239). Humana Press.

Callado, C. S.-C., N. Núñez-Sánchez, S. Casano and C. Ferreiro-Vera (2018) The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: A comparative study. *Talanta 190*: 147-157.

de Meijer, E. P. M., & K. M. Hammond (2016) The inheritance of chemical phenotype in *Cannabis sativa* L.(V): regulation of the propyl-/pentyl cannabinoid ratio, completion of a genetic model. *Euphytica*, *210*(2): 291-307.

Liu, M., Toth, J.A., Childs, M., Smart, L. B., Abbaspourrad, A. (submitted manuscript, 2022). Composition and functional properties of hemp seed protein isolates from various cultivars. Manuscript submitted for publication.

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

Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *Journal of Biotechnology, 284*, 17-26.