DISSERTATION

PLASTICITY, ALLELIC DIVERSITY, AND GENETIC ARCHITECTURE OF INDUSTRIAL HEMP (*CANNABIS SATIVA L.*)

Submitted by

Brian J. Campbell

Department of Bioagricultural Sciences and Pest Management

In partial fulfillment of the requirements

For the Degree Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2019

Doctoral Committee:

Advisor: John McKay

Courtney Jahn
Stephen Pearce
Mark Brick

ABSTRACT

PLASTICITY, ALLELIC DIVERSITY, AND GENETIC ARCHITECTURE OF INDUSTRIAL HEMP (*CANNABIS SATIVA L.*)

The first time in United States history that hemp was legally distinguished from high-THC *Cannabis* (marijuana) was in 2014 when the Farm Bill was passed. Although the two crops had been distinguished by their usage for thousands of years, their monospecific nature led to both psychoactive and non-psychoactive forms being legislated in tandem from the time that Americans began regulating *Cannabis* cultivation and usage. A simple statement in the 2014 Farm Bill distinguished hemp as *Cannabis sativa L.* with a tetrahydrocannabinol (THC) content of 0.3% or less. A second sentence enabled research into the crop and production within pilot programs in states where it is legal. This minor change in legality, followed by subsequent relaxation of laws surrounding hemp in the 2018 Farm Bill, has allowed a burgeoning hemp industry to form in the United States and enabled the return of a relict crop.

Due to the long period of prohibition, hemp did not undergo the same type of crops research as other staple American crops. Consequently, little is known about the genetic mechanisms that control many of the key traits in hemp production. Understanding basic information about how traits are affected by environmental factors is highly important when regulation of the crop is based on a stringent and arbitrarily set threshold for chemical content. In 2016, we performed field trials of a diverse set of industrial hemp cultivars in multiple growing environments and assessed a wide range of traits. Expression of some traits, like days to maturity and THC content, were strongly influenced by genotype. Other traits, such as grain yield and

plant height, exhibited large proportions of variance due to environmental factors and genotype-by-environment interactions. There were also varying ranges of plasticity exhibited between cultivars, underscoring the importance of selecting the right cultivar for target production environments. This highlights the importance of thoroughly characterizing genotype-by-environment interactions when breeding locally adapted hemp cultivars. Understanding genetic control of important traits and their range of plasticity enables the development of locally adapted cultivars for a wide range of end uses.

Another aspect of *Cannabis* that is understudied is the genetic basis for differentiating hemp and high-THC *Cannabis*. Since the legal distinction is based on a strict threshold placed on a quantitative trait and not any known geographic or biological reproductive barriers, it is unclear whether or not there is genetic evidence to support the distinction or if the two groups are simply divergent phenotypes. A joint-site frequency and F_{ST} analysis show that individuals of the two groups mainly share common polymorphisms, with a small number of loci where differentiation occurs. These loci serve as the basis for distinguishing the two groups, but more study is needed to determine if alleles in these regions were driven to fixation via genetic drift and selection on unrelated traits, or if there is an evolutionary basis for the observed differences. When heterozygosity was assessed in these samples, the hemp group had higher overall heterozygosity levels, but the high-THC *Cannabis* group had more outliers which lead to a wider distribution with more extreme minimum and maximum values. Although it is clear that there are genetic differences distinguishing the two groups, extensive human vectoring and admixture between the groups, both historically and currently, makes it difficult to differentiate causes for the differences. A lack of centralized germplasm makes large-scale genomic studies of the species difficult, but, as more samples are surveyed over time, a more detailed picture of the genomic

variation will emerge. These types of studies will be able to provide a more nuanced picture of the evolutionary history and current state of allelic variation within the species.

In addition to plasticity and allelic diversity, genetic architecture of traits has also largely been ignored until recently. The first QTL study in *Cannabis* was performed in 2015 and was limited by legal restraints. Since understanding how economically relevant traits function is important to breeding improved hemp cultivars, we developed a genetic mapping population that captured variation for a wide range of traits. Utilizing whole-genome sequencing and phenotype data from a replicated field trial, we were able to detect 121 QTL associated with 38 agronomic and biochemical traits. Some traits, like days to maturity, had single loci of large effect accounting for the majority of trait variance, while other traits, like α-Pinene production, exhibited more complex polygenic architecture with epistatic interactions. Colocalization of QTL and significant trait correlations showed that there were positive relationships within both agronomic and biochemical trait groups. Although this study was limited by assessment of the population in a single environment, detecting these putative QTL serves as a substantial step forward in characterizing many relevant production traits.

ACKNOWLEDGEMENTS

Until you have lived through a doctorate program, it is hard to understand the level of determination it takes. It is something that I certainly took for granted and, appropriately, this program has been the single most formative educational experience of my life. I frequently learned lessons where I didn't expect to and, as a major takeaway, found that it is hard to differentiate failure from success until you take time and gain perspective. So much of what I learned along the way is due to interacting with others. There is plenty of hard work you have to apply in solitude, but the real learning comes from exchanging ideas and being open to true shifts in paradigm. First and foremost, I would like to thank my mentor, John McKay. John's fierce independence, encyclopedic knowledge of the literature, and reserved but firm leadership enabled me to feel like I was in charge of my own fate, but never without counsel. You took a chance on me and I hope that I held up my end of the bargain. To my committee members Courtney, Stephen, and Mark, thank you for always challenging me. The expertise that is contained in this group humbles and inspires me regularly. I thank my collaborators and coauthors: CSU's AES and CSU Pueblo's ICR for funding, Dong Zhang, Patrick Woods and Kayla Hardwick for bioinformatics wizardry; Stefano Amaducci for sharing germplasm and hemp expertise; Chris Hudalla and Kevin McKernan for sample analysis; Abdel Berrada for always working in tandem with me; and Jack Mullen for everything from planting crops to running code. Thank you to my awesome assistants and mentees, especially Casey Gildea and Hayley Park. To everyone in the McKay Lab, Anne, Kyle, Grey, and Nicolas for everything from hands-on help to cutting critiques. But mostly, thank you to Eva. You are my reason for existing and I never could have made it to this point without you as my biggest inspiration.

TABLE OF CONTENTS

CHAPTER 1:

LITERATURE REVIEW

INTRODUCTION

Cannabis sativa L. is an economically important crop that has been surrounded by controversy over the last 100 years. Despite its widespread use as an intoxicant and an industrial crop, governments worldwide have struggled to appropriately regulate *Cannabis* use and production. The lack of uniformity in *Cannabis* law, both spatially and temporally, has made research on this plant difficult through traditional channels. With increasing public support for medical marijuana and growing interest in applications for industrial hemp, laws are changing and doors are opening providing for long-overdue research. The purpose of this chapter is to review the current body of knowledge relative to *Cannabis* genetics, specifically regarding speciation/classification, evolutionary origin, genetics of industrial hemp breeding, genetic diversity and population structure of the species, and *Cannabis* genomics, as well as a look at important areas of genetics and genomics research for the future of industrial hemp.

The origin and the taxonomy of *Cannabis* is surrounded by uncertainty and academics debate questions of both where its evolutionary roots lie, as well as whether or not the diversity observed in *Cannabis* warrants distinguishing common types as separate species or subspecies. Eurasia has been proposed as its evolutionary center of origin, with more specific recommendations of central Asia, in the Himalayas, or possibly two distinct centers of Hindustani and European-Siberian origin (Clarke and Merlin, 2016; Hillig, 2005; Andre et al., 2016). As defined in the previous chapter, *Cannabis* has been utilized since ancient times. Archeological findings date the use of hemp rope to 10,000 years ago in Taiwan, China (Laursen, 2015). This

early human use led to vectoring of seed thousands of years ago (Hillig, 2005) that has made clearly distinguishing evolutionary origins by genetic analyses difficult, if not practically impossible.

The diverse morphology of *C. sativa,* combined with a multitude of uses, has caused no shortage of confusion over the classification of this species. The debate of whether or not *Cannabis* is a single species began long ago. The species was first labeled *Cannabis sativa* by Carl Linnaeus in 1753 (Watts, 2006). This monospecific viewpoint was challenged in 1785 when Jean-Baptiste Lamarck found that some *Cannabis* specimens from India exhibited distinctly different morphology from those described by Linnaeus and created a new classification, *Cannabis indica* (Watts, 2006). Sativa samples were tall, with long intermodal spacing, and narrow leaflets, while indica-types are shorter and bushier plants with wide leaflets. Although several other classifications have been proposed, the only generally accepted third possible species was proposed in 1924 by Dmitri Janischevsky as *Cannabis ruderalis* (Hillig and Mahlberg, 2004). This newest classification of *Cannabis* was created to describe Russian samples that did not exhibit the same "domestication syndrome" traits as indica and sativa samples and were essentially small, wild (ruderal; or generally occurring in regularly disturbed ecosystems) plants (Small and Cronquist, 1976). Since there are no reproductive barriers between these three types of *Cannabis*, we consider this as a single species, but the debate between the "splitters" and the "lumpers" remains very active (Watts, 2006).

Chromosome number and ploidy level

Although the taxonomy of *Cannabis* is debated, it is agreed that plants in the species are diploid and have 9 pairs of autosomes, as well as a pair of sex chromosomes that are cytogenetically heteromorphic in male plants (XY system), but homomorphic in female and

monoecious plants (XX system) (Divashuk et al., 2014; Faux et al., 2016; Razumova et al., 2016). However, the roles of Y chromosomes and X-to-autosome ratio in the sex determination system are still open questions in *Cannabis* (Sakamoto et al., 1998; Ming et al., 2011; Faux et al., 2014; reviewed by Vergara et al., 2016). Using DAPI/C-banding staining and FISH, Divashuk et al. (2014) demonstrated the karyotype in dioecious hemp. The cytogenetic study showed that all chromosomes appear to be submetacentric and metacentric (meaning the centromere is located below or near the center of the chromosome, respectively. The location of the centromere assists in definitions of karyotypes and provides for appropriate descriptions of the locations of genes on chromosomes) and the Y chromosome is larger than the X chromosome (Divashuk et al., 2014). In contrast, the Y chromosome was reported to be shorter than X chromosome in male samples of the closely related species; hops, *Humulus lupulus* (Karlov et al., 2003). The genetic degeneration of Y chromosome after divergence of the ancestors of *Cannabis* and *Humulus* deserves further investigation.

GENETICS OF HEMP BREEDING

Breeding targets

There are thousands of potential products that can be made from hemp, but breeding objectives generally fall within three main categories: fiber, seed, and, more recently, secondary metabolite production (non-THC cannabinoids and terpenes) (Salentijn et al., 2015). Due to this diversity of end uses, hemp breeding can progress in many different directions depending on the goals of the individual breeding program. It is important, as in any breeding program, to have clear goals that align with the production of specific end products.

Hemp's use as a durable fiber dates back thousands of years and archeological evidence shows that it was one of the first fiber plants domesticated by humans (Lynch et al., 2016). As such, improving fiber yield and quality have been primary breeding goals for hemp. Length of vegetation period is directly correlated to fiber yield, so selection based on this trait allowed for steady improvement of stem biomass in early cultivars (Ranalli, 2004). The proportion of bast fiber content to biomass, however, is a more complex trait to improve and little improvement was made until the Bredemann method was employed starting in 1942, which used bast fiber content as a primary criterion for selecting males (Salentijn et al., 2015). This *in vivo* method involved splitting the main stem to measure bast fiber content on living plants and only allowing males with the highest bast fiber content to flower (Ranalli, 2004). This enabled increased genetic gain and created plants with three times higher bast fiber content over the next 30 years (Ranalli, 2004). In 1953, Jakobey noted a negative correlation between bast fiber content and stem weight, so he developed a technique called the "normal axis" method to identify plants that broke pattern with the common correlation (Bócsa, 1999). Adding to this work, Horkay (1982) found that there was a strong negative relationship between bast fiber content and fiber quality, where increases in bast fiber were almost entirely secondary fibers of lower quality. After this period, however, breeding for fiber quality was no longer a priority because of the advent of new fiber processing techniques, such as steam explosion processing and ultrasonic refining (Bócsa, 1999).

Hemp's historic breeding efforts have largely focused on increasing fiber yield, but its potential as an oilseed crop is also being considered. Traditional dioecious hemp is grown as a seed crop, but these cultivars often exhibit significant variation and only produce seed on half of the plants. However, intersex plants are a common occurrence in dioecious hemp and can be selected for, with ideal stabilization of sex ratios within six to eight generations (Bócsa, 1999).

This creates a crop that has higher uniformity than dioecious types, and they have become common for oilseed or dual-purpose (seed and fiber) cultivars, although many dioecious oilseed varieties are utilized as well (Salentijn et al., 2015). The first true oilseed variety, "FIN-314", was developed in Finland using germplasm from the Vavilov Research Institute Gene Bank and was put into production in Canada in 1998 (Ranalli, 2004) after bans on hemp production were lifted (Salentijn et al., 2015). Canadian farmers in particular have turned to hemp as an alternative oilseed crop. With the support of government programs and a pre-existing oilseed infrastructure, oilseed hemp production and breeding have flourished in Canada, with 39 Canadian approved oilseed/dual-purpose cultivars listed as of 2014 (Salentijn et al., 2015).

While breeding for fiber and seed traits began long ago, a much newer breeding target has emerged in the form of cannabinoid profiles. The practical side of breeding for cannabinoid content lies in the legal restrictions on THC content (<0.3% dry weight worldwide, <0.2% in Europe). Any breeder that sells seed needs to be entirely sure that their variety is in compliance with standards in the production region. There is also current interest in medical hemp, which is harvested for *Cannabis*' other primary cannabinoid, cannabidiol or CBD. CBD is one of the major constituents that makes *Cannabis* medicine, with reported benefits as an antiepileptic, anticonvulsant, neuroprotectant, antioxidant, and as an anti-anxiety and anti-inflammatory agent (Devinsky et al., 2014). This has led to a boom in breeding high-CBD varieties of hemp. There are both qualitative and quantitative aspects to breeding for chemical phenotype (chemotype) and cannabinoid content, which will be discussed more extensively later. In Colorado and other states with legal marijuana it has become common to cross hemp varieties with marijuana strains to produce a predominantly CBD chemotype with high levels of overall cannabinoid production. The resulting progeny often has unstable THC levels and requires extensive breeding for uniformity and stability before seed

can reliably produce a hemp phenotype and be sold as industrial hemp. Most CBD farmers are circumventing this latter restriction by planting tested clones in the field, but this approach is cost and labor intensive and ultimately not scalable or sustainable in the same way that it is possible to produce crops from seed.

Along with these three major breeding objectives, hemp fills many specialty niche roles for which breeding will be integral. If industries develop around the use of hemp as paper, concrete, composites, textiles, specialty foods, bio-plastics, and more, industry-driven breeding of locally adapted cultivars that maximize specific plant components will become increasingly important (Salentijn et al., 2015).

Breeding Methods

Hemp is a naturally cross-pollinated crop, which, in the absence of strict selection, maintains high levels of natural genetic variation and heterozygosity within populations (Salentijn et al., 2015). As a result, most available hemp cultivars are populations that exhibit phenotypic variation. This can be a challenge for farmers, for instance, when plants are different heights, harvesting grain heads with a combine is problematic. Differences in maturity times can also result in seed loss. Because many hemp varieties were initially bred for fiber (and harvested prior to maturity), it is unclear how much effort was put into breeding for uniform height and reproductive maturity.

Historically, the most common approach to hemp breeding has been recurrent mass selection, where each generation's plants or seeds are selected to create the next generation based on a predetermined trait threshold (Ranalli, 2004). This approach has created many productive cultivars such as Bolognese, Toscana and YunMa 1 by improving landrace varieties, but is limited

by the intensity of selection and the heritability of the target traits (Salentijn et al., 2015; Hennink, 1994). Hennink (1994) reported that no studies had even reported estimates of heritability and it appears that maximizing response to selection has not traditionally been a primary goal for hemp breeders.

Despite a generally primitive approach to breeding, individual breeders recognized the potential of exploiting heterosis relatively early. Dewey (1927) is credited with creating the first intervarietal, or synthetic, hybrid by crossing Kymington and Ferrara. The resulting F_1 hybrid, Ferraramington, had excellent fiber characteristics and was one of several successful cultivars developed by Dewey (Bócsa, 1999). Unfortunately, all of Dewey's germplasm was lost due to *Cannabis* prohibition in the United States (Ranalli, 2004). Eventually, crossing varieties became more common as a way to generate new genetic variation in breeding populations and resulted in improved cultivars like the Chinese varieties YunMa 2 and YunMa 4 (Ranalli, 2004; Salentijn et al., 2015).

To maximize heterotic response, it is necessary to first develop inbred lines which are subsequently crossed to make true hybrid cultivars (Bernardo, 2002). One common difficulty in producing hybrids is pollen control, but utilizing self-fertilization in monoecious or subdioecious hemp can produce all female progeny which can act as a proxy for a male sterility system (Salentijn et al., 2015). This method has been used in Hungary and China to produce both single and double cross hybrids that greatly outperformed the parents (Salentijn et al., 2015; Bócsa, 1999). Despite these successes, true hybrid varieties of hemp are still relatively rare and there will need to be a paradigm shift in hemp breeding to realize gains similar to what was achieved in maize throughout the 1900's.

RECENT *CANNABIS* GENETIC STUDIES

Genetic Basis for Production of Secondary Metabolites

Due to the previously restricted ability to grow and handle *Cannabis* plants, research of *Cannabis* genetics and the development of genetic resources lags far behind other economically important crops. With the recent relaxation of these restrictions, a corresponding increase in all types of *Cannabis* research has emerged, particularly regarding industrial hemp. Research on the genetic basis of many traits has begun in the last decade, with a heavy initial focus on the genetics of cannabinoid production, as well as sex determination and agronomic traits (e.g., fiber quality).

Since the distinction between marijuana and hemp depends (legally) on the level of THC found in plant material, it has become of primary importance to understand the genetics of cannabinoid biosynthesis. Potential medical uses of *Cannabis* have also generated a significant amount of interest in THC's non-intoxicating isomer, CBD, as a pharmacological compound to treat a range of ailments from epilepsy to anxiety (van Bakel et al., 2011; Grotenhermen and Müller-Vahl, 2016, Felberbaum and Walsh, 2018). For many years it was thought that cannabidiolic acid (CBDA) was the direct precursor to tetrahydrocannabinolic acid (THCA) (Mechoulam et al., 1970; Shoyama et al., 1975). However the correct biochemistry of this process was elucidated more recently when it was discovered that cannabigerolic acid (CBGA) acts as a precursor to multiple compounds and produces either THCA or CBDA via enzymatic conversion with THCA synthase or CBDA synthase (Taura et al., 1995).

Cannabis has been described as having three common chemotypes distinguished by cannabinoid ratios: high THC/low CBD (marijuana), low THC/high CBD (hemp), and an intermediate ratio of the two compounds (hybrid) (de Meijer et al., 2003). At first glance, the

genetics of these categories seem straightforward. A single, codominant locus (*B*) appears to establish the chemotype with B_T and B_D alleles producing predominantly THC or CBD, respectively (de Meijer et al., 2003). It was also reported that other rare alleles, B_C and B_O, create rare chemotypes that produce mainly cannabichromenic acid (CBCA) and a non-functional, cannabinoid-free phenotype (de Meijer et al., 2003). In the same study, de Meijer et al. (2009) proposed that the B_O allele may actually be a linked second locus (*O*) where homozygous *O/O* combinations produce a normal range of cannabinoids, heterozygous *O/o* combinations severely reduce cannabinoid production, and homozygous recessive *o/o* combinations produce a cannabinoid-free phenotype. The B_O allele or *o/o* recessive genotype is particularly interesting as it represents a mechanism for creating industrial hemp cultivars that have nil levels of THC and can be easily utilized via molecular marker-assisted selection.

Weiblen et al. (2015) performed a quantitative trait locus (QTL) study which supported the single locus model of chemotype inheritance, but the distribution of cannabinoid synthase homologs in their mapping population indicated that two or more tightly linked loci could be controlling the trait, an idea initially proposed by de Meijer et al. (2009) regarding rare chemotypes and van Bakel et al. (2011) specifically regarding THCA/CBDA production. Despite previous efforts to categorize cannabinoid production as a qualitative trait, in reality the quantity of cannabinoids present in *Cannabis* flowers has proven to be a quantitative, polygenic trait. Weiblen et al. (2015) found significantly different quantities of cannabinoids in their study, with marijuana-type *Cannabis* averaging 4.5 times total cannabinoid levels compared to hemp. However, the small population size (N=62) did not allow for detection of any significant QTL for cannabinoid quantity. The linkage map reported in this study was created from an F_2 population, derived from

a cross between a single staminate hemp plant (Carmen) and a single pistillate marijuana plant (Skunk #1) and consists of 9 linkage groups, including 103 AFLP and 16 SSR markers, spaced 6.10 cM on average (Weiblen et al., 2015). Ultimately, the study detected only one significant QTL for qualitative chemotype characterization and one putative QTL for log THCA/CBDA content, located in two separate linkage groups. This supports the idea of separate loci affecting cannabinoid type and content, but the authors acknowledge an inability to detect sufficient QTL to fully characterize the genetic architecture of cannabinoid production and expect that with higher map density and larger populations it will be possible to detect more QTL and a tenth linkage group will emerge to properly reconcile the linkage map with known chromosome number (Weiblen et al., 2015).

Terpenoid production

Terpenoids, contributing to the scent and taste of *Cannabis* and commonly called essential oils, are a source of interest as phytotherapeutic agents, as well as for their hypothesized non-additive interactions with cannabinoids (Russo, 2011). These compounds are produced in terpene-rich resin, which is mainly synthesized and accumulated in glandular trichomes of female inflorescences in *Cannabis* (Booth et al., 2017). To date, over 100 terpenoids have been identified in *Cannabis*, prompting questions of both how these compounds are produced and what possible uses they could fulfill (Andre et al., 2016).

The first study investigating the genetics of terpene synthesis (Booth et al., 2017) showed that transcripts associated with terpene biosynthesis are expressed in glandular trichomes more than in non-resin producing tissues, agreeing with chemical analyses of these tissues. Genomic and transcriptomic data from the hemp variety 'Finola' enabled the identification of nine *Cannabis* terpene synthases (CsTPS) that account for the majority of terpene production, with the exception

of terpinolene which proved elusive (Booth et al., 2017). Similar to cannabinoid production, it appears that quantity of terpenes produced is polygenic and involves the production of competitive enzymes (Booth et al., 2017). Due to an intense interest in characterizing important pharmaceutical interactions in medical marijuana and an emerging interest in hemp as a source of medical and wellness products, research in this area is likely to expand rapidly in the near future.

Sex Expression in Hemp

Although *Cannabis* is mainly dioecious, monoecious plants are often observed in natural populations and can be intentionally induced via treatment with chemicals or environmental stress (Ram and Sett, 1982). These monoecious plants lack a Y chromosome, but are still able to produce staminate inflorescences. One interesting aspect of *Cannabis* is that "sex expression" in monoecious plants has been defined as a quantitative trait rather than a binary trait. A recent study (Faux et al., 2016) quantified sex expression in three hemp F_1 populations by the ratio of female and male flowers. Faux et al. (2016) utilized 71 AFLP markers to identify 5 QTL in each of three maps, and showed genetic correspondence of QTL across three maps. However, the study provided relatively low mapping resolution for sex expression due to the low number of markers.

Genetics of Agronomic Traits

There is an emerging picture of the genetics behind the production of secondary compounds and sex expression in *Cannabis*, but other traits remain unexplored. Although agronomic performance of hemp has been relatively well characterized, exploration of genetics behind major agronomic traits is just beginning. For hemp breeding and production to advance, it is necessary to understand how major quantitative traits are controlled.

Despite an historical breeding focus on fiber quality and quantity, the genetics of these traits are poorly understood. One initial study on fiber quality by van den Broeck et al. (2008) explored the molecular processes underlying cell wall synthesis to lay the groundwork for manipulating content of cellulose and lignin in hemp stem tissue. The authors looked at genes that were differentially expressed in the bast and hurd fibers using a cDNA microarray and found 110 clones with higher expression in bast tissue and 178 clones more highly expressed in hurd tissue. The genes preferentially expressed in the bast tissue were, expectedly, many genes associated with photosynthesis, chlorophyll, and chloroplast production, as well as arabinogalactan proteins. Most of the genes more highly expressed in the hurd tissue were directly related to enzymatic conversion of fructose-6-phosphate to various forms of lignin (van den Broeck et al., 2008). This is relatively unsurprising since the core is the woody section of the stem, but is an important characteristic. For instance, when using hemp fiber for making composite materials,. lignin can function as a useful binder, whereas the same compounds lower the quality for textile applications by adding undesired stiffness (van den Broeck et al., 2008). This study provides information about genes and gene families that are important to biosynthesis of commercially relevant traits, however, utilizing this information is difficult without further study of the degree of impact of individual genes or haplotypes.

Hemp has a long history as a fiber crop, but hemp grain (seed) has been utilized for at least 6,000 years as well (Li, 1973). One relevant area of study that has been explored in hemp seed is the genetics of fatty acid production. Hemp seed contains over 80% polyunsaturated fatty acids, with a desirable ratio of linoleic acid and alpha-linolenic acid, making it an common source of oil and protein for human and animal nutrition dating back to Neolithic times (Li, 1973; Bielecka et al., 2014). In a first of its kind study, Bielecka et al. (2014) created a TILLING population

(Targeting Induced Local Lesions in Genomes, Till et al., 2006) of industrial hemp from the oilseed cultivar Finola. This reverse genetics approach, which induces point mutations throughout the genome, allows researchers to observe altered phenotypes in mutant progeny and determine which gene sequences changed to produce these aberrant phenotypes. This particular study focused on Δ12 and Δ15 desaturase genes by comparing expressed sequence tags (ESTs) that showed homology to known desaturase genes. This approach identified 12 genes with membrane-bound expression in the FAD2, FAD3, and Δ6/Δ8 sphingo-lipid families and five genes for soluble Δ9 stearoyl-ACP desaturases. Utilizing M_2 plants with mutations in these genes, function of these oil metabolism genes was confirmed and a pathway was laid to produce specialized oil profiles in hemp, such as high-oleic hemp (Bielecka et al., 2014). This can have important commercial applications as oilseed varieties of hemp can be used for specific and unrelated end-uses like human consumption or production of biofuel.

The genetics of other important agronomic traits such as seed yield, biomass production, crop uniformity, photoperiod sensitivity, and flowering time have no published studies at the time of this writing. However, the Bielecka et al. (2014) study showed that despite a lack of major genetic resources such as an annotated and anchored genome in *Cannabis*, it is possible to use homology with other well-studied crops as a shortcut to understanding gene function. The availability of modern tools like affordable Next Generation Sequencing will help allow *Cannabis* researchers to rapidly catch up to other major crops in the coming decades.

MultiHemp

The TILLING study by Bielecka et al. (2014) was partially funded by the EU Framework Programme 7, MultiHemp. Multihemp is the first major government funded *Cannabis* research

initiative. The program ran from September 2012 to February 2017 with the goal of using "cutting-edge genomic approaches to achieve rapid targeted improvements in hemp productivity and raw material quality for end-user requirements, whilst also advancing scientific understanding of gene-to-trait relationships in this crop" (MultiHemp, 2017). The scope of the project was expansive and includedengineering for harvest and processing, hemp agronomy, crop modeling, and genetics/genomics. The Bielecka et al. (2014) study is the first genetics paper to be published from this project, but more genetics projects are underway describing the first Genome-wide Association Study (GWAS) and Heteroduplex mapping in hemp (MultiHemp, 2017). This project was an important step in breaking long-held stereotypes about *Cannabis*. If the United States is to properly contribute to hemp research, it is important for federal granting agencies like the United States Department of Agriculture (USDA) to rapidly create a path for publicly funded hemp research. Additionally, a permanent change in law regarding *Cannabis* and the distinction between marijuana and hemp would allow researchers to undertake projects without overly burdensome regulation.

GENOMICS AND GENETIC DIVERSITY

Reference genome and transcriptomes

To determine gene function in any species and understand the relationships between genes and haplotypes with phenotype, an accurate reference assembly is essential (Stemple, 2013). The marijuana strain Purple Kush (PK) was the first published genome in *Cannabis*, using a combination of Illumina and Roche 454 sequencing with ~130X coverage of the estimated ~820 Mb haploid genome. De novo assembly generated 136,290 scaffolds with a total size of 786.6 Mb, accounting for approximately ~96% of the estimated haploid genome size (van Bakel et al., 2011).

However, the genome coverage could be overestimated due to high proportion of redundant scaffolds of homologous regions with high heterozygosity rates (Vergara et al., 2016). Ongoing efforts to accurately assemble and annotate the genome are necessary to more clearly establish full genomic coverage.

A total of 30,074 transcript isoforms were constructed from the transcriptome assembly of PK, in which 83% have homologous counterparts in other plants. The remaining 17% may represent some unique gene models in *Cannabis*, but also likely represents assembly error and erroneous gene model prediction. Characterization of the transcriptome was paralleled by identification of differential gene expression in root, stem, shoot and three flowering stages. The expression profiles exhibit similar patterns in the 6 tissues because of widespread expression of photosynthetic processes and primary metabolic pathways in the plants (van Bakel et al., 2011). The authors also explored expression of THCA synthase and CBDA synthase and showed that they are expressed in opposite ways in the marijuana type (PK) and the hemp type (Finola), supporting that qualitative aspects of cannabinoid production are primarily determined by the presence or absence of these enzymes.

The released draft genomes and transcriptomes of marijuana types provide references for genetic variant detection and accelerate progress in genetic mapping and relating *Cannabis* genes to their functions. Additional genomic resources are being developed and are well reviewed by Vergara et al. (2016). These forthcoming resources will help answer questions such as: what is the content and function of repetitive regions, is there any evidence of ancestral whole-genome duplications, and what are the over- and under-represented gene families? The abundance of repetitive sequences and level of heterozygosity represent challenges in making a chromosome-scale assembly in *Cannabis*, which could be mitigated with targeted approaches utilizing inbred

lines and structured populations. It is also important to characterize sex chromosomes to clarify current pseudo-autosomes, male-specific loci, and the fate and consequences of genes on sex chromosomes.

Comparative genomics

Genomic resources for *Cannabis* are relatively sparse compared to model species such as *Arabidopsis thaliana*, so utilizing comparative genomics as a natural extension of *Cannabis* genomics research will answer questions regarding how *Cannabis* gene function is both similar to and different from other plant species. Due to strong preservation of homoeologous regions, "translation genomics" has been a successful approach for cross-utilization of genetic knowledge of closely related species (Paterson, 1995; Kim et al., 2012). Therefore, an obvious starting point for this approach is to utilize species most closely related to *Cannabis*. Using four plastid loci (atpB-rbcL, rbcL, rps16 and trnL-trnF), a molecular phylogenetic study confirmed the close genetic relationship between *Humulus* and *Cannabis* (Yang et al., 2013), two genera in Cannabaceae family) which diverged around 21-27.8 million years ago (Divashuk et al., 2014; Laursen, 2015). The group shows variation with regard to genome size and chromosome numbers among *C. sativa* (male: ~0.84Gb, 2*n*=20; female: ~0.81Gb, 2*n*=20), *H. japonicus* (male: ~1.7Gb, 2*n*=17; female: 2*n*=16) (the primitive type of *Humulus*) and *H. lupulus* (male: ~2.9Gb, 2*n*=20; female: ~2.57Gb, 2*n*=20) (van Bakel et al., 2011; Divashuk et al., 2014; Natsume et al., 2014). Understanding the patterns of evolution of genome size and structure among the members in Cannabaceae provides clues about the path of speciation and selection, and the fates of gene families, especially for sex expression.

The prevalence of atypical meiotic configuration, such as translocation heterozygosity, has been implicated in *Humulus* (Sinotô, 1929; Neve, 1958; Haunold, 1991; Shephard et al., 2000; Zhang et al., 2016). The findings shed light on questions on the unusual transmission genetics and phenotypic variation in hop, yet the abnormal meiotic events have not been reported in cytogenetic studies in *Cannabis* (Divashuk et al., 2014; Razumova et al., 2016). Due to a shared genetic origin with *Humulus* species, however, the possibility of atypical meiotic configuration may not automatically be ruled out in *Cannabis*.

Cannabis and *Humulus* are frequently characterized by their secondary metabolite systems, producing a variety of chemical compounds contributing to plant growth and human uses. One example of convergent breeding targets can be found in the selection of terpene profiles, which is commercially relevant for both and likely have similar genetic bases. Although this approach has only begun to be explored in *Cannabis*, important information can be gleaned in this manner, both by comparison with *Humulus* species as well as more distantly related but better characterized species like *A. thaliana* or maize (*Zea mays*).

Genetic Diversity and Population Structure

Debates over the speciation and classification of types of *Cannabis* are largely rooted in the phenotypic diversity that is apparent in the species. This has led to questions about what the total genetic diversity of *Cannabis* encompasses and how it is possible to understand and classify this diversity. It is also important to understand genetic patterns in the species that will allow variety and cultivar identification, purity inspection, lineage, and characterization of drug (or medicinal) and non-drug (or non-medicinal) strains. Identification can be addressed by a combination of morphology, chemistry, and genetic testing. To our best knowledge, three studies have investigated genetic diversity and population structures among hemp and marijuana. Sawler

et al. (2015) assessed the genetic patterns of 81 marijuana and 43 hemp samples using 14,031 SNPs characterized by genotype-by-sequencing (GBS). Lynch et al. (2016) investigated genetic structure of 340 accessions, which were a mixture of publicly available sequence (WGS and GBS) data and newly sequenced plants, representing three proposed categories based on reported ancestry and/or reported leaf shape: hemp, narrow-leaf drug-type (NLDT, i.e. sativa) and broad-leaf drug type (BLDT, i.e. indica). Dufresnes et al. (2017) conducted genetic analysis of 1324 samples collected from 24 hemp varieties and 15 marijuana strains using 13 SSR markers.

The three studies agreed on statistically significant population differentiation between hemp and marijuana types. However, results were not in agreement concerning whether hemp was more closely related to sativa or indica-types, or the comparison of heterozygosity rates between hemp and marijuana. Sawler et al. (2015) indicated that a hemp population collected from Canada, Europe and Asia is more genetically related to C. indica-type marijuana than to C. sativa strains and the hemp population exhibits higher heterozygosity rates than drug-types. Conversely, Lynch et al., (2016) demonstrated that European hemp varieties are more closely related to NLDT than to BLDT, with one exception of a Chinese hemp sample clustering with BLDT and that hemp varieties show less heterozygosity than drug-types, clearly divergent conclusions.

Although the studies differed in their conclusions about hemp's relatedness to indica or sativa groups, both agreed that there is a correlation between genetic structure and reported indica or sativa ancestry using a principal component approach (Sawler et al., 2015) and fastSTRUCTURE (Raj et al., 2014), and FLOCK (Duchesne and Turgeon, 2012) analyses (Lynch et al., 2016) and that these data support a genetic distinction between indica, sativa, and hemp groups. Dufresnes et al. (2017) took a forensic approach and did not attempt to draw a distinction between types of marijuana, simply comparing marijuana with hemp. Their analysis supported that

hemp and marijuana are genetically distinct (relating to cannabinoid production, which is generally acknowledged) and that unknown samples could be classified using these markers, but 13 SSR markers in a small population are insufficient to analyze genetic diversity in a way that can ascribe generalizations to the species or understand if there is a genetic basis for distinction outside of THC content.

Both of the studies that utilized substantial genomic coverage used small numbers of hemp samples, 22 in the Lynch et al. (2016) study and 43 in the Sawler et al. (2015) study, which does not fully encompass the genetic diversity of the group. There is also added confusion about distinguishing types of *Cannabis* because both groups found significant evidence of admixture between all three groups due to natural and human-directed hybridization and reported that marijuana strain names and ancestry data are inherently unreliable. Phylogenetic analyses are helpful in group comparisons, but can be misleading if generalizations are made when species diversity is underrepresented and quality genomes of ancestral species are not available. As more information is added to this debate, a clearer consensus will emerge on true allelic diversity throughout the genome, as well as characterizing population structure. Due to the extensive admixture of these groups and the fact that the basis of their distinction is rooted in a qualitative description of a quantitative phenotype, it is unlikely that population structure will neatly fall into the historically proposed "sativa" and "indica" subgroups.

Germplasm Resources

One of the problems with characterizing genetic diversity and population structure of *Cannabis* is a lack of access to diverse germplasm. Unlike most crops, no centralized germplasm repository exists for hemp. Americans and Canadians have long histories of producing hemp, but

North American germplasm resources were destroyed during *Cannabis* prohibition. Specifically, a coordinated effort was made to remove *Cannabis* accessions from gene banks in both the US and Canada around 1980 (Small and Marcus, 2003). Even in countries where hemp was not prohibited, many accessions were lost during periods of political turmoil or through displacement by other crops (Grigoryev, 2017; MultiHemp, 2017). There are a small number of gene banks that store *Cannabis* germplasm and a few working collections, but all of these organizations act independently and there has not been a collaborative effort of any kind to preserve the *Cannabis* gene pool.

The Vavilov Institute in Russia maintains the largest collection of approximately 500 accessions of hemp, representing many fiber and seed varieties as well as Chinese landraces (Ranalli, 2004). These are available for research and breeding, but a lack of funding has made maintenance of these accessions difficult for the Institute (Clarke, 1998). Another major gene bank has recently started preserving *Cannabis* germplasm as well. The Institute of Plant Genetics and Crop Plant Research Gatersleben (IPK) in Germany has a small collection of hemp accessions that are available for research and preservation. This collection contains approximately 55 accessions of cultivated and wild hemp (Graner, 2017). There are also a handful of gene banks that preserve limited collections of mostly local hemp accessions in Hungary, Turkey, Japan, and Italy (Ranalli, 2004). The largest of these collections is in Hungary where 70 local accessions are held, but the others have less than 20 accessions each (Ranalli, 2004).

In addition to gene bank preservation there are some working collections of *Cannabis* germplasm, but these are not freely available to the public and are not intended for long-term preservation (Ranalli, 2004). The most notable of these is the Dutch Center for Plant Breeding and Reproduction Research (CPRO)/Private Plant Research International (PRI) collection for the

Dutch 'National Hemp Programme'. It contains 204 accessions comprised of 74 cultivars, 51 landraces, 17 feral samples, and 65 accessions of unknown classification (Bas et al., 2015).

A generous estimate of extant *Cannabis* accessions would be around 1,000 samples total, and it would nearly impossible to access all of these. Additionally, not all of these accessions qualify as (or are) hemp which makes access or possession legally problematic. In comparison, a single germplasm bank at the International Maize and Wheat Improvement Center (CIMMYT) in Mexico maintains approximately 150,000 accessions of wheat (*Triticum aestivum*), which are publicly available for research and breeding (Pixley, 2017). Although a direct comparison with a major staple food crop is perhaps unfair, the sheer magnitude of the difference in germplasm resources highlights the fact that *Cannabis* researchers, breeders, and even farmers face significant challenges in obtaining or creating locally adapted germplasm. It is imperative that a collaborative international effort is undertaken in the near future to preserve the genetic diversity of this potentially important crop before more genes disappear permanently. For more on *Cannabis* gene bank accessions and the need for coordinated efforts to preserve the *Cannabis* gene pool, refer to Clarke and Merlin (2016).

FUTURE DIRECTIONS FOR *CANNABIS* GENETICS RESEARCH

One of the primary obstacles to advancing functional genetics research in *Cannabis* is the lack of an anchored and annotated genome. The highest quality draft genome for *Cannabis* was published by van Bakel et al. (2011) for PK. Assembly of this genome is difficult because of the lack of a linkage map and the fact that no closely related species have assembled genomes. Even with a genome for a related model species available, recent gene duplications and translocations reduce the accuracy of syntenic alignment (Wicker et al., 2011) and limits assembly to gene coding

regions due to rapid evolution of repetitive sequences in non-coding regions (Brunner et al., 2005). It has been recommended that for proper genome assembly at least one segregating population should be sequenced using a whole-genome shotgun sequencing approach to properly align sequences and correct for common errors (Gao et al., 2013; Mascher and Stein, 2014). Although the van Bakel et al. (2011) genome is the only publicly available reference genome, researchers are currently working on implementing structured populations and other approaches in industrial hemp to promote further exploration into gene function validation and genomic studies.

An important goal for breeding high-performance hybrid hemp is properly characterizing heterotic pools within the species. Although a number of genetic diversity studies have been published on *Cannabis* (Sawler et al., 2015; Lynch et al., 2016; Dufresnes et al., 2017), the limited access to representative samples has not allowed for a full characterization of the germplasm pool, by molecular marker analysis or otherwise. In the United States that problem is exacerbated by the fact that little historical data on hemp performance exists and most currently available germplasm is imported rather than developed locally. This lack of information can be viewed as an opportunity to characterize and curate representative *Cannabis* populations in a highly documented and organized fashion. Part of this organization should be the development of heterosis breeding. Determining relatedness using molecular markers, along with measuring mid-parent heterosis and assessing parent and F_1 performance in mega-environments, has successfully improved classification of maize varieties into heterotic pools (Livini et al., 1992; Reif et al., 2010). Similar strategies could be applied to hemp. As more trial data is collected and next-generation sequencing becomes standard, a comprehensive approach to forming heterotic pools can take place. This will put hemp breeding at an advantage to other crops that developed heterotic pools before these modern tools existed. Although it will still be necessary to develop high-performing inbred lines

and regularly test combining ability to assess true heterotic potential, access to modern tools allows breeders to make more rapid and informed choices that can both preserve genetic diversity in the species and maximize heterotic breeding efficiency.

In addition to characterization of agronomic performance, a deeper understanding of genotype by environment interactions (GEI or GxE) should be pursued. Hemp is very sensitive to environmental conditions and its inherent plasticity leads to different phenotypes when soil moisture status, temperatures, or day length change (Salentijn et al., 2015). This is particularly important regarding cannabinoid content since under the current regulatory framework, production of higher levels of THC can leave a farmer with a crop that must be destroyed rather than a marketable, hemp-based commodity. Using crop modelling to predict the effect of environment on cultivar performance has been successful (Amaducci et al., 2008) but is not a replacement for multi-environment trials (MET). Whether or not research on GEI and MET for hemp occurs in the private or public sector will depend on funding for these types of studies, but a push toward public research in this area would help to advance the understanding of GEI in hemp and hasten the development of locally adapted cultivars.

As was previously mentioned, our understanding of the genetics of important agronomic traits is woefully inadequate in *Cannabis*. Some initial studies utilizing QTL, GWAS, and TILLING have been performed (Weiblen et al., 2015; Faux et al., 2016; Salentijn et al., 2015; Bielecka et al., 2014), but these are merely first steps in truly understanding genotype to phenotype relationships. More of these types of studies as well as other functional genetic approaches should be used to further our understanding of the control of major traits, including: development of transgenic knockout lines, near-isogenic lines, and comparative genomics with model, crop, and

closely related species. Only when the biochemical pathways and genetic architecture of quantitative traits are understood will we be able to fully customize and utilize industrial hemp.

Along with traditional approaches to functional genetics, modern tools may be implemented to advance hemp breeding without identifying causal genes. Marker-assisted selection has become common in many crops, but is primarily only useful for qualitative traits and is limited to QTL that have been verified in breeding populations (Heffner et al., 2009). Genomic selection is a "black box" method that bypasses functional genetics and uses genotype and phenotype data from a training population to predict breeding values and performance of subsequent offspring (Bernardo and Yu, 2007). This approach is able to utilize all genomic information in a way that captures both major and minor allele effects and can more rapidly improve quantitative traits (Chakradhar et al., 2017). It is important to carefully design training populations to mitigate effects of population structure and composition, but genomic selection has been a qualified success in maize breeding programs (Chakradhar et al., 2017). Since hemp faces many similar breeding challenges to maize and high-quality sequence data continues to become more affordable, genomic selection has excellent potential as a breeding method to improve complex traits in hemp.

CONCLUSION

The future of research efforts with industrial hemp and *Cannabis* in general is promising. Although thirty to fifty years ago we saw a massive worldwide effort to eradicate both hemp and marijuana, and its legal status is still variable from place to place today, there has been recent and high-level acknowledgement and acceptance of the medical and industrial uses of the *Cannabis* plant. New molecular tools are allowing us to look into how the plant functions as well as delve

into the origins of the species and classify the wide breadth of diversity observed in the species. Although it is still too early to say whether or not describing official subspecies is warranted, initial studies have supported a genetic basis to distinguishing major gene pools of hemp and marijuana, as well as indica and sativa heritages. Genomics research in *Cannabis* is still in its infancy, but access to new technologies, combined with less restrictive rules governing industrial hemp research, promises a wealth of information to come. More collaborative, government-funded research, like the European Union MultiHemp project, is an absolute necessity to advance hemp research in a rigorous way that contributes to the evolution of the nascent industry. The multitude of uses possible for *Cannabis* warrants a methodological approach to fully understanding and characterizing the genetic architecture of important traits, so that the plant can be optimized for a variety of tasks. Only by modernizing our approach to understanding *Cannabis* genetics and genomics will it be possible to utilize and regulate production of this plant in a way that is truly beneficial, and in the most efficient manner possible. Every other crop of significant economic importance has been characterized in this way. We propose that it is time for *Cannabis* research to catch up, so that impactful, plant-based solutions are not overlooked or under-utilized

REFERENCES

Amaducci, Stefano, et al. "Modelling post-emergent hemp phenology (*Cannabis sativa L.*): theory and evaluation." *European Journal of Agronomy* 28.2 (2008): 90-102.

Andre, Christelle M., Jean-Francois Hausman, and Gea Guerriero. "*Cannabis* sativa: the plant of the thousand and one molecules." *Frontiers in plant science* 7 (2016).

van Bakel, Harm, et al. "The draft genome and transcriptome of *Cannabis sativa*." *Genome biology* 12.10 (2011): R102.

Bas, Noor, Marcel Toonen, and Luisa Trindade. "Current status of the Dutch Hemp collection." *Korea* 3.1: 4 (2015).

Bernardo, Rex. Breeding for quantitative traits in plants. No. 576.5 B523. Stemma Press, 2002.

Bernardo, Rex, and Jianming Yu. "Prospects for genomewide selection for quantitative traits in maize." *Crop Science* 47.3 (2007): 1082-1090.

Bielecka, Monika, et al. "Targeted mutation of Δ12 and Δ15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil." *Plant biotechnology journal* 12.5 (2014): 613-623.

Bócsa, Ivan. "Genetic improvement: conventional approaches." Advances in Hemp Research, Haworth Press, Binghamton, NY (1999): 153-184.

Booth, Judith K., Jonathan E. Page, and Jörg Bohlmann. "Terpene synthases from *Cannabis sativa*." *PloS one* 12.3 (2017): e0173911.

van den Broeck, Hetty C., et al. "Differential expression of genes involved in C 1 metabolism and lignin biosynthesis in wooden core and bast tissues of fibre hemp (*Cannabis sativa L.*)." *Plant science* 174.2 (2008): 205-220.

Brunner, Stephan, et al. "Evolution of DNA sequence nonhomologies among maize inbreds." *The Plant Cell* 17.2 (2005): 343-360.

Chakradhar, Thammineni, Vemuri Hindu, and Palakolanu Sudhakar Reddy. "Genomic-based-breeding tools for tropical maize improvement." *Genetica* (2017): 1-15.

Clarke, Robert C. "Maintenance of *Cannabis* germplasm in the Vavilov Research Institute gene bank-five year report." *Romania*1.4 (1998): 4.

Clarke, Robert C., and Mark D. Merlin. "*Cannabis* Domestication, Breeding History, Present-day Genetic Diversity, and Future Prospects." *Critical Reviews in Plant Sciences* 35.5-6 (2016): 293-327.

Devinsky, Orrin, et al. "Cannabidiol: pharmacology and potential therapeutic role in epilepsy and other neuropsychiatric disorders." *Epilepsia* 55.6 (2014): 791-802.

Dewey, L.H., 1927. Hemp varieties of improved type are result of selection. In: Yearbook of the USDA. (1927).

Divashuk, Mikhail G., et al. "Molecular cytogenetic characterization of the dioecious *Cannabis* sativa with an XY chromosome sex determination system." *PLoS One* 9.1 (2014): e85118.

Duchesne, Pierre, and Julie Turgeon. "FLOCK provides reliable solutions to the "number of populations" problem." *Journal of Heredity* 103.5 (2012): 734-743.

Dufresnes, Christophe, et al. "Broad-scale genetic diversity of *Cannabis* for forensic applications." *PloS one* 12.1 (2017): e0170522.

Faux, Anne-Michelle, et al. "Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis* sativa L.)." *Euphytica* 196.2 (2014): 183-197.

Faux, A-M., et al. "Identification of QTLs for sex expression in dioecious and monoecious hemp (*Cannabis sativa L.)." Euphytica* 209.2 (2016): 357-376.

Felberbaum, Michael and Sandy Walsh. "FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy". FDA News Release (2018). https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm611046.htm

Gao, Zhen-Yu, et al. "Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences." *Proceedings of the National Academy of Sciences* 110.35 (2013): 14492-14497.

Graner, Andreas. "Genebank Department". http://www.ipk-gatersleben.de/en/genebank/. IPK Gatersleben. Web. 18 Sept. 2017.

Grigoryev, Sergey. "Hemp (*Cannabis sativa L.*) genetic resources at the VIR: From the collection of seeds, through the collection of sources, towards the collection of donors of traits". http://www.vir.nw.ru/hemp/hemp1.htm. Vavilov Institute. Web. 18 Sept. 2017.

Grotenhermen, Franjo, and Kirsten Müller-Vahl. "Medicinal Uses of Marijuana and Cannabinoids." *Critical Reviews in Plant Sciences* 35.5-6 (2016): 378-405.

Haunold, Alfred. "Cytology and cytogenetics of hops." *Developments in Plant Genetics and Breeding (Netherlands)* (1991).

Heffner, Elliot L., Mark E. Sorrells, and Jean-Luc Jannink. "Genomic selection for crop improvement." *Crop Science* 49.1 (2009): 1-12.

Hennink, Sebastiaan. "Optimisation of breeding for agronomic traits in fibre hemp (*Cannabis* sativa L.) by study of parent-offspring relationships." *Euphytica* 78.1-2 (1994): 69-76.

Hillig, Karl W., and Paul G. Mahlberg. "A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae)." *American Journal of Botany* 91.6 (2004): 966-975.

Hillig, Karl W. "Genetic evidence for speciation in *Cannabis* (Cannabaceae)." *Genetic Resources and Crop Evolution* 52.2 (2005): 161-180.

Horkay, Mrs. "sorttartalom novelesere iranyulo szelekcio hatasa a kender primer es szekunder rostsejtaranyara." Novenytermeles *Crop production* (1982).

Karlov, G. I., et al. "Molecular cytogenetics in hop (*Humulus* lupulus L.) and identification of sex chromosomes by DAPI-banding." *Euphytica* 132.2 (2003): 185-190.

Kim, Changsoo, et al. "SSR-based genetic maps of Miscanthus sinensis and M. sacchariflorus, and their comparison to sorghum." *Theoretical and Applied Genetics* 124.7 (2012): 1325-1338.

Laursen, Lucas. "Botany: The cultivation of weed." *Nature* 525.7570 (2015): S4-S5.

Li, Hui-Lin. "An archaeological and historical account of *Cannabis* in China." *Economic Botany* 28.4 (1973): 437-448.

Livini, C., et al. "Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLPs." *TAG Theoretical and Applied Genetics* 84.1 (1992): 17-25.

Lynch, Ryan C., et al. "Genomic and chemical diversity in *Cannabis*." *Critical Reviews in Plant Sciences* 35.5-6 (2016): 349-363.

Mascher, Martin, and Nils Stein. "Genetic anchoring of whole-genome shotgun assemblies." *Frontiers in genetics* 5 (2014).

Mechoulam R, Shani A, Yagnitinsky B, Ben-Zvi Z, Braun P, Gaoni Y. 1970.
Some aspects of cannabinoid chemistry. In: Joyce CRB, Curry SH, eds. The
botany and chemistry of *Cannabis*. London, UK: Churchill, 93–115.

de Meijer, Etienne PM, et al. "The inheritance of chemical phenotype in *Cannabis sativa L.*" *Genetics* 163.1 (2003): 335-346.

de Meijer, E. P. M., K. M. Hammond, and A. Sutton. "The inheritance of chemical phenotype in *Cannabis sativa L.*(IV): cannabinoid-free plants." *Euphytica* 168.1 (2009): 95-112.

Ming, Ray, Abdelhafid Bendahmane, and Susanne S. Renner. "Sex chromosomes in land plants." *Annual review of plant biology* 62 (2011): 485-514.

MultiHemp, http://multihemp.eu/project/. Web. 2017.

Natsume, Satoshi, et al. "The draft genome of hop (*Humulus lupulus*), an essence for brewing." *Plant and Cell Physiology* 56.3 (2014): 428-441.

Neve, R. A. "Sex chromosomes in the hop *Humulus lupulus*." *Nature* 181.4615 (1958): 1084-1085.

Paterson, Andrew H. "Molecular dissection of quantitative traits: progress and prospects." *Genome Research* 5.4 (1995): 321-333.

Pixley, Kevin. "Germplasm Research". http://www.cimmyt.org/germplasm-bank/. CIMMYT. Web. 18 Sept. 2017.

Raj, Anil, Matthew Stephens, and Jonathan K. Pritchard. "fastSTRUCTURE: variational inference of population structure in large SNP data sets." *Genetics* 197.2 (2014): 573-589.

Ram, HY Mohan, and R. Sett. "Induction of fertile male flowers in genetically female Cannabis sativa plants by silver nitrate and silver thiosulphate anionic complex." *Theoretical and Applied Genetics* 62.4 (1982): 369-375.

Ranalli, Paolo. "Current status and future scenarios of hemp breeding." *Euphytica* 140.1 (2004): 121-131.

Razumova, Olga V., et al. "Molecular cytogenetic analysis of monoecious hemp (*Cannabis* sativa L.) cultivars reveals its karyotype variations and sex chromosomes constitution." *Protoplasma* 253.3 (2016): 895-901.

Reif, Jochen C., et al. "Broadening the genetic base of European maize heterotic pools with US Cornbelt germplasm using field and molecular marker data." *Theoretical and applied genetics* 120.2 (2010): 301-310.

Russo, Ethan B. "Taming THC: potential *Cannabis* synergy and phytocannabinoid-terpenoid entourage effects." *British journal of pharmacology* 163.7 (2011): 1344-1364.

Sakamoto, Koichi, et al. "Characterization; genome sizes and morphology of sex chromosomes in hemp (*Cannabis* sativa L.)." *Cytologia* 63.4 (1998): 459-464.

Salentijn, Elma MJ, et al. "New developments in fiber hemp (*Cannabis sativa L.*) breeding." *Industrial Crops and Products* 68 (2015): 32-41.

Sawler, Jason, et al. "The genetic structure of marijuana and hemp." *PloS one* 10.8 (2015): e0133292.

Shephard, H. L., et al. "Sexual development and sex chromosomes in hop." *New Phytologist* 148.3 (2000): 397-411.

Shoyama Y, Yagi M, Nishioka I, Yamauchi T. Biosynthesis of cannabinoid acids. *Phytochemistry* 14 (1975): 2189–2192.

Sinotô, Yosito. "On the tetrapartite chromosome in *Humulus lupulus*." *Proceedings of the Imperial Academy* 5.1 (1929): 46-47.

Small, Ernest, and Arthur Cronquist. "A practical and natural taxonomy for *Cannabis*." *Taxon* (1976): 405-435.

Small, Ernest, and David Marcus. "Tetrahydrocannabinol levels in hemp (*Cannabis* sativa) germplasm resources." *Economic botany*57.4 (2003): 545-558.

Stemple, Derek L. "So, you want to sequence a genome..." *Genome biology* 14.7 (2013): 128.

Taura, Futoshi, et al. "First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis." *Journal of the American Chemical Society* 117.38 (1995): 9766-9767.

Till, Bradley J., et al. "A protocol for TILLING and Ecotilling in plants and animals." *Nature protocols* 1.5 (2006): 2465.

Vergara, Daniela, et al. "Genetic and genomic tools for *Cannabis* sativa." *Critical Reviews in Plant Sciences* 35.5-6 (2016): 364-377.

Watts, Geoff. "Science commentary: *Cannabis* confusions." *BMJ: British Medical Journal* 332.7534 (2006): 175.

Weiblen, George D., et al. "Gene duplication and divergence affecting drug content in *Cannabis sativa*." *New Phytologist* 208.4 (2015): 1241-1250.

Wicker, Thomas, et al. "Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives." *The Plant Cell Online* 23.5 (2011): 1706-1718.

Yang, Mei-Qing, et al. "Molecular phylogenetics and character evolution of Cannabaceae." *Taxon* 62.3 (2013): 473-485.

Zhang, Dong, et al. "Non-Mendelian inheritance of SNP markers reveals extensive chromosomal translocations in dioecious hops (*Humulus lupulus L.*)." *bioRxiv* (2016): 069849.

CHAPTER 2:

GENOTYPE-BY-ENVIRONMENT INTERACTIONS OF INDUSTRIAL HEMP (*CANNABIS SATIVA L.*) CULTIVARS HIGHLIGHT DIVERSE RESPONSES TO ENVIRONMENTAL FACTORS

SUMMARY

Little is known regarding the genetic mechanisms controlling economically relevant traits. Particularly, with Federal legality of the crop hinging on a stringent tetrahydrocannabinol (THC) content of 0.3% or less, it is necessary to assess variance in this trait due to environmental effects and genotype-by-environment interactions (GEI) to avoid running afoul of Federal law. Understanding how physical and biochemical traits respond to the environment also plays a strong role in selecting and developing appropriate cultivars for production in diverse growing regions. In 2016 we performed cultivar trials in multiple environments in Colorado to assess performance characteristics of a diverse set of germplasm from breeding programs across Europe and Asia. From these data, we were able to identify traits nearly entirely controlled by genetic factors, like days to maturity and THC and CBD production. We also identified traits strongly influenced by the environment and GEI, like grain yield, plant height, and water use. Individual cultivars also exhibited widely varying degrees of sensitivity to the environment. This underscores the importance of continued work to characterize genetic control of hemp traits to expedite breeding of cultivars that are well-adapted to target growing regions.

INTRODUCTION

Industrial hemp (*Cannabis sativa L.*) has been primarily a fiber or a dual purpose (fiber and grain) crop (Amaducci et al., 2015), but it is generally recognized as a multipurpose crop having a vast diversity of actual and potential end uses, including food, fuel, textile, non-intoxicating medicine, and many industrial products like lubricants, bioplastics, paper, and concrete (Amaducci & Gusovious, 2010, Salentijn et al., 2015; Schluttenhofer and Yuan, 2017). The United States placed de facto prohibition on hemp farming until 2014. In 2014 the US Farm Bill (U.S. Govt., 2014) defined hemp as *Cannabis sativa* with a tetrahydrocannabinol (THC) content of 0.3% or less and permitted cultivation of the crop in states with amenable laws under the oversight of University research or State Departments of Agriculture (U.S. Govt., 2014). Since then, most US states have passed legislation to allow cultivation of hemp. The 2018 Farm Bill removes hemp from the Controlled Substances Act and will force various federal agencies to regulate hemp like "normal" crops.

Since no US germplasm has been maintained through the 77 years of prohibition and coordinated efforts were made to remove hemp seeds from American gene banks during the Reagan Drug War era (Small and Marcus, 2003), only feral hemp remains in the US and testing unadapted germplasm from abroad is a logical starting point for the reintroduction of hemp. This study represents the first attempt in Colorado to perform hemp trials using germplasm developed abroad. These types of cultivar trials help to establish both the fitness of current cultivars for production in a novel target population of environments, as well as to assess stability of commercially important traits. Due to inherent plasticity, it is important that these newly introduced cultivars are tested in multiple environments to assess the degree of trait variance in the phenotypes produced by a cultivar. These multi-environment trials are common in plant

breeding, helping to discriminate cultivars with wide or specific adaptation, as well as determining environmental factors that are strongly influencing traits of interest (Makahiya et al., 2017).

Utilizing these types of multi-environment trials, it is possible to discriminate trait variation that is due to genetic effects, environmental effects, and genotype-by-environment interactions (GEI). This is useful information to ascertain whether observed phenotypes are a product of genetic factors, which may be heritable, versus environmental factors which are not. Trait variance due to GEI is also interesting because it demonstrates a range of plasticity in a population. Visible or measurable phenotypes are a complex product of many factors and dissecting trait variance in this way is an important first step in understanding how specific phenotypes are produced, as well as the overall stability or plasticity of a trait or cultivar. Cultivars that are less sensitive to environmental changes exhibit broad adaptation and may be suitable for production in a wide range of environments, while more plastic cultivars may be highly adapted to a specific region, but an unsuitable choice for other areas.

Yield of all crops in Colorado is limited by available precipitation. This is particularly true for summer row crops, where lack of soil moisture affects stand establishment, plant development, and yield. In this experiment, we manipulated irrigation as a major environmental factor to which genotypes may differ in response and which plays a significant role in crop management. One of the primary traits of interest to farmers is cannabinoid content, particularly THC content due to its legal implications. THC imparts the psychoactive effects that cannabis is known for and if a crop of industrial hemp tests above the stringent threshold of 0.3%, it is in violation of Federal law. There is also current interest in cannabidiol (CBD) as THC's non-psychoactive counterpart with potential therapeutic uses covering a wide range of ailments

(Gallily et al., 2015). Little has been published on the genetic control of cannabinoids in general. A single quantitative trait locus (QTL) study characterized genetic control of chemotype (high THC/low CBD, high CBD/low THC, or intermediate types), but did not detect any significant QTL for cannabinoid content (Weiblen et al., 2015). Although this is an important first step in understanding the genetic architecture of cannabinoid production, understanding the relative influence of genetics and environment warrants further exploration to mitigate risk and inform efforts to create high yielding, Federally compliant cultivars. In addition to major cannabinoids hemp also produces a wide range of terpenes that contribute aroma and flavor to the hemp flower, act as a natural insecticide, and have potential medicinal and therapeutic uses (Russo, 2011). These compounds can be extracted from the hemp flower to create additional value-added byproducts for hemp farmers.

However, the chemical makeup of hemp is not its only important feature. Depending on markets and infrastructure the most valuable parts of the plant may vary, but the majority of its potential products are produced in the flowers, stalks, and grain. Therefore, a host of agronomic characteristics are important for successful crop production, including grain and fiber yield and crop phenology. We assessed many of these agronomic traits, as well as measuring carbon isotope ratio in leaves, which is directly related to water use efficiency in plants (Donovan and Ehleringer, 1994). There is a clear need for all performance-related and physiological traits to be explored to help hemp breeding and production "catch up" to other crops that have not had the same legal restrictions and stigmatization. GEI studies are a first step to selecting or breeding appropriate hemp cultivars for new environments, in the United States and abroad. The aim of this study is to assess the performance of a diverse set of cultivars from Europe and Asia in multiple environments representative of Colorado's population of target environments and to

determine the extent to which genetic and environmental factors influence economically important traits.

MATERIALS AND METHODS

Plant Materials

Seed from a set of 13 cultivars of industrial hemp collected in the frame of the European project Multihemp (multihemp.eu) were imported from Italy to Colorado for the 2016 cultivar trials. Supplementary Table 2.1 lists the cultivars assessed, as well as their country of origin, latitude, and whether or not the cultivars were dioecious or monoecious types. Monoecious types have been bred to be "dual-purpose" i.e. useable for both grain and fiber applications, while dioecious types are generally intended for fiber production (Tang et al., 2017).

Environments

Trials were planted at two separate locations, representing distinct growing regions of Colorado. One location was at Colorado State University's primary research farm, the Agricultural Research Development and Education Center (ARDEC). This facility is in Fort Collins, Colorado at latitude 40.65 and longitude -105.00, with an elevation of 1,557 m and average annual precipitation of 408 mm. The soil texture at this location is a sandy clay loam. ARDEC is equipped with an overhead linear sprinkler irrigation system, which was used to irrigate these trials. The second location was at the Southwestern Colorado Research Center, near Yellow Jacket, Colorado at latitude 37.53 and longitude -108.72. This site has an elevation of

2,103 m and an average annual precipitation of 405 mm. The soil at this research facility is Wetherill silty clay loam and irrigation was applied by an overhead pivot sprinkler system.

To reduce visibility of the trials, dent corn (*Zea mays*) was planted as borders on all sides of the experiments. The seedbed was prepared by disking and soil tests were performed to determine the levels of minerals present in the soil. Target fertilizer levels were 112 kg ha^{-1} nitrogen, 90 kg ha^{-1} phosphorous, 112 kg ha^{-1} potassium, and 39 kg ha^{-1} sulfur. To achieve these levels, 54 kg ha^{-1} nitrogen, 45 kg ha^{-1} phosphorous, and 13 kg ha^{-1} sulfur were incorporated into the soil at the Fort Collins site and 61 kg ha^{-1} nitrogen and 39 kg ha^{-1} of phosphorous were added at Yellow Jacket. Glyphosate (RoundUp Powermax, Monsanto) and ethalfluralin (Sonolan, Dow Agrosciences) were applied before planting to remove weeds.

Planting methods were the same at both sites. A seed drill was used to plant at a depth of approximately 2.5 cm, with a target plant density of 120 plants per square meter (Tang et al., 2017). Seeding rates were increased from that target density proportional to germination rates among cultivars and this rate was then doubled to account for expected seedling mortality due to challenging soil conditions in Colorado. Weeds were controlled by hand and no pesticides were applied during either growing season. Weather data for both locations were obtained from the Colorado Agricultural Meteorological network (COAGMET) and the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC).

Precipitation and Irrigation

Yellow Jacket received more rainfall than Fort Collins. For the purpose of analyzing environmental data, we defined the growing season as the date of planting to the date that the last plot was harvested. In 2016, minimal rainfall was experienced at Fort Collins, with growing

season rainfall of 53 mm (27.3% of the 30-year average for the growing season). Two irrigation treatments were utilized at Fort Collins (described below). A "limited irrigation" treatment received 147 mm of overhead sprinkler irrigation and "fully irrigated" treatment received 398 mm of supplemental irrigation throughout the season. Yellow Jacket also experienced lower than average rainfall, with a season total of 121 mm representing 73.6% of the 30-year average. Irrigation at this site was also delivered by overhead sprinkler irrigation in the amount of 203 mm. Precipitation for the 2016 growing season is compared with 30-year averages in Supplementary Table 2.3.

Experimental Design

At Fort Collins, a Latinized row-column design was used, while a randomized complete block design was used at Yellow Jacket. In both locations, cultivars were replicated four times in the experiment, but with slightly different plot sizes. The plots at Fort Collins were six row plots that were 6.1 m in length with approximately 0.25 m spacing between rows. Plots at Yellow Jacket were six row plots, 9.1 m in length with 0.20 m spacing between rows. Because previous data for hemp evapotranspiration and irrigation rates in Colorado were not available, irrigation rates were adjusted as necessary throughout the season.

At Fort Collins, two different irrigation treatments were employed in 2016, one "fully irrigated" treatment approximating 100% of evapotranspiration and one "limited irrigation" treatment where irrigation was only applied to establish seedlings and if severe wilting was observed, effectively doubling the size of the experiment, i.e. four replications of the set of cultivars in each treatment. These treatments are referred to as "wet" and "dry", respectively and were treated as discrete environments. The differential irrigation was employed to explore the

effects of drought stress on hemp productivity. At Yellow Jacket, only a "fully irrigated" treatment was planted due to resource limitations. The Fort Collins trial was planted on May 31, 2016 and the Yellow Jacket trial was planted on June 7, 2016.

Data Collection

Plant height was measured as the vertical distance from the soil surface to the apical tip of a plant. Plant heights were measured once weekly from mid-July until plants stopped growing near maturity. Five random plants were chosen per plot each week and all four replications were measured. Plant heights reported are final plant heights.

The procedure used for determining flowering stage and maturity is fully described in Supplementary Table 2.2. The date that more than 50% of plants in a plot reached a given growth stage was scored and the number of days that elapsed between planting and that stage were calculated. Observations were taken three times weekly for these traits and were based on visual observations. Plant maturity was only measured for female/monoecious plants and was considered as seed maturity, i.e. when bracts began to dehisce and darkening of the seed coat was visible. Plots were harvested within three days of being scored for seed maturity and plants were bundled and hung to dry.

Total plant biomass, referred to as biomass or dry biomass, was measured as the mass of the aboveground portion of all of the plants in a plot. Plants were cut at the soil surface and air-dried for a minimum of 30 days. The plants were then weighed before threshing.

Grain was separated from the flowers using a mechanical thresher (Almaco, Nevada, Iowa) and seed was cleaned using a column blower (Agriculex, Guelph, Ontario, Canada). Grain

was air-dried to approximately 8-10% seed moisture, as determined by a GAC 500XT grain moisture tester (Dickey-John, Auburn, IL) and weighed to obtain grain yield.

Stand establishment was calculated as the number of plants standing at harvest, divided by the total number of seeds planted, and multiplied by 100. Since this trait was based on an "over-seeding" rate, a stand establishment rate of 50% is considered optimal.

Biochemical Trait Analysis

Biochemical traits were analyzed from flower and leaf samples collected at Fort Collins only. The top 10 cm of female flowers were collected from three random plants per plot at maturity. Seeds were removed from the flowers by hand and composite samples were made with the flower chaff from each plot. Cannabinoid and terpene profiles were analyzed using high-performance liquid chromatography (HPLC) and gas chromatography with flame ionization detector (GC-FID) by ProVerde Labs (Milford, MA). Carbon isotope composition was measured by mass spectrometry of dried and ground leaf samples as described by Ehleringer and Osmond (1989).

Sample Preparation for Cannabinoid Analysis

Sample preparation for the analysis of cannabinoid profiles was performed by extraction of the cannabinoids in organic solvent. Approximately 300 mg of homogenized plant material was extracted with 4 ml of isopropanol with sonication for 20 minutes. The resulting extract was filtered with a syringe filter, and further diluted with 71% acetonitrile (ACN) to the appropriate concentration for LC analysis, and transferred to an auto-sampler vial.

Chromatographic Cannabinoid Analysis

The liquid chromatographic analyses were performed using an ultra-high-pressure liquid chromatographic system (Waters UPLC) with Photo Diode Array, UV Detection (PDA), with a Cortecs C18 column (2.7 µm, 2.1 mm x 100 mm) (Waters Corporation, MA). Mobile phases were water (A) and acetonitrile (B), both acidified with 0.1% formic acid. Separation was achieved under gradient conditions of 59-100% mobile phase B over 2.5 min at a flow rate of 0.56 ml min^{-1} at 40°C. Samples were introduced with a 3.5 µl injection, with chromatographic data collected at 225 nm.

Cannabinoid certified reference materials (Cerilliant, Sigma-Aldrich and Cayman Chemicals) were used for peak identification and generation of calibration curves used for quantitation, and included: THC, THCA, CBD, CBDA, CBG, CBGA, CBC, CBN, THCV and CBDV. Data was recorded and processed using Empower Software (Version 3, Waters Corporation).

Sample Preparation for Terpene Analysis

Analysis of terpene profiles was performed using Full Evaporative Technique GC-FID Chromatography (FET-GC-FID). The Full Evaporative Technique is a form of head-space sampling, for which standards or samples are placed and sealed directly in a head space vial. The sealed vial is equilibrated at elevated temperatures to vaporize volatile compounds for head-space sampling. For these evaluations, samples were homogenized and sealed directly in to the head-space vials, then equilibrated for 30 minutes at 140°C prior to injection using a Hewlett Packard head-space autosampler (HP G1290A).

Chromatographic Terpene Analysis

The GC analyses were performed using Shimadzu GC-2014 gas chromatograph with Flame Ionization Detection (FID), with an Rxi-624Sil MS column (30 m x 0.25 mm x 1.4 µm) (Restek, Bellefonte, PA). Samples were introduced directly from the head-space auto sampler via a transfer line held at 160°C to prevent condensation of sample vapors prior to injection.

Nitrogen was used as the GC carrier gas at a flow rate of ~80 ml min^{-1}. Hydrogen and compressed air were used as the combustion gases. The following instrument parameters were employed: air, 50 psi; hydrogen, 70 psi; nitrogen, 60 psi; linear velocity flow control, 33 cm s^{-1}; split ratio, 20:1; injector temperature, 250°C; detector temperature, 320°C; oven program, 75°C (hold 0.4 min) to 160°C at 8°C min^{-1}; ramped to 250°C at 20°C min^{-1}; ramped to 300°C at 12.5°C min^{-1} (hold 3 min); run time, 22.2 min.

Terpene certified reference materials (Restek CRMs #34095 and 34096) were used for peak identification and generation of calibration curves used for quantitation. Data was recorded and processed using Clarity Software (Version 5.0.4.158).

Statistical Analysis

Analysis of variance (ANOVA) was used to determine effects due to genotype, environment, and genotype-by-environment interactions by linear regression using the aov function in R (R Core Team, 2013), with genotype and environment treated as fixed effects to determine significance and as random effects to calculate variance explained. Traits in common were analyzed across all environments and traits only measured at Fort Collins were analyzed across the wet and dry irrigation treatments. Data were organized and visualized using the tidyverse and ggplot2 packages in R (Wickham, 2017; Wickham, 2009). Before the regression

analysis, traits were tested for normality with a Shapiro-Wilk test in R (R Core Team, 2013). Non-normally distributed data were transformed using the bestNormalize package in R (Peterson, 2017) to determine the optimal transformation for each trait. Percent variance explained was calculated as the sum of squares for each variable divided by the total sum of squares. Pearson's product moment correlation analysis was performed using the corrplot package in R (Wei, 2013). The Additive Main Effect and Multiplicative Interaction (AMMI) biplot was created using the agricolae package in R (de Mendiburu, 2015).

RESULTS

Physical Traits

Grain Yield

Mean grain yield for the cultivars tested varied dramatically, from 27 kg ha^{-1} to 2366 kg ha^{-1}. These observed extremes were a nearly equal contribution of genetics, environment, and GEI (Table 2.1). Genetic effects explained 28% of trait variance, with 25% of variance attributed to environmental effects and 35% attributed to GEI. As expected, grain yield increased with additional irrigation, with a limited irrigation mean yield of 404 kg ha^{-1} at Fort Collins and a fully irrigated mean yield of 782 kg ha^{-1} and 1123 kg ha^{-1} at the Yellow Jacket and Fort Collins sites, respectively. However, not all cultivars were affected equally by increased access to water. While every cultivar produced more grain with increased irrigation, some exhibited a much more plastic response than others (Figure 2.1). The yield reaction norm shows characteristic divergent GEI (Malosetti et al., 2013). The cultivar 'Félina 32' produced the most grain under both drought and full irrigation at Fort Collins and also showed the largest response to increased irrigation.

This cultivar produced an average of 611 kg ha^{-1} of grain under limited irrigation and 2337 kg ha^{-1} of grain under full irrigation. The cultivar 'Uso 31' was not included in the reaction norm due to grain loss in the limited irrigation treatment plots caused by birds during the harvest process (Figure 2.1). Grain yield GEI was also visualized using an AMMI biplot (Figure 2.2). Cultivars closest to specific environmental vectors exhibited higher yields in those environments and the direction of the vectors shows whether those environments produced yields that were above or below average. Means and standard deviation for all traits are reported by cultivar in Supplementary Tables 2.4.1-2.4.12.

Total plant biomass

Aboveground total plant biomass showed a more substantial influence of environmental factors than grain yield, with 61% of trait variance due to environmental effects. Genotype and GEI were responsible for a similar amount of variance, at 11% and 13%, respectively. Although the fully irrigated plots at Fort Collins produced more grain than at Yellow Jacket, the opposite was true for biomass. Under limited irrigation at Fort Collins, plants produced an average of 2,482 kg ha^{-1}. Under full irrigation at Fort Collins, average dry biomass was 6,239 kg ha^{-1} and, under full irrigation at Yellow Jacket, produced an average of 6,834 kg ha^{-1}. However, biomass was more variable at Yellow Jacket than at Fort Collins, with a standard deviation of 1,545 kg ha^{-1}. Comparatively, standard deviation at Fort Collins was 516 kg ha^{-1} under limited irrigation and 767 kg ha^{-1} under full irrigation.

Days to Maturity

Due to resource limitation, days to maturity was not measured at Yellow Jacket, so these data represent the differentially irrigated environments at Fort Collins. The number of days from planting to maturation of grain had a very strong genetic component, with 97% of variation observed at Fort Collins attributed to genetic effects. There was not a significant effect of environment (irrigation) for this trait, but GEI accounted for slightly over 1% of trait variation. There was a 40-day spread for maturity at Fort Collins in 2016, from 93 days to 133 days. Mean days to maturity were nearly identical in the limited and fully irrigated treatments at 116 and 117 days, respectively. However, crossover GEI led some cultivars to take longer to mature under full irrigation, while others matured more quickly with increased irrigation. In general, this trait was largely determined by genotype with only small differences in maturity due to interactions with the environment.

Plant Height

Plant height was strongly influenced by both genetic and environmental factors. Genotype accounted for 36% of variance in height, while environment and GEI accounted for 38% and 9%, respectively. Plant height followed the expectation that increasing irrigation would result in taller plants. Mean plant height at maturity was 135 cm under limited irrigation at Fort Collins. This increased to 153 cm at Yellow Jacket and to 181 cm at Fort Collins when fully irrigated. Values for mean plant height by cultivar ranged from 113 cm to 210 cm, with later flowering/maturing cultivars exhibiting taller phenotypes.

Stem Diameter

Although stem diameter showed nearly the same degree of genetic effects (36%) as plant height, a much smaller proportion of variance was attributed to environmental effects (8%). There was a statistically significant and slightly higher amount of variance attributed to GEI at 15%. There was a large amount of residual variance for this trait as well (41%). Cultivar means for stem diameter had a minimum of 4.52 mm and a maximum of 11.70 mm. Similar to grain yield and plant height, stem diameter in general increased with more access to water. The limited irrigation treatment at Fort Collins had the smallest mean stem diameter at 5.77 mm. Yellow Jacket produced a mean stem diameter of 6.61 mm, while the fully irrigated treatment at Fort Collins had a mean diameter of 7.06 mm.

Stand Establishment

The composite trait of stand establishment showed highly significant effects of genotype, environment, and GEI, with the highest proportion assigned to environment. Genetic effects accounted for 24% of trait variance. Environmental effects were the strongest contributing factor at 51%, while GEI accounted for a smaller, but significant (p-value = 0.02), 7% of trait variance. Stand establishment was lowest in the limited irrigation treatment at Fort Collins, with a mean of 14%. Establishment was higher under full irrigation at 19%, while the highest stand establishment was observed at Yellow Jacket with a rate of 29%. GEI for this trait changed rankings dramatically, but some cultivars were relatively consistent across locations and treatments. For instance, ‘Diana’ had the lowest stand establishment rates in all environments. However, other cultivars were highly variable. For example, the cultivar ‘Bialobrzeskie’ was ranked at 12/13 at Fort Collins Dry (5%), 11/13 at Fort Collins Wet (16%) and 1/13 at Yellow

Jacket Wet (38%). Plant density, from which stand establishment was calculated, ranged from 8-111 plants per square meter, with a mean of 48 plants per square meter.

Biochemical traits

Cannabinoids

To account for the loss of the carboxylic acid group during decarboxylation, total tetrahydrocannabinol (THC) levels were calculated as the sum percentage by dry weight of tetrahydrocannabinolic acid (THCA)*0.877 plus delta-9-tetrahydrocannabinol (Δ9-THC). THC content was primarily an effect of genotype, with a small, but significant effect of environment. Over 80% of trait variance was attributed to the effect of genotype, while 1.7% of variance was explained by environment. The mean across treatments at Fort Collins in 2016 for total THC was 0.14%. The mean THC content by cultivar ranged from 0.002% to 0.63%, with a median value of 0.09%. A single cultivar, 'Tiborszállási', tested over the 0.3% threshold in 2016.

The total cannabidiol (CBD) content was also calculated as the sum of the acid form and decarboxylated form of the compound, cannabidiolic acid (CBDA) multiplied by 0.877 and cannabidiol (CBD), respectively. Similar to THC, the regression model detected significant effects of genotype and environment. The magnitude of effects was also very similar. Genotype explained 83% of variation in CBD content, with 6% of variation attributed to environmental effects. The overall mean content of total CBD varied by treatment, with a fully irrigated mean value of 2.24% and a limited irrigation mean of 1.43% of dry flower weight. The highest mean CBD content was found in the Italian fiber cultivar 'Carmagnola Selezione' (CS) at 5.95% in the fully irrigated treatment and the lowest mean CBD content was found in the French cultivar 'Santhica 27', at 0.003% in the limited irrigation treatment.

Although THC and CBD exhibited similar patterns of genetic control, cannabichromene (CBC) showed a different pattern. Genotype explained 50% of trait variance and GEI explained 17%. This cannabinoid did not have a detectable influence of environment. However, there was also a much larger error variance for this trait of 33%, indicating that other factors contributed to variance in this trait that were not included in our model. The levels of this compound were lower than the other cannabinoids but showed a clear quantitative range. On average, cultivars in the limited irrigation treatment had slightly higher levels at 0.0052% of dry flower weight. The fully irrigated samples had mean CBC content of 0.0045% by weight. However, the range of CBC content was quite similar between the two treatments, from 0.0003% to 0.0133% under drought stress and from 0.0001% to 0.0112% under full irrigation. Despite the very similar ranges, rankings of the individual cultivars changed considerably which contributed to the significant GEI that were observed.

Terpenes

Of the 23 terpenes found in measurable amounts in the samples, only two showed evidence of significant GEI, α-pinene (p-value = 0.007) and β-pinene (p-value = 0.049). Although α-pinene and β-pinene levels were largely influenced by genetic effects (54% and 50% variance explained), there was still a significant effect of GEI (12% of trait variance for both terpenes). The genetic control of these traits was quite similar overall, but there was a small and significant (p-value = 0.011) effect of environment for α-pinene (3%) that was not detected for β-pinene. The observed GEI in α-pinene and β-pinene were particularly dramatic, as can be seen in a reaction norm of α-pinene content (Figure 2.3). Levels of these compounds did not show any clear trend, increasing or decreasing to varying degrees based on a particular genotype's

response to each environment (Table 2.1). Although the levels appear to be quite low when described as percent mass, there was measurable quantitative variation, particularly when viewed on a ppm scale. We chose to report these values as percent mass for uniformity and ease of comparison with cannabinoid measurements.

Carbon Isotope Ratio

Although cannabinoid and terpene production appeared to be largely products of genetic effects, carbon isotope ratio (δ13C) exhibited a different pattern. The variance in this trait showed significant effects of genotype, environment, and GEI, with the largest proportion attributed to environmental effects at 46%. A substantial 29% of variance is attributed to GEI effects, while 24% is due to genotype. Mean δ13C was -24.23 μg mg^{-1} under limited irrigation and -26.44 μg mg^{-1} under full irrigation, with a higher degree of isotope discrimination under full irrigation.

Correlations

To reduce estimates of correlation due to plasticity, Pearson's product moment correlations were calculated using cultivar means from each environment. For physical traits, this included data from all three of the environments measured in 2016. Since the biochemical traits were only assessed in the differential irrigation experiment at Fort Collins in 2016, those correlations reflect data from two environments rather than three. Correlation coefficients and significance levels for each trait are reported (Table 2.2) and significant relationships are represented as a heat map to visualize strength and directionality (Figure 2.4).

Grain yield was statistically significantly correlated with four of the traits analyzed: total plant biomass, plant height, days to maturity, and CBD. Yield was strongly and positively correlated with biomass ($r = 0.62$, p-value = 6.18×10^{-5}). Plant height was also positively correlated with grain yield, but to a lesser extent than biomass. The r-value for this relationship was 0.33, with a narrowly significant p-value of 0.046. These results support that larger plants yielded more grain, with biomass being a stronger indicator of grain yield than height.

Days to maturity was also correlated to grain yield, with a negative r-value of -0.53 (p-value = 0.007). This indicates that, in general, grain yield was reduced in cultivars that took longer to mature. However, this was not an entirely linear relationship. A bivariate scatterplot of grain yield versus days to maturity (Figure 2.5) shows that late maturing phenotypes led to decreased yield, while the highest yields were obtained from moderately early flowering cultivars.

The only biochemical trait that shared a genetic correlation with grain yield exhibited a negative relationship. The correlation between grain yield and CBD, with $r = -0.51$and p-value = 0.009, showed that levels of this cannabinoid tended to decrease when grain yield increased.

Total plant biomass was correlated with three other traits, in addition to grain yield. It was expected that taller plants would accumulate more biomass and the data supported that expectation. The correlation between biomass and plant height had an r-value of 0.58 and a p-value of 2.48×10^{-4}. Stand establishment was also positively correlated with biomass ($r = 0.63$, p-value = 5.29×10^{-5}). Plots that had more plants emerge and survive produced more biomass. Another correlation that confirmed an expected relationship was between biomass and carbon isotope ratio. A correlation of $r = 0.63$ and p-value = 4.66×10^{-4} shows that as the carbon isotope ratio grows, more biomass is produced. This was also reflected in the correlations between carbon isotope ratio and the other indicators of plant growth, plant height and stem diameter.

Positive correlations between carbon isotope ratio with plant height and stem diameter ($r = 0.74$, p-value = 1.71×10^{-5} and $r = 0.46$, p-value = 0.02, respectively) show that a less conservative pattern of water use (higher δ13C) is associated with more vigorous plant growth and mass.

Phenology, in particular, affected an entire suite of traits. Days to maturity was significantly correlated with grain yield, plant height, stem diameter, CBD, α-pinene, and β-pinene. With the exception of grain yield, which was previously reported, all of these traits were significantly positively correlated with days to maturity (Table 2.2).

All cannabinoid and terpene traits were positively correlated with one another, with the exception of CBC which appeared to be independent from relationships with any trait. Overall, increasing cannabinoid levels corresponded to increasing terpene levels and vice versa. The strongest of these observed relationships was between α-pinene and β-pinene with $r = 0.91$ and p-value = 1.53×10^{-10}.

DISCUSSION

Grain yield and biomass rankings were much more similar between the two treatments at Fort Collins than they were at Yellow Jacket, indicating that soil moisture alone does not account for the GEI that was observed with these cultivars. In general, increased precipitation led to higher yields, but genotypes did not respond in a uniform manner. A wider range of test locations and years of testing these same cultivars would help to elucidate the factors responsible for affecting grain yield and other important traits. These could include a number of direct environmental factors, such as latitude, temperature, elevation, humidity, and soil type, or management practices like fertilizer rates, planting dates, planting density, etc. Hemp cropping is

amenable to a wide range of target environments, but each environment will have different management needs and will maximize productivity using locally adapted germplasm.

An optimal phenology was apparent, with moderately early cultivars maturing around 110 days after sowing yielding the most grain. This agrees with a recent study by Long et al. (2017) that determined that corn cultivars that mature between 107 and 118 days will yield the highest amount of grain at latitudes between 35 and 40. Developing cultivars with adapted phenology should be relatively straightforward since this trait shows a strong genetic component and little discernable environmental effects. However, the environmental effect on phenology may become more pronounced with larger differences in latitude due to variable photoperiod sensitivity among genotypes (Amaducci et al., 2008). Grain yield was also positively correlated with biomass and, to a lesser extent, plant height. This implies that higher biomass is a stronger indicator of grain yield than height alone. In fact, the tall, late-flowering fiber cultivars fared the worst by far for grain yield and did not produce more biomass than earlier maturing cultivars. This may, at least partially, be driven by the fact that the fiber cultivars in these trials were all dioecious and only half of the plants produced grain. However, there is a distinction to be drawn between total plant biomass and stem biomass. For fiber applications, it may make sense to select a cultivar that produces less grain, flower, and leaf biomass, and select cultivars based on stem biomass. It should be noted that, for a complex trait like yield, a single season of data is not sufficient to draw generalized conclusions and more research should be performed to fully understand the relationships between grain yield, yield components, and other traits.

Stand establishment was strongly and positively correlated with biomass, which is in line with expectations and indicates that we did not surpass ideal planting density in this experiment. The lack of a direct significant correlation with grain yield is consistent with recent results by

Tang et al. (2017) that plant density, between 30 and 240 plants m^{-2}, did not significantly affect grain yield. It is possible that yield may have been more negatively impacted by low plant density in the absence of aggressive weed control. Keeping weed pressure artificially low allowed plots with lower emergence to fill in gaps and form consistent canopies. In an agricultural setting where mechanical control of weeds is not feasible, interspecies competition may hinder the crop's ability to form complete canopies. While the composite trait of stand establishment does reflect seeds that were planted that made it to maturity and has some basis in germination rates, it does not fully explain the roles of seedling mortality or self-thinning. Despite these values all being under our target of 50%, it was noted during the season that some plots appeared to be overly dense, exhibiting interplant competition and self-thinning. Distribution of plants throughout the plots, which was not measured, may play as strong a role as stand establishment in determining yield potential. Previous studies have shown that planting density and self-thinning can significantly affect yields and that different optima may be chosen based on maximizing stem yields versus flower or grain yields (van der Werf et al., 1995; Amaducci et al., 2002; Campiglia et al., 2017). These studies were also consistent with our results that cultivar choice interacts with these factors substantially and does not lead to a uniform seeding rate.

Water use efficiency (WUE) is another important aspect of crop performance and management. Carbon isotope ratio has shown to be a reliable proxy for WUE, where δ13C and WUE are inversely correlated (Donovan and Ehleringer, 1994). The δ13C measurements showed that, although a large proportion of trait variance was a product of environmental effects, there are substantial effects of genotype and GEI that show certain cultivars have a more efficient pattern of water use than others. Twenty-nine percent of trait variation for δ13C was due to GEI,

which shows substantially different responses to increased irrigation/precipitation. Differential access to water led to crossover GEI, with some cultivars showing a more conservative pattern of water use under full irrigation compared to drought conditions and other cultivars showing less conservative water use in the same conditions. While our data showed that less conservative water use was correlated with increased plant height, δ13C was not significantly correlated to grain yield. It is not possible to draw generalizations about δ13C /WUE of hemp in general with the small data set used for this particular analysis, however, our results support the idea that yield *per se* is a better measurement of drought tolerance than δ13C alone. Full characterization of this trait and its implications should be explored more extensively in the future to assist in the development of drought tolerant cultivars, with an emphasis on genotype specific responses to water limitation.

The data on THC content supported that the trait is largely controlled by genetic factors, but there was a significant effect of environment and levels were slightly higher under full irrigation on average. Due to the legal threshold of 0.3% THC content in hemp this single trait can make or break a farmer's season, so it is very important to understand its interaction with environmental factors. Although the small sample size and low levels of THC in these samples does not allow for a highly precise estimation of variance components, these data support that cultivar selection is an important factor in mitigating risk for farmers. Despite the fact that many of these cultivars were grown in environments drastically different from where they were developed, the vast majority tested below the legal threshold for THC. This is a testament to the breeders that developed the seed, as well as some assurance that this trait can be relatively stable. However, environmental factors outside of soil moisture may influence the expression of this trait, which could be problematic and certainly warrants further investigation.

Interest in farming hemp for CBD is growing and there is a positive outlook for that as well. CBD showed a positive correlation with later flowering, but 83% of variance in CBD content was a result of genetic effects. Again, this makes cultivar selection by far the most important decision when seeking a specific CBD content. Increasing irrigation did slightly increase CBD content, which seems in agreement with the positive correlation between CBD content and rainfall recently found by Calzolari et al. (2017), but it was a small enough change that flower biomass yield is likely far more important to overall yield than a slight change in CBD percentage. One caveat about CBD cropping is that none of the cultivars tested in these trials were considered "high CBD" cultivars. Much of the germplasm in Colorado that is being marketed as such has been developed by crossing drug-type cannabis with industrial hemp and is anecdotally known to be less stable for THC content. It was observed in this study that THC content was positively correlated with CBD content. This could be problematic when breeding to increase CBD content. Although the genetics of THCA versus CBDA production have begun to be elucidated, cannabinoid content is thought to be a complex, quantitative trait (Weiblen et al., 2015). Until the genetic mechanisms controlling cannabinoid content are better characterized, it remains a risk to utilize seed developed for high cannabinoid production unless rigorous testing has proven stability of THC levels for a particular cultivar in its target environment(s).

In addition to non-psychoactive cannabinoids, terpenes are also receiving new attention as a value-added byproduct of industrial hemp. These smell and flavor compounds have a wide range of uses, including natural pesticides, aromatherapy, brewing, and as therapeutic agents in medicine (Russo, 2011). An interesting result is that the strongest correlation observed among the biochemical traits was between α-pinene and β-pinene. These compounds are enzymatically converted from the same biochemical precursor, geranyl pyrophosphate (Croteau et al., 1989),

but do not appear to be competitively exclusive. Despite increasing interest in hemp terpenes, breeding for specific terpene profiles will add in a distinct element of complexity with its polygenic nature and high degree of GEI (Booth et al., 2017). But breeding for certain terpenes may dovetail nicely with other goals like breeding for pest resistance. With an emerging awareness of these compounds and a financial incentive for their production, it is very likely that customizing terpene profiles will become a more common goal in hemp breeding programs.

The current study represents a first attempt at characterizing genetic, environmental, and GEI effects caused by water limitation in hemp, as well as the first GEI study to be performed on hemp in the United States. Previous hemp studies of GEI have focused largely on fiber characteristics, agronomic management, and phenology (Struik et al., 2000; Tang et al., 2016). Although we recognize that logistical restraints surrounding acquisition of germplasm and running multiple testing sites hindered our ability fully characterize GEI in complex traits, this experiment contributes information to an understudied area to encourage academic discussion of an important topic. The potential of hemp to move beyond its historic role as a fiber crop is beginning to be realized and information about environmental impacts on traits with legal or medical implications is essential for consistent field production of this unique crop. These data offer some insight into which types of cultivars will perform well locally and, more importantly, contribute to a broader understanding of the plasticity and stability of both physical and biochemical traits in hemp. The substantial level and different types of GEI observed in this population tells a complicated story in which certain cultivars are more plastic than others, but, also, traits themselves exhibit a wide range of plasticity. As legalization of hemp continues to expand cultivation into new areas, more information about the impact of environmental factors on end-use traits will become available. These types of data will be germane to breeding

programs and will help in developing a host of highly adapted cultivars with improved stability and uniformity.

REFERENCES

Amaducci, Stefano, Michele Colauzzi, Gianni Bellocchi, and Gianpietro Venturi. 2008. Modelling post-emergent hemp phenology (*Cannabis sativa L.*): theory and evaluation. European Journal of Agronomy 28.2: 90-102.

Amaducci, S., Scordia, D., Liu, F.H., Zhang, Q., Guo, H., Testa, G., Cosentino S.L., 2015. Key cultivation techniques for hemp in Europe and China. Ind. Crops Prod., 68, 2-16.

Amaducci, S., Errani, M., Venturi, G., 2002. Plant population effects on fibre hemp morphology and production. Journal of industrial hemp, Vol. 7 (2), 33-60.

Amaducci, S. and H.J. Gusovious, 2010. Hemp - Cultivation, Extraction and Processing. In: Industrial Applications of Natural Fibres: Structure, Properties and Technical Applications (ed J. Müssig), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9780470660324.ch5

Booth, Judith K., Jonathan E. Page, and Jörg Bohlmann. 2017. Terpene synthases from *Cannabis sativa*. PloS one 12.3: e0173911.

Calzolari, D., Magagnini G., Lucini, L., Grassi, G., Appendino, G., Amaducci, S., 2017. High Added-Value Compounds from Cannabis threshing residues. Ind. Crops Prod, 198, 558-563.

Campiglia, Enio, Emanuele Radicetti, and Roberto Mancinelli. 2017. Plant density and nitrogen fertilization affect agronomic performance of industrial hemp (*Cannabis sativa L.)* in Mediterranean environment. Industrial crops and products 100: 246-254.

Croteau, Rodney, D. Michael Satterwhite, Carl J. Wheeler, and N. Mark Felton. 1989. Biosynthesis of monoterpenes. Stereochemistry of the enzymatic cyclizations of geranyl pyrophosphate to (+)-alpha-pinene and (-)-beta-pinene. Journal of Biological Chemistry 264.4: 2075-2080.

Donovan, Lisa A., and James R. Ehleringer. 1994. Carbon isotope discrimination, water-use efficiency, growth, and mortality in a natural shrub population. Oecologia 100.3: 347-354.

Ehleringer, J. R., and C. Osmond. 1989. Stable isotopes. In: RW Pearcy, JR Ehleringer, HA Mooney and PW Rundel, editors, Plant Physiological Ecology: Field Methods and Instrumentation. Kluwer Academic Publishers, Boston, MA. p. 281-300.

Gallily, R., Yekhtin, Z. and Hanuš, L.O., 2015. Overcoming the bell-shaped dose-response of cannabidiol by using cannabis extract enriched in cannabidiol. *Pharmacol Pharm*, *6*(2), pp.75-85.

Long, Nguyen V., Yared Assefa, Rai Schwalbert, Ignacio Ciampitti. 2017. Maize Yield and Planting Date Relationship: A Synthesis-Analysis for US High-Yielding Contest-Winner and Field Research Data. Frontiers in plant science 8: 2106.

Makahiya, Hazel Anne F., Kim Nyka C. Perdiguerra, Ivy Rose M. Marajas, Ma. Lourdes S. Edaño, Jose E. Hernandez, and Pompe C. Sta Cruz. 2017. Analysis of genotype by environment interaction in irrigated lowland rice (Oryza sativa L.) varieties under diverse agroclimatic environments. Philippine Journal of Crop Science 42.2: 28-36.

Malosetti, Marcos, Jean-Marcel Ribaut, and Fred A. van Eeuwijk. 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Frontiers in physiology 4: 44.

de Mendiburu, Felipe. 2015. agricolae: Statistical Procedures for Agricultural Research. R package version 1.2-3. http://CRAN.R-project.org/package=agricolae

Peterson, Ryan A. 2017. bestNormalize: A suite of normalizing transformations. R package version 3.4.1. https://github.com/petersonR/bestNormalize

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Russo, Ethan B. 2011. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. British journal of pharmacology 163.7: 1344-1364.

Salentijn, Elma MJ, Qingying Zhang, Stefano Amaducci, Ming Yang, and Luisa M. Trindade. 2015. New developments in fiber hemp (*Cannabis sativa L.*) breeding. Industrial Crops and Products 68: 32-41.

Schluttenhofer, Craig, and Ling Yuan. 2017. Challenges towards Revitalizing Hemp: A Multifaceted Crop. Trends in plant science.

Small, Ernest, and David Marcus. 2003. Tetrahydrocannabinol levels in hemp (*Cannabis* sativa) germplasm resources. Economic botany 57.4: 545-558.

Struik, P. C., S. Amaducci, M.J. Bullard, N.C. Stutterheim, G. Venturi, and H.T.H Cromack. 2000. Agronomy of fibre hemp (*Cannabis sativa L.*) in Europe." Industrial crops and products 11.2-3: 107-118.

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

Tang, K., Struik, P.C., Yin, X., Calzolari, D., Musio, S., Thouminot, C., Bjelková, M., Stramkale, V., Amaducci, S., 2017. A comprehensive study of planting density and nitrogen fertilization effect on dual-purpose hemp (*Cannabis sativa L.*) cultivation. Industrial Crops and Products 107: 427-438

U.S. Govt. Print. 2014 (enacted). Farm Bill – Section 7606. Print.

Van der Werf, Hayo MG, M. Wijlhuizen, and J. A. A. De Schutter. 1995. Plant density and self-thinning affect yield and quality of fibre hemp (*Cannabis sativa L.*). Field Crops Research 40.3: 153-164.

Wei, Taiyun. 2013. corrplot: Visualization of a correlation matrix. R package version 0.73. http://CRAN.R-project.org/package=corrplot

Weiblen, George D., Jonathan P. Wenger, Kathleen J. Craft, Mahmoud A. ElSohly, Zlatko Mehmedic, et al. 2015. Gene duplication and divergence affecting drug content in *Cannabis sativa*. New Phytologist 208.4: 1241-1250.

Wickham, Hadley. 2017. tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. http://CRAN.R-project.org/package=tidyverse

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

FIGURES

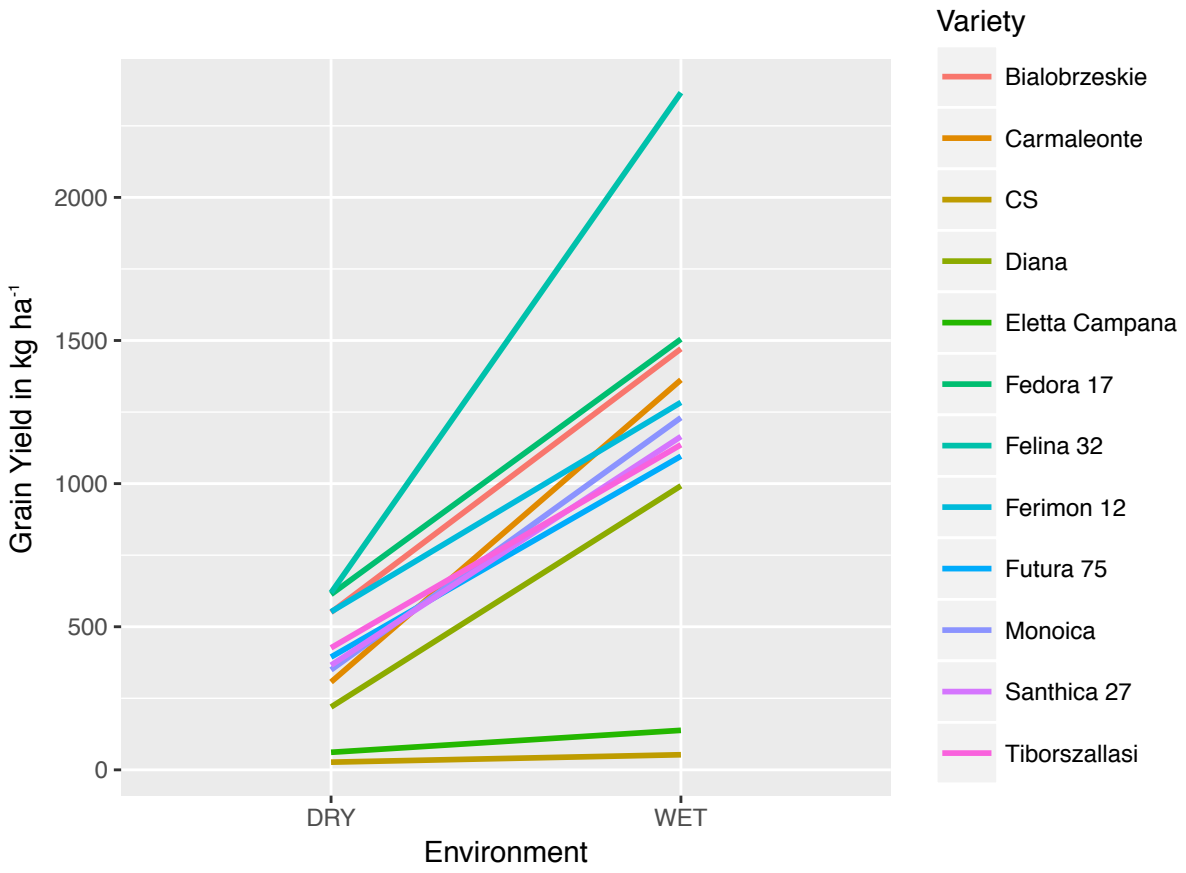


Figure 2.1. Reaction norm showing Grain Yield plasticity of industrial hemp cultivars under limited and full irrigation at Fort Collins, CO. All of the cultivars exhibited higher yield with increased irrigation, but with changing ranks, a profile that is characteristic of divergent genotype-by-environment interactions.

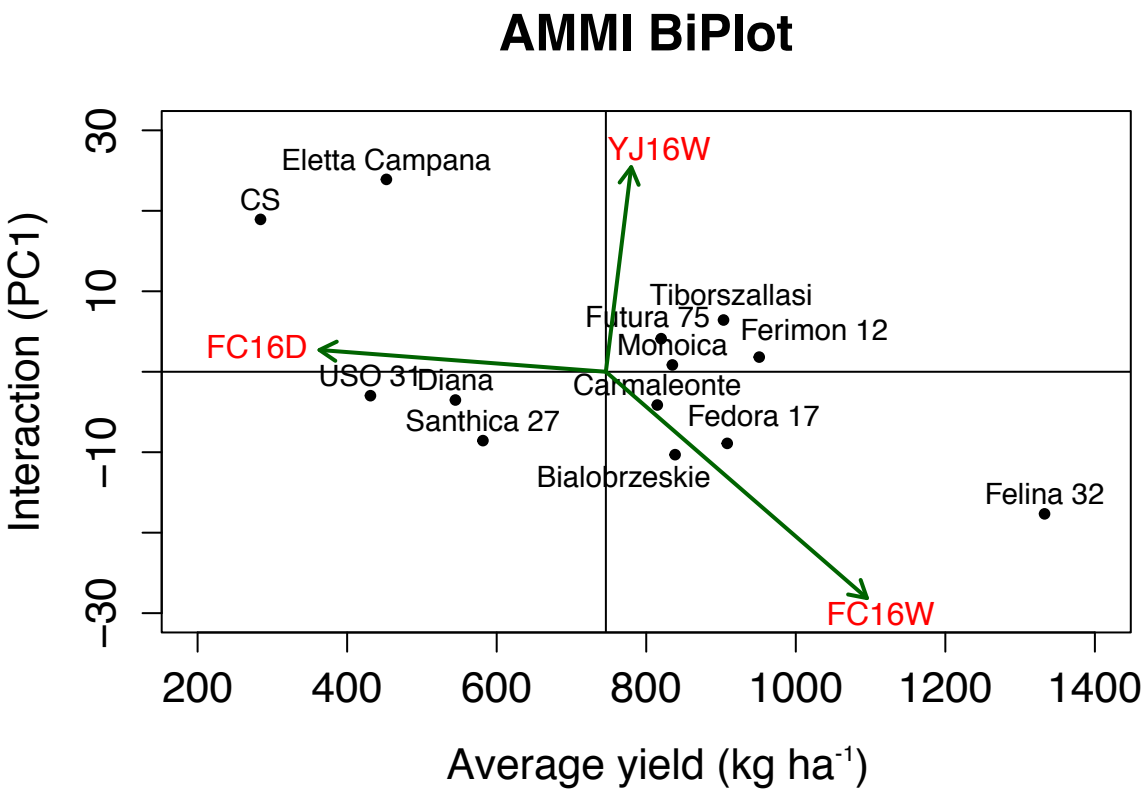


Figure 2.2: Additive Main Effect and Multiplicative Interation (AMMI) biplot, showing genotype-by-environment interactions for grain yield in three environments. The direction of the environmental vectors shows whether the cultivars performed above or below trial averages in those environments. FC16D = Fort Collins, CO, 2016, limited irrigation (Dry), FC16W = Fort Collins, CO, 2016, full irrigation (Wet), YJ16W = Yellow Jacket, CO, 2016, full irrigation (Wet).

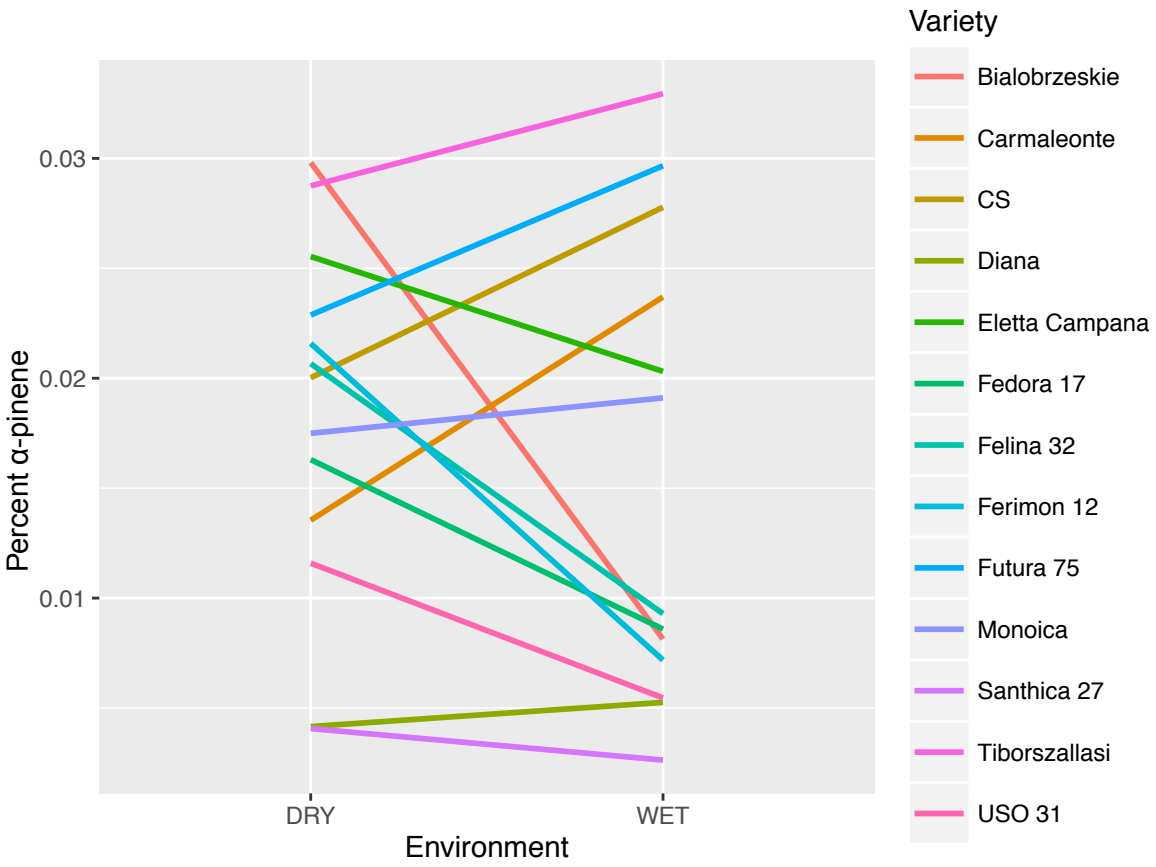


Figure 2.3. Reaction norm showing α-pinene content plasticity of industrial hemp cultivars under limited and full irrigation at Fort Collins, CO. The extreme rank changes are characteristic of crossover genotype-by-environment interactions.

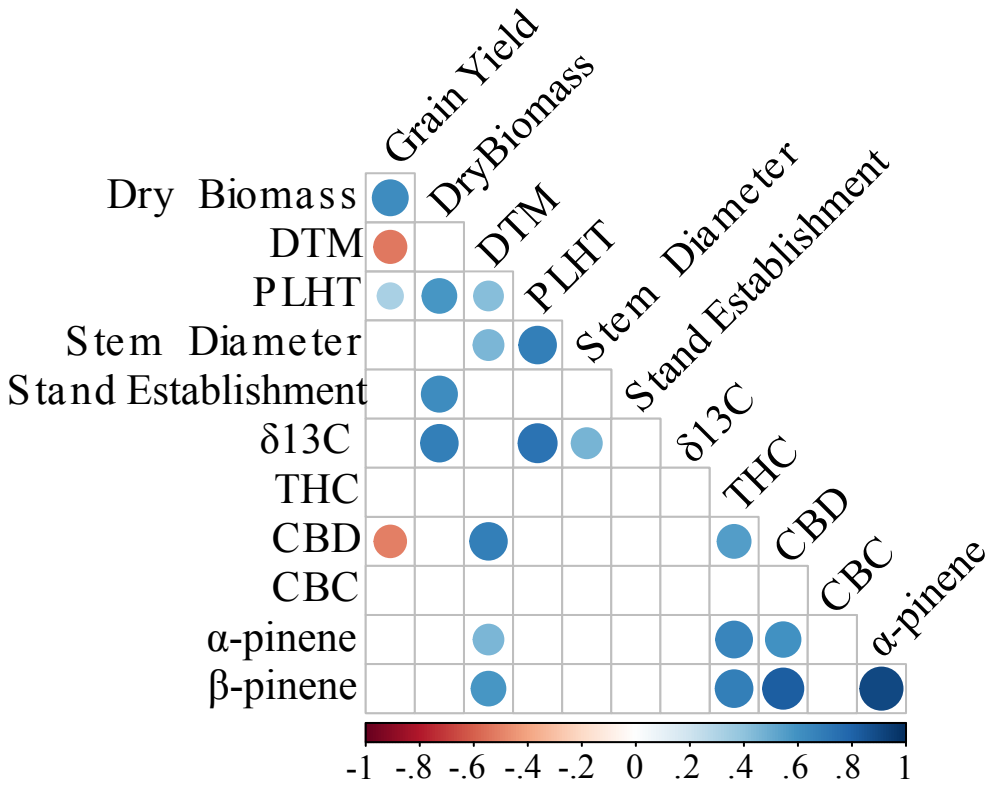


Figure 2.4. Trait correlation heat map of significant trait correlations at Fort Collins, CO and Yellow Jacket, CO in 2016. Only statistically significant correlations are shown, with the size and color of the circles representing the strength and direction of the correlations.
Trait Abbreviations: DTM=Days to Maturity, PLHT=Plant height, Stand Est=Stand Establishment, δ13C= Carbon isotope ratio, THC=Total Tetrahydrocannabinol, CBD=Total Cannabidiol, CBC=Cannabichromene

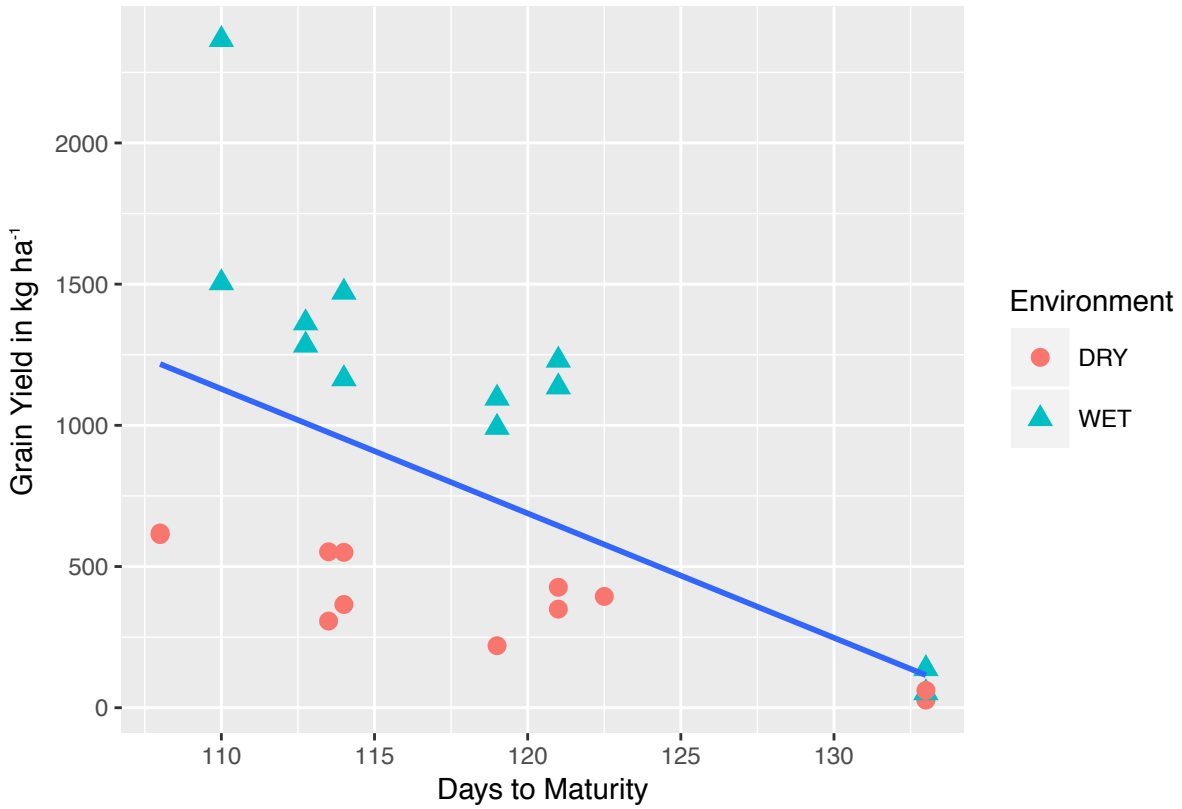


Figure 2.5. Grain Yield Response to Days to Maturity at Fort Collins, 2016. A best fit line is drawn on a scatterplot comparing maturity dates with grain yield data. The color-coded data points demonstrate that the relationship between phenology and yield is consistent due to maturity being relatively unaffected by environmental factors. DRY = Fort Collins, 2016 under limited irrigation, WET = Fort Collins, 2016 under full irrigation.

TABLES

Table 2.1: ANOVA and Summary Statistics at Fort Collins, CO and Yellow Jacket, CO in 2016.

	Percent Variance Explained			Significance		
Trait	**Genotype**	**Environment**	**G x E**	**Genotype**	**Environment**	**GEI**
Grain Yield (kg ha^{-1})	28.4	24.8	34.6	***	***	***
Dry Biomass (kg ha^{-1})	11.5	61.5	12.9	***	***	***
DTM (Days)	97.3	0.0	1.0	***	NS	***
PLHT (cm)	35.9	37.8	9.0	***	***	***
Stem Diameter (mm)	36.1	8.3	14.9	***	***	*
Stand Establishment (%)	24.2	51.1	6.6	***	***	*
δ13C (µg mg^{-1})	23.9	46.4	28.9	***	***	***
Total THC (%)	80.4	1.7	2.1	***	**	NS
Total CBD (%)	82.7	6.0	1.7	***	***	NS
CBC (%)	50.0	0.0	16.8	***	NS	*
α-pinene (%)	54.2	2.7	12.2	***	*	**
β-pinene (%)	50.3	0.2	12.0	***	NS	*

DTM=Days to maturity, PLHT=Plant Height at maturity, δ13C=carbon isotope ratio, THC=total potential Δ9-tetrahydrocannabinol, CBD=total potential cannabidiol, CBC=cannabichromene

p-values:*= <.05 **=<.01 ***=<.001

Table 2.2: Trait correlations (r-values) at Fort Collins, CO and Yellow Jacket, CO in 2016.

Trait	Seed Yield	Biomass	DTM	PLHT	Stem Diameter	Stand Est.	δ13C	THC	CBD	CBC	α-pinene
Seed Yield	-										
Biomass	0.62 ***	-									
DTM	-0.53 **	-0.15 NS	-								
PLHT	0.33 *	0.58 ***	0.42 *	-							
Stem Diameter	0.05 NS	0.29 NS	0.46 *	0.68 ***	-						
Stand Est.	0.29 NS	0.63 ***	-0.20 NS	-0.07 NS	-0.17 NS	-					
δ13C	0.35 NS	0.68 ***	0.19 NS	0.74 ***	0.46 *	0.28 NS	-				
THC	-0.20 NS	-0.13 NS	0.38 NS	0.05 NS	-0.11 NS	0.12 NS	0.04 NS	-			
CBD	-0.51 **	-0.28 NS	0.68 ***	0.01 NS	0.02 NS	-0.02 NS	0.00 NS	0.55 **	-		
CBC	0.26 NS	0.35 NS	-0.32 NS	-0.05 NS	-0.25 NS	0.18 NS	0.11 NS	-0.22 NS	-0.15 NS	-	
α-pinene	-0.01 NS	0.16 NS	0.45 *	0.26 NS	0.00 NS	0.33 NS	0.30 NS	0.66 ***	0.61 ***	-0.04 NS	-
β-pinene	-0.25 NS	-0.02 NS	0.58 **	0.24 NS	0.09 NS	0.15 NS	0.24 NS	0.69 ***	0.82 ***	-0.12 NS	0.91 ***

"NS" non significant,* Significant at the 0.05 probability level; ** Significant at the 0.01 probability level; *** Significant at the 0.001 probability level
Trait Abbreviations: DTM=Days to Maturity, PLHT=Plant height, Stand Est=Stand Establishment, δ13C= Carbon isotope ratio, THC=Total Tetrahydrocannabinol, CBD=Total Cannabidiol, CBC=Cannabichromene

CHAPTER 3:

DIVERSITY, DIVERGENCE, AND HETEROZYGOSITY IN *CANNABIS SATIVA*

SUMMARY

The usage and regulation of *Cannabis sativa* has led humans to classify two distinct subgroups within the species, hemp and high-THC *Cannabis* (often referred to as marijuana). While hemp has a legal definition that distinguishes it by its low levels of THC, whether or not there is a genetic basis for differentiation of these two groups remains debated. We used a genotype-by-sequencing (GBS) approach to analyze samples of both industrial hemp cultivars and high-THC *Cannabis* to better understand population structure, assess allelic diversity of these groups, compare heterozygosity levels within and between groups, and identify rare and common alleles of the two groups. Our findings show there is weak support for hemp and high-THC *Cannabis* being distinct groups using genome-wide F_{ST} and phylogenetic analyses. Average heterozygosity levels were higher in the hemp group, however, there were more outliers in the high-THC *Cannabis* group which exhibited more extreme minimum and maximum heterozygosity. Overall, both groups shared large amounts of common variation with a small number of loci where alleles were fixed in one group and variable in the other. These findings indicate that further exploration is warranted in determining which genes are driving the population structure to elucidate the particular mechanisms responsible for distinguishing the two groups.

INTRODUCTION

Cannabis (*Cannabis sativa L.)* has a long and storied history of interaction with humans, with documented use dating back 10,000 years (Laursen, 2015). Due to extensive human vectoring and gene flow between types of *Cannabis*, evolutionary origins of the species, as well as speciation itself, are highly difficult to discern and are the subject of academic debate (Small and Cronquist, 1976; Hillig and Mahlberg, 2004; Hillig, 2005; Watts, 2006; Merlin and Clarke, 2013; etc.). Additionally, the fluctuating legal status of *Cannabis* has made research surrounding the plant more restrictive than with other crops. On top of this, illicit movement and unknown provenance and pedigrees of a large number of uniquely named public sector Cannabis genotypes makes predictions of relatedness difficult. Finally, prohibition is likely to have caused bottlenecks in the domesticated lineages of Cannabis. Despite these challenges, the advent of affordable molecular tools has enabled new methods to explore *Cannabis*' genomic composition and history. Genetic diversity studies of *Cannabis* are still in their infancy but have begun to contribute information to describe diversity in the species beyond the range of observable phenotypes.

To date, several studies have attempted to classify genetic diversity present in *Cannabis* using molecular tools (Gao et al., 2014; Zhang et al., 2014; Sawler et al., 2015; Lynch et al., 2016; Hu et al., 2016; Soorni et al., 2017). An initial study by Gao et al. (2014) genotyped 115 hemp varieties (*Cannabis sativa L.* with low tetrahydrocannabinol (THC) content) using simple-sequence repeat (SSR) markers. An interesting result from this study showed groups exhibiting higher genetic similarity separating into clusters that reflected a gradient of latitude and day length. Although this clustering is likely a result of selection for local adaptation, it represents the first study to show a genetic basis for subgroups within the species. Similarly, Zhang et al.

(2014) genotyped 27 native Chinese hemp accessions with inter-simple-sequence repeat (ISSR) markers and detected a latitudinal gradient. Sawler et al. (2015) and Lynch et al. (2016) both attempted to compare allelic frequencies across the culturally and legally distinguished subgroups of hemp and high-THC *Cannabis* (*Cannabis sativa L.* with THC content greater than 0.3%, also referred to as marijuana), as well as the extensively colloquially used high-THC *Cannabis* subgroups of "indica" and "sativa". Sawler et al. (2015) used a genotype-by-sequencing approach to detect single-nucleotide polymorphisms (SNP), while Lynch et al. (2016) used a combination of whole-genome sequence (WGS) data and GBS data. Both groups were able to distinguish three subgroups from their population structure analyses, supporting the hemp/indica/sativa distinction within the species. However, the two studies disagreed on whether hemp was more related to "indica" or "sativa" high-THC *Cannabis* and whether hemp or high-THC *Cannabis* harbored more inherent heterozygosity. Extensive overlap and the attempt to create dichotomous categories from the quantitative traits used to distinguish the putative "indica" and "sativa" subgroups further complicate the issue.

Many questions are still unanswered regarding the extant diversity and heterozygosity present in the species. One of the major challenges of these types of studies is the decentralized nature of *Cannabis* germplasm. While all major crops have germplasm banks dedicated to preserving genetic diversity, *Cannabis*, due to its history of prohibition and displacement by modern fiber crops, has only a handful of disconnected germplasm repositories around the world with accession numbers generally in the realm of dozens rather than thousands (Small and Marcus, 2003, Ranalli, 2004). In light of this challenge, it is unlikely that a single study will be able to provide a comprehensive picture of the genetic diversity present in the entire species. Rather, independent studies will contribute information to the growing body of literature to form

an emergent consensus over time. An important part of characterizing the extant diversity is identifying where in the genome hemp and high-THC groups differ, and which specific genomic regions play a role in differentiating any such groups. The purpose of the current study is to compare allelic diversity within samples from the legally defined groups of hemp and high-THC *Cannabis* and identify common and rare alleles in these groups. In addition, we explore whether there is a genetic basis for this legal differentiation, as well as report heterozygosity/homozygosity levels within a diverse collection.

MATERIALS AND METHODS

Plant Materials and Data Collection

Leaf samples were collected from two different sets of germplasm for these analyses. Dr. Stefano Amaducci (Università Cattolica del Sacro Cuore, Milano, Lombardy, Italy) provided 17 varieties of industrial hemp for field trials in 2015 and leaves were collected during the vegetative growth stage and lyophilized (Campbell et al., 2019). High-THC *Cannabis* leaf samples were collected from 60 unique strains, donated by vendors at the 2014 Emerald Cup (Santa Rosa, CA). DNA was extracted from both sets of leaves using a Qiagen DNEasy kit (Qiagen Inc., Valencia, CA). Genotype-by-Sequencing (GBS) was performed on the extracted DNA using Illumina (Illumina Inc., San Diego, CA) sequencing, identifying approximately 800,000 single-nucleotide polymorphisms (SNPs) at 10x coverage. SAMTools (Li et al., 2009) was used to call single-nucleotide polymorphisms (SNPs), generating approximately 800,000 SNPs for these data.

Statistical Analysis

Population parameters including population differentiation, F-statistics, linkage disequilibrium (LD) and a folded joint site frequency spectrum, as well as heterozygosity levels for individual samples, were computed using the scikit-allel package for Python developed by Miles and Harding (2017). Scaffolds identified in the joint site frequency spectrum that contained alleles in high frequency in each group were compared to known gene sequences in the National Center for Biotechnology Information database, using the Nucleotide BLAST tool (NCBI BLAST). We extracted scaffolds of interest and used the BLASTN algorithm to query them against the complete NCBI nucleotide database. We filtered the resulting alignments to remove those where the percent sequence identity was less than 75%, and where the proportion of the subject sequence covered by the alignment was less than 0.75. When scaffolds had multiple hits within the same region, we chose the alignment that maximized percent identity and alignment length. This resulted in four scaffolds with similarity to sequences in the nucleotide database, two of which had multiple regions with significant alignments.

Values for the F_{ST} analysis were calculated using methods developed by Hudson (1992) and refined by Bhatia et al. (2013) to correct for the influence of rare variants. A T-distributed Stochastic Neighbor Embedding (t-SNE) analysis was used to visualize population structure comparing monoecious hemp, dioecious hemp, and high-THC *Cannabis*-type samples using the scikit-learn package for Python (Van Der Maaten and Hinton 2008, Pedregosa et al., 2011). As population structure in the whole *Cannabis* population can significantly increase the extent of LD, we calculated pairwise LD in subpopulations of hemp and high-THC *Cannabis* individually. We used the method of Rogers and Huff (2008) to infer LD between loci with unknown phase.

RESULTS

Allelic Diversity

A folded joint site frequency spectrum is shown in Figure 3.1, comparing allelic frequencies in the two populations. The vast majority of alleles in the joint site frequency analysis are shared by both hemp and High-THC *Cannabis*. Most SNPs were in low frequency among both groups. We identified a small number of SNPs that were in high frequency in one group and low in the other (>0.9, <0.1). The sequences flanking these divergent SNPs were compared with other species sequences to determine if alleles for certain genes were fixed or nearly fixed in either population. A list of homologous gene sequences can be found in Table 3.1. Alleles for four common genes were found that had a high frequency in the hemp group and a low frequency in the high-THC *Cannabis* group that had sequence homology greater than 75%. Conversely, alleles that were more exclusive to the high-THC *Cannabis* group did not match any known sequences from the NCBI database other than a single gene that has been noted as a male-associated DNA sequence in *Cannabis*.

Population Structure

The results of the T-distributed Stochastic Neighbor Embedding (t-SNE) analysis (Figure 3.2) shows clustering of the hemp and high-THC *Cannabis* subpopulations, with dioecious hemp more similar to high-THC *Cannabis* than monecious hemp based on the SNP data despite some admixture. This result is supported in the phylogenetic neighbor joining tree (Figure 3.3), where the Chinese dioecious hemp varieties Mengma and Jianji are clustered with the hemp group but also related to the high-THC *Cannabis* group Northern Lights samples.

Heterozygosity Rates

Figure 3.4.1 shows the overall heterozygosity rates for the hemp and high-THC *Cannabis* subpopulations. The heterozygosity rates for hemp, based on unambiguous SNP calls, ranged from 11.7% to 19.4%, with a mean value of 15.0%. The high-THC *Cannabis* group exhibited lower overall heterozygosity, ranging from 8.2% to 20.2, with a mean value of 12.6%. Although there is overlap between these two groups, a t-test shows that the two groups differ significantly, with a p-value of $1.38e^{-5}$. Figure 3.4.2 shows the heterozygosity levels of the individual samples from the analysis. The high-THC *Cannabis* subpopulation has more extreme outliers than the hemp group, but lower overall levels of heterozygosity.

F_{ST} was calculated as a measure of genomic differentiation between the two subpopulations assayed. These results, shown in Figure 3.5, indicate that there is a small proportion of the genome that is highly differentiated in these two groups, while most of the genome contains polymorphisms that are present in both subpopulations. This supports the results from the joint-site frequency spectrum, where the vast majority of the hemp and high-THC *Cannabis* genomes shared common allelic variants. Linkage disequilibrium decay (shown in Supplemental Figures 3.1.1 and 3.1.2) was not found to be significantly different between the two groups.

DISCUSSION

The results of the joint-site frequency spectrum analysis and the F_{ST} analysis show that individuals in the hemp and high-THC *Cannabis* subpopulations mainly share common polymorphisms. There are some loci throughout the genome where differentiation occurs, and these loci serve as the basis of higher within group similarity. Although there is evidence of

population structure within the species, it is not possible to say whether or not that structure is rooted in evolutionary backgrounds or an artifact of breeding and artificial selection. This is evident in the neighbor-joining tree, where known fiber hemp varieties are as similar in identity-by-state to high-THC *Cannabis* samples as they are to other industrial hemp samples. This is not an entirely surprising finding, due to the fact that the two subpopulations are distinguished by phenotypes based on usage and legal thresholds and extensive admixture of the two groups has occurred historically and continues to occur.

If we adhere to viewing hemp and high-THC *Cannabis* as distinct subpopulations, an important question is how do the two groups actually differ? The two ways in which we explored this topic were to compare loci in which alleles in common with the Purple Kush reference genome were rare in one group or the other, as well as to compare heterozygosity levels of the samples between groups. Since the reference genome was created from a high-THC *Cannabis* sample and the data used in this study were generated using a GBS approach, not all of the diversity of the *Cannabis* genome is represented. However, we were able to identify a number of scaffolds that contained alleles nearly fixed in one population while occurring in very low frequency in the other. Alleles that were common in the hemp group and rare in the high-THC *Cannabis* group, shown in Table 3.1, were related to genes controlling standard plant processes such production of isomerase and helicase enzymes, and cytoplasmic tRNA. It is our hypothesis that these alleles were in high frequency in the hemp group either by chance or as an indirect effect of selection on unrelated traits. Expanding this assay to a wider group of samples would help to understand whether or not this differentiation is occurring by chance and warrants further study. The alleles that were in high frequency in the high-THC *Cannabis* group and in low frequency in the hemp group, however, were more ambiguous. Only one scaffold had variation

in homologous sequences, identifying a putative transcriptase gene and a male-associated DNA sequence from *Cannabis* itself. The large proportion of scaffolds that did not exhibit sequence homology with known genes is interesting. These alleles may be related to traits that distinguish high-THC *Cannabis* from hemp, but without further functional characterization it is impossible to know if this is the case or if these scaffolds contain relatively common genes that had alleles in common by chance in a small sample population or by selection on unrelated traits. It is also possible that extensive bottlenecking and limited sexual reproduction in high-THC *Cannabis* populations has led to the loss of variation in important genes in the high-THC *Cannabis* gene pool. Understanding the broader allelic states that are unique to one group or the other will help to establish the basis of differentiation, i.e. whether or not the groups have an evolutionary basis for distinction or whether the groups are solely distinguished by phenotypes that are a result of artificial selection and/or genetic drift. This is an area that deserves more attention, as it could have important implications in breeding improved *Cannabis* varieties despite the intended use.

Two previous studies have attempted to directly answer whether or not hemp or high-THC Cannabis have higher overall levels of heterozygosity (Sawler et al., 2015; Lynch et al., 2016). The two studies had conflicting results, with Sawler et al. (2015) concluding that hemp varieties are more heterozygous on average, while Lynch et al. (2016) concluded that high-THC Cannabis samples had higher average heterozygosity throughout the genome. There are logical arguments to be made as to why either group may be more heterozygous. Breeding methods have been quite different for the two groups, with a focus on maintaining open-pollinated populations in hemp and repeated poly-hybridization of distinctly bottlenecked populations in high-THC *Cannabis*. Both of these techniques can lead to maintenance or purging of diversity depending on effective population size and chance recombination. In our study, the hemp group was found

to have higher average heterozygosity, which was consistent with the findings of Soorni et al. (2017). However, despite having lower average heterozygosity, the high-THC *Cannabis* group had more outliers which created a wider distribution with lower minimum and higher maximum levels of heterozygosity. Further studies with larger samples and georeferenced landrace collections are needed to provide insight into the evolutionary history of Cannabis genetics and the resulting effects on the genome. The basis of these groups (hemp versus high-THC *Cannabis*) is based on a legal threshold in a polygenic trait, and not any known geographic or biological barriers to gene flow. It is our opinion that the early human vectoring of *Cannabis* germplasm (Hillig, 2005) and extensive admixture between the two groups will make it extremely difficult, if not impossible, to determine whether or not these two subpopulations of the species have distinct evolutionary origins.

REFERENCES

Bhatia, Gaurav, et al. 2013. "Estimating and interpreting FST: the impact of rare variants." *Genome research* gr-154831.

Campbell, Brian J., et al. 2019. "Genotype-by-Environment Interactions of Industrial Hemp (*Cannabis sativa L.*) Cultivars Highlight Diverse Responses to Environmental Factors" *Agrosystems, Geoscience & Environment* (in press).

Clarke, Robert C., and Mark D. Merlin. *Cannabis: evolution and ethnobotany*. Univ of California Press, 2013.

Hillig, Karl W., and Paul G. Mahlberg. 2004. "A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae)." *American Journal of Botany* 91.6: 966-975.

Hillig, Karl W. 2005. "Genetic evidence for speciation in *Cannabis* (Cannabaceae)." *Genetic Resources and Crop Evolution* 52.2: 161-180.

HU, Zun-Gong, et al. "Genetic diversity research of hemp (Cannabis sativa L) cultivar based on AFLP analysis." *Journal of Plant Genetic Resources* 13.4 (2012): 555-561.

Hudson R.R., Slatkin, M., Maddison, W.P. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–589.

Laursen, Lucas. 2015. "Botany: The cultivation of weed." *Nature* 525.7570: S4-S5.

Li, Heng, et al. "The sequence alignment/map format and SAMtools." *Bioinformatics* 25.16 (2009): 2078-2079.

Lynch, Ryan C., et al. "Genomic and chemical diversity in *Cannabis*." *Critical Reviews in Plant Sciences* 35.5-6 (2016): 349-363.

Miles, Alistair and Nick Harding. (2017, July 4). cggh/scikit-allel: v1.1.8 (Version v1.1.8). Zenodo. http://doi.org/10.5281/zenodo.822784

NCBI BLAST, http://www.ncbi.nlm.nih.gov/BLAST/

Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830.

Ranalli, Paolo. "Current status and future scenarios of hemp breeding." *Euphytica* 140.1 (2004): 121-131.

Rogers A.R., and C. Huff (2009) Linkage disequilibrium between loci with unknown phase. Genetics 182: 839–844

Sawler, Jason, et al. "The genetic structure of marijuana and hemp." *PloS one* 10.8 (2015): e0133292.

Small, Ernest, and Arthur Cronquist. "A practical and natural taxonomy for *Cannabis*." *Taxon* (1976): 405-435.

Small, Ernest, and David Marcus. "Tetrahydrocannabinol levels in hemp (*Cannabis* sativa) germplasm resources." *Economic botany*57.4 (2003): 545-558.

Soorni, Aboozar, et al. "Assessment of Genetic Diversity and Population Structure in Iranian Cannabis Germplasm." *Scientific reports* 7.1 (2017): 15668.

Van Der Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9:2579–2605.

Watts, Geoff. "Science commentary: *Cannabis* confusions." *BMJ: British Medical Journal* 332.7534 (2006): 175.

Zhang, L. G., et al. "Analysis of the genetic diversity of Chinese native Cannabis sativa cultivars by using ISSR and chromosome markers." *Genet. Mol. Res* 13.4 (2014): 490-500.

FIGURES

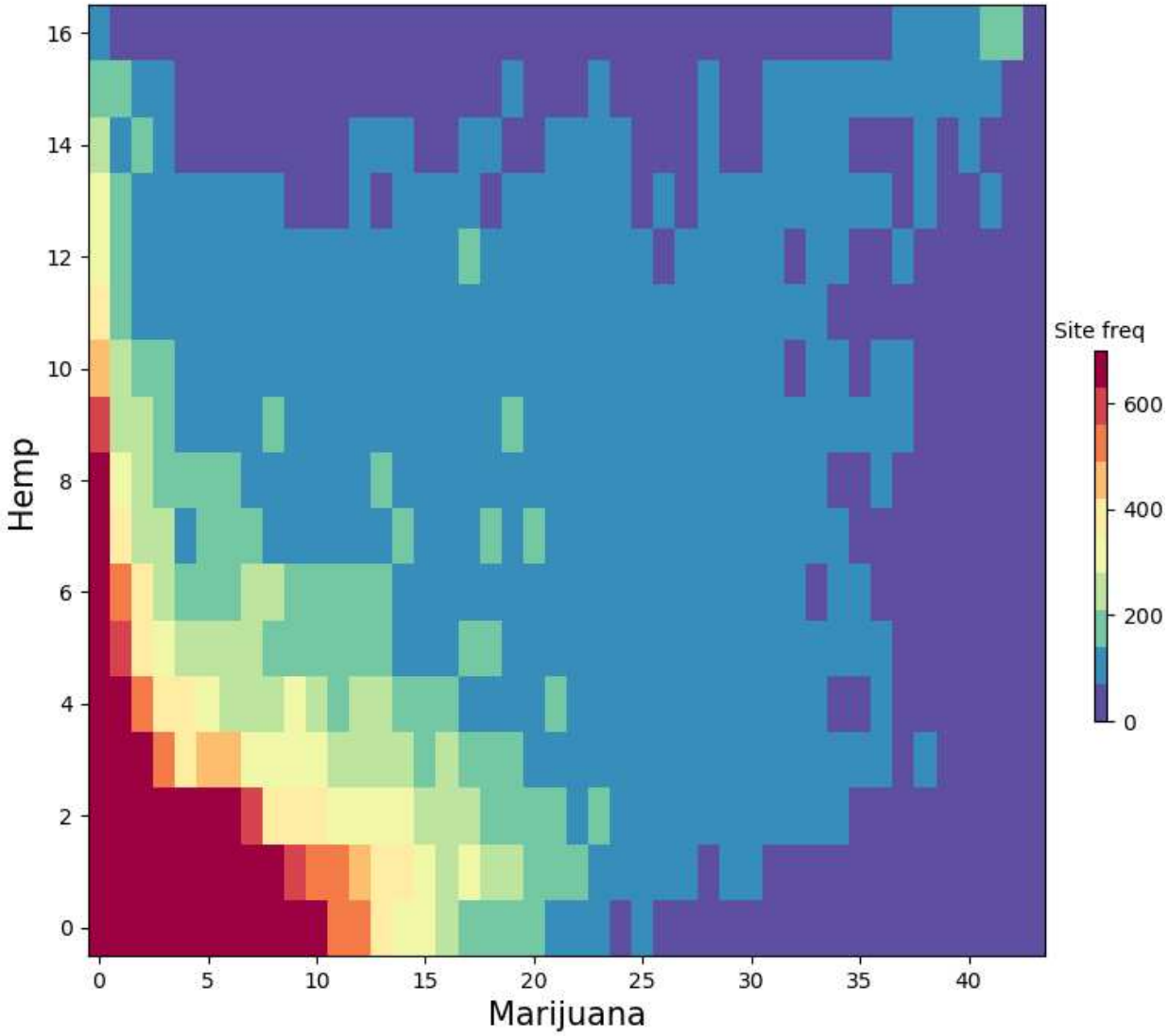


Fig. 3.1. Folded joint site frequency spectrum comparing hemp and high-THC *Cannabis* (marijuana) samples. Allelic frequencies are represented by colors, based on loci identified via GBS.

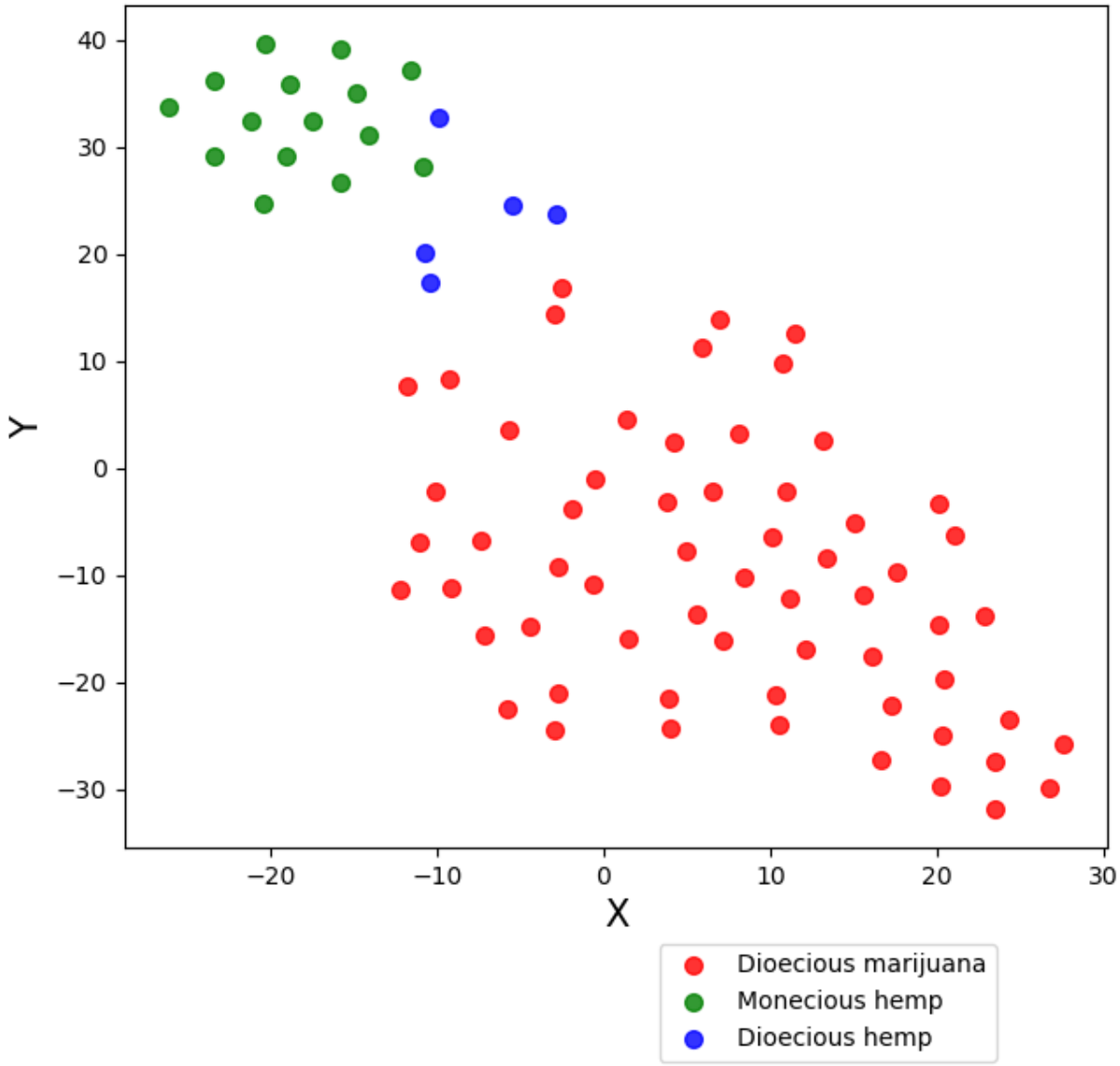


Figure 3.2: Population Clustering of Hemp and High-THC Cannabis Subpopulations. T-distributed Stochastic Neighbor Embedding (t-SNE) plot for GBS data. High-THC *Cannabis* (marijuana), monecious hemp and dioecious hemp are color-coded to distinguish from one another.

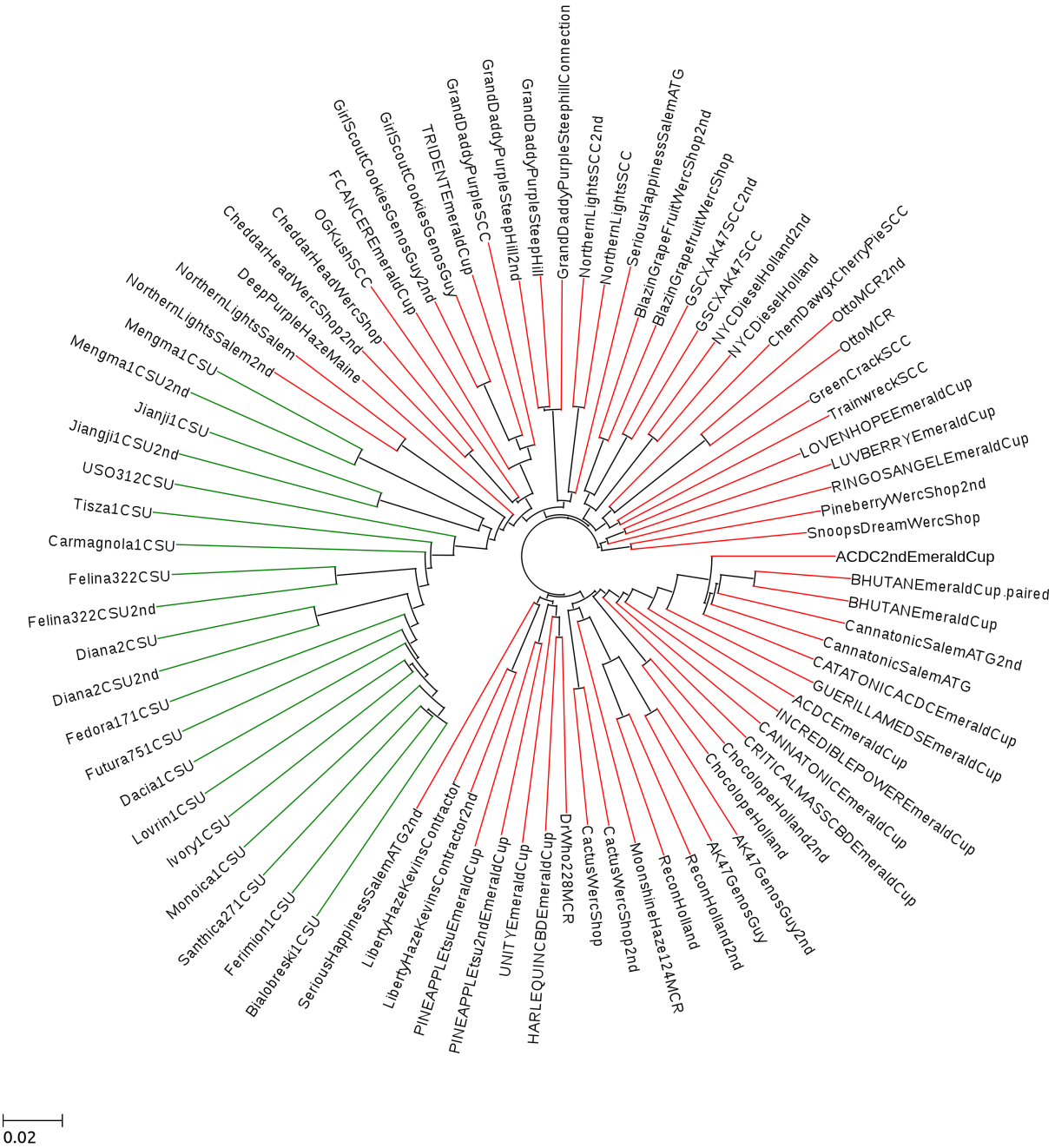


Fig. 3.3. Phylogenetic neighbor-joining tree of 81 *Cannabis* accessions. Hemp and high-THC *Cannabis* are color-coded by green (hemp) and red (high-THC *Cannabis*) branches.

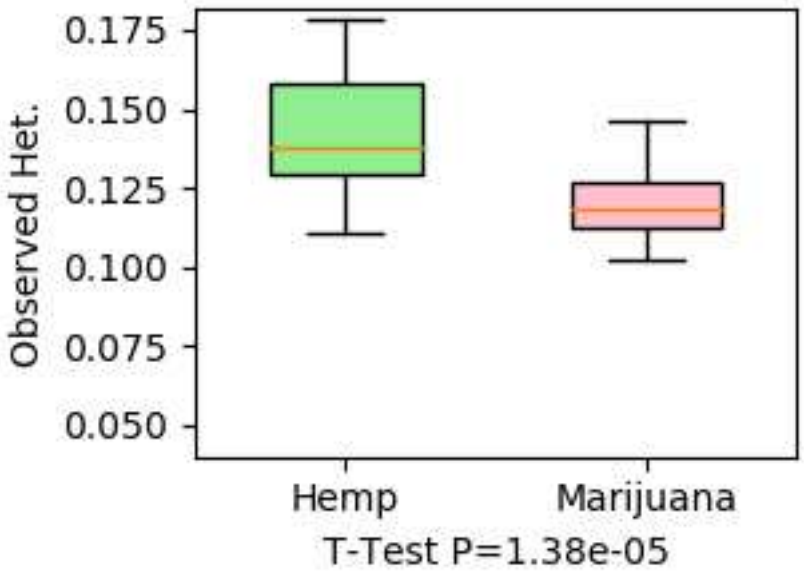


Figure 3.4.1: Observed Heterozygosity Rate of *C. sativa* Subpopulations. Each box represents the mean and interquartile range (IQR). P value of t-test = 1.38e-5 was used to reject the null hypothesis that hemp and high-THC *Cannabis* (marijuana) heterozygosity rates are not significantly different.

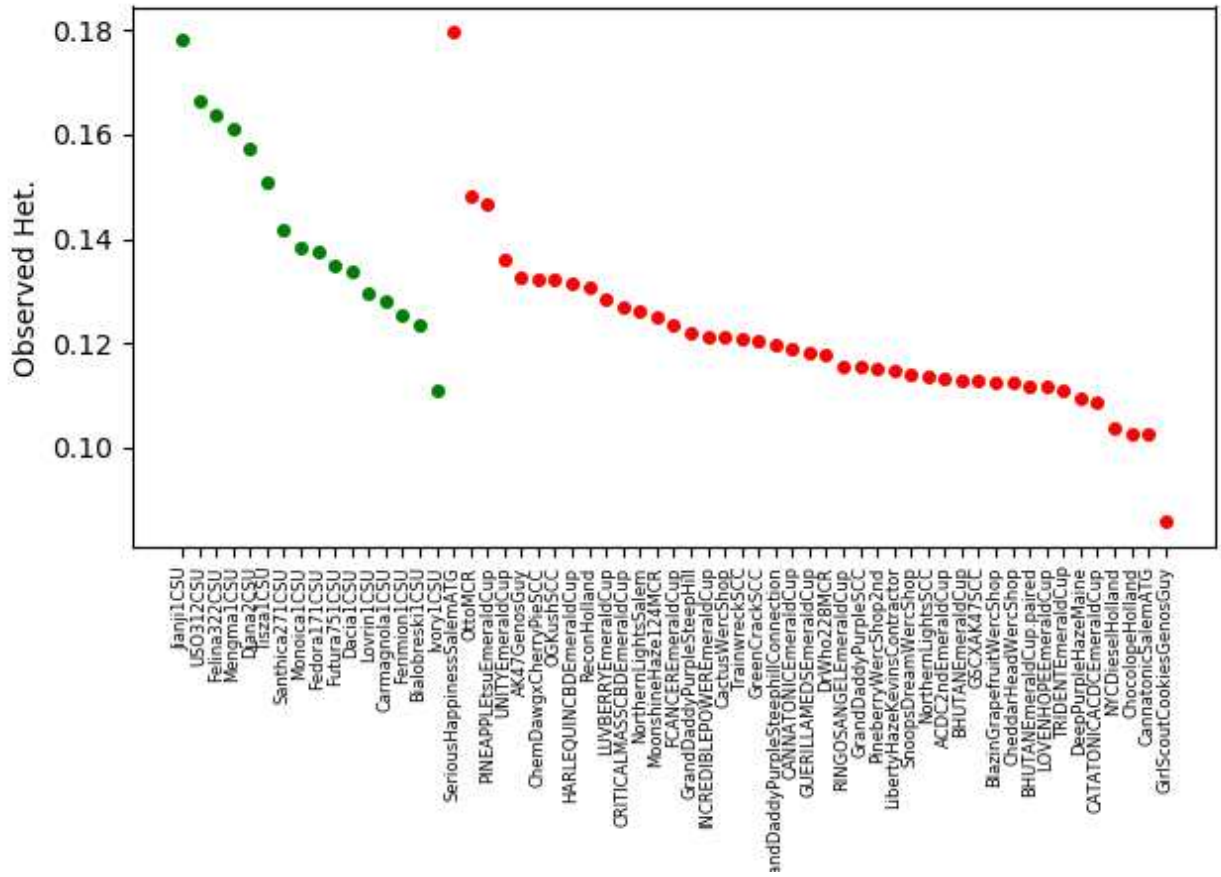


Fig. 3.4.2. Observed heterozygosity rate of individual *C. sativa* samples.

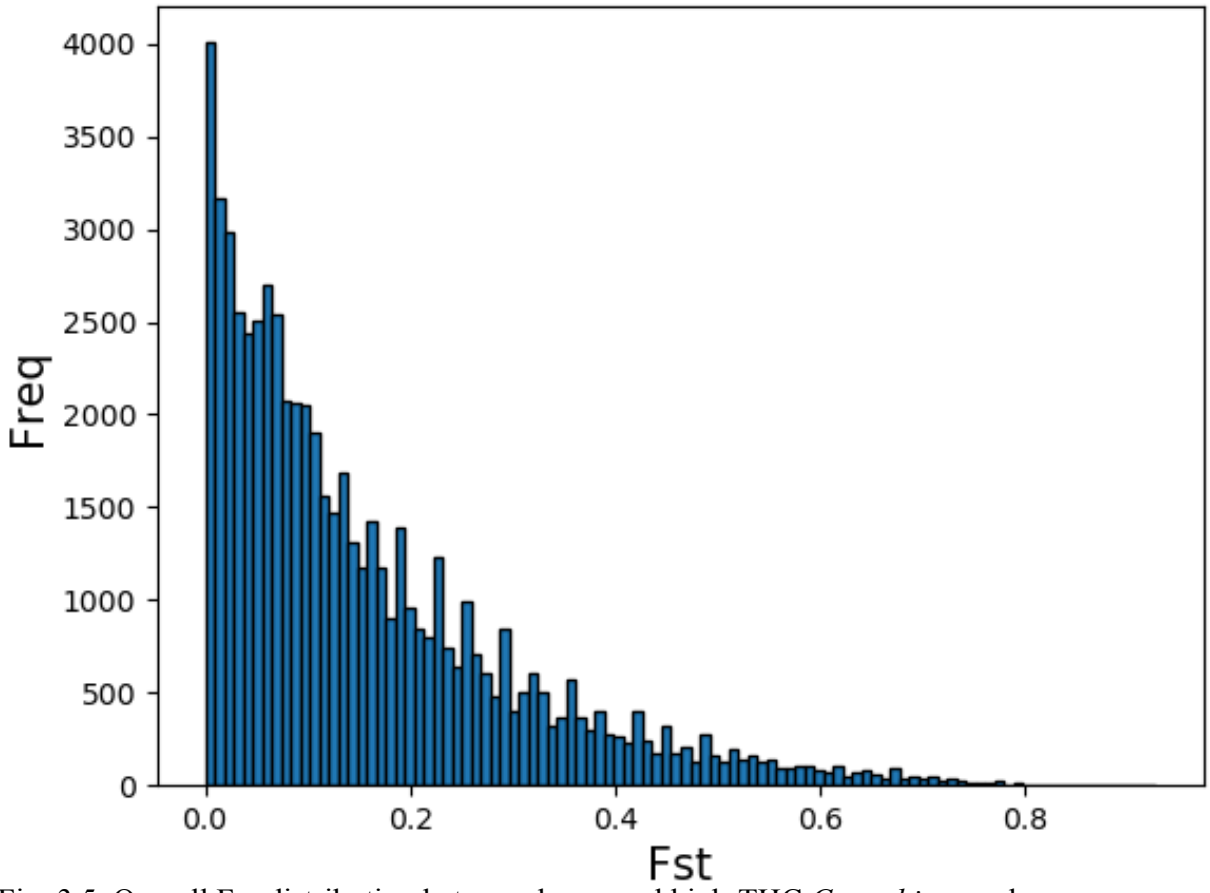


Fig. 3.5. Overall F_{ST} distribution between hemp and high-THC *Cannabis* samples.

TABLES

Table 3.1: Homologous genes in high/low frequency in the hemp and high-THC Cannabis groups

Population	Scaffold ID	Percent Identity	Alignment Length	Alignment Start Position	Alignment End Position	E-value	Accession ID of Subject Sequence	Potential Genes
Hemp	scaffold3458	75.989	733	95837	96563	4.59E-95	XM_024176681.1	enoyl-CoA delta isomerase 2, peroxisomal-like [*Morus notabilis*]
Hemp	scaffold11297	77.568	370	878	1240	5.97E-47	XM_015306116.1	uncharacterized LOC107059539 [*Solanum tuberosum*]
Hemp	scaffold11297	81.798	2258	87886	90121	0	XM_024174529.1	DEAD-box ATP-dependent RNA helicase 28 [*Morus notabilis*]
Hemp	scaffold6040	97.561	82	5886	5967	2.02E-27	X72899.1	cytoplasmic tRNA Ser (gcu) [*Nicotiana rustica*]
High-THC *Cannabis*	scaffold10728	81.081	185	13517	13701	2.07E-29	AF223328.1	retrotransposon Tib10 reverse transcriptase gene [*Ipomoea batatas*]
High-THC *Cannabis*	scaffold10728	85.892	723	61785	62507	0	D50414.1	male-associated DNA sequence [*Cannabis sativa*]

CHAPTER 4:

BI-PARENTAL QUANTITATIVE TRAIT LOCI (QTL) MAPPING DEMONSTRATES GENETIC ARCHITECTURE OF AGRONOMIC AND BIOCHEMICAL TRAITS OF INDUSTRIAL HEMP (*CANNABIS SATIVA L.*)

SUMMARY

Since industrial hemp's change in legal status with the 2014 and 2018 Farm Bills, many farmers are looking at hemp as an alternative crop with many potential end-uses and product markets. In addition to the traditional fiber and oilseed markets, there is current interest in cannabinoids and terpenes for their use in medicine, wellness products, flavorings, pesticides, etc. Although interest in hemp is at an all-time high, very little is currently understood about the genetic architecture of economically relevant traits in industrial hemp. This represents a challenge to breeders who could benefit from understanding how these traits function on a genetic level, as well as how they interact with one another. A small number of QTL studies have been performed in the species, but none have directly addressed the quantitative aspects of economically relevant traits. In this study, we developed a bi-parental QTL mapping population from parents that were divergent for a wide range of agronomic and biochemical traits. Field assessment, combined with whole-genome sequencing of the population allowed us to detect 121 significant QTL associated with 38 agronomic and biochemical traits. Some traits, such as days to maturity, exhibited a single QTL of large effect and several other QTL of minor effects. Other traits, such as α-Pinene production, were highly polygenic with epistatic interactions. Interesting correlations were also detected, like a very strong positive correlation between THC and CBD levels, which may be problematic for breeders trying to produce products in compliance with

Federal and State regulations. This study contributes actionable information to breeders of industrial hemp and also highlights the need for continued characterization of genetic architecture of quantitative traits in the species.

INTRODUCTION

Interest in cultivating and researching industrial hemp (*Cannabis sativa L.*) has steadily increased in the United States since its initial change in legal status in the 2014 Farm Bill and subsequent removal of hemp from the Controlled Substances Act in the 2018 Farm Bill (U.S. Govt., 2014; U.S. Govt., 2018). The first of these two Farm Bills legally defined hemp by a firm cutoff for content of potential delta-9-tetrahydrocannabinol (THC) at 0.3%. Since this threshold determines whether or not hemp and its products can enter the market, production of this compound is a highly important trait to understand for researchers and farmers. However, THC content is not the only biochemical trait that is important in hemp. There is a growing market for cannabinoids and terpenes, outside of THC, that are prized for their medicinal and therapeutic properties, use as flavor and smell additives, ability to make natural pesticides, and many other potential uses (Gallily et al., 2015; Russo, 2011). In particular, cannabidiol (CBD) has garnered interest as a high-value medicinal and wellness compound that can make hemp an attractive alternative crop for farmers. In addition, agronomic performance traits like grain yield and stalk/plant biomass play a role in whether hemp is a competitive crop. A number of agronomic studies have been performed on industrial hemp (Van der Werf et al., 1995a; Van der Werf et al., 1995c; Struik et al., 2000; etc.), but, despite an expanding market for hemp products, little is

understood about the genetic mechanisms controlling important agronomic and biochemical traits.

Two studies to date have utilized quantitative genetics approaches to understand traits in *Cannabis*, with one study exploring overall chemotype ratios (Weiblen et al., 2015) and another to characterize sex expression (Faux et al., 2016). Quantitative Trait Loci (QTL) studies are a first step in determining genetic architecture of traits and can lead to detection of causal genes that play a role in producing specific phenotypes. Due to legal restrictions limiting the number of plants, the study on chemotype ratios was admittedly underpowered and the only trait for which it was possible to detect a significant QTL was the determination of high-THC versus high-CBD chemotypes (Weiblen et al., 2015). Faux et al. (2016) utilized dioecious and monoecious hemp cultivars and, not suffering the same onerous restrictions of working with high-THC *Cannabis*, were able to detect five QTL associated with sex segregation and suggested that regions on both X and Y chromosomes contribute to the quantitative expression of sex in the species regardless of the presence of a Y chromosome. Both of these studies were research landmarks and support continued study into the genetic control of quantitative traits in *Cannabis*.

The purpose of the current study is to detect QTL for a broad range of physical and biochemical traits to better understand the genetic architecture of agronomically relevant traits. To enable this approach, parents that exhibited dramatically different phenotypes for many characteristics were used to create a diverse, segregating mapping population. To the author's knowledge, this study is the first of its kind to characterize QTL for agronomic traits and quantitative aspects of biochemical traits in the species.

MATERIALS AND METHODS

Plant Materials

Seed of two cultivars of industrial hemp, 'Carmagnola' and 'Uso 31', which are divergent for a range of physical and biochemical traits, were collected in the frame of the European project Multihemp (multihemp.eu) and imported from Italy to Colorado in 2015 (Campbell et al., 2019). Seed of both cultivars was planted in a greenhouse and two healthy plants were selected to create a bi-parental QTL mapping population. 'Carmagnola' is a dioecious fiber cultivar and 'Uso 31' is a monoecious dual-purpose cultivar, bred for both grain and fiber production (Tang et al., 2017). A monoecious 'Uso 31' plant was the pollen donor and a female 'Carmagnola' plant was the pollen receiver. A monoecious F_1 plant was then selfed to make the F_2 mapping population. A total of 372 F_2 individuals were genotyped and assessed in the 2017 field trial. The F_2 seeds were germinated in a greenhouse in rockwool plugs (Grodan, Roermond, the Netherlands) on May 8, 2017. Clones were taken from each of the seedlings on June 8, 2017 so that each individual was replicated in the experiment three times and the clones were transplanted into the field on June 30, 2017.

Environment

The trial was planted at Colorado State University's Agricultural Research Development and Education Center (ARDEC). This research facility is in Fort Collins, Colorado at latitude 40.65 and longitude -105.00, with an elevation of 1,557 m and average annual precipitation of 408 mm. The soil texture at this location is a sandy clay loam. ARDEC is equipped with an overhead linear sprinkler irrigation system, which was used to irrigate this trial.

The study was planted in a standing field of dent corn (*Zea mays*), which limited our ability to fertilize for hemp specific conditions. The seedbed was ripped and disked and then fertilizer was incorporated into the soil. According to soil tests and optimizing for corn, the field was amended with 372 kg ha^{-1} of nitrogen as urea. An additional 44 kg ha^{-1} of phosphorous (monoammonium phosphate) and 9 kg ha^{-1} of sulfur (granules) were also added. Glyphosate (RoundUp Powermax, Monsanto) and dicamba (Sterling Blue, Winfield United) were applied before planting to clear the field of existing weeds.

Clones were allowed to harden off outside of the greenhouse for a day and then transplanted into the field by hand. The planting density was one plant per 2.25 m^2 to avoid interplant competition. Weed pressure was controlled manually and no pesticides were applied during the growing season. Weather data were collected from the Colorado Agricultural Meteorological (COAGMET) network and the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC).

Precipitation and Irrigation

For calculating precipitation, the growing season was defined as the date of transplant into the field until the harvest of the last plot. The trial received a total of 157 mm of precipitation as rainfall and an additional 254 mm was applied as irrigation. Information comparing precipitation for the 2017 with 30-year averages is available in Supplementary Table 4.1.

Experimental Design

A Latinized row-column design was utilized to minimize spatial bias. The experiment was replicated three times, with one clone of each F_2 line and 3 clones of each parental line represented in each replicate block. The plots were 1.5m in length and width with a single plant in the center of the plot.

Data Collection

The date that each plant reached a given growth stage, e.g. initiation of flowering or maturity, was noted and the number of days that elapsed between when the clones were propagated and when the initiation of that stage was noted were calculated. Dates for initiation of flowering were noted for each plant when bracts could first be identified. Plant maturity was considered as seed maturity, i.e. when bracts began to dehisce and darkening of the seed coat was visible. Plants were harvested within three days of being scored for seed maturity.

Proportion of female flowers was scored as a visual estimate of the percentage of female flowers versus male flowers over the entire plant during the flowering stage.

Leaf water content (LWC) was measured for each plant. One leaf was collected from the middle of each plant on a single day and placed in airtight containers. The leaves were weighed, lyophilized, and then weighed again. The calculated difference in mass is reported as leaf water content.

Plant height was measured as the vertical distance from the soil surface to the tallest naturally occurring part of a plant. This measurement was taken for each plant at maturity before harvest.

Total plant biomass, referred to as biomass or dry biomass, was measured as the mass of the aboveground portion of all of the plants in a plot. Plants were cut at the soil surface and air-dried for a minimum of 30 days. The plants were then weighed before threshing. Stems were also weighed separately after threshing to determine stem biomass. The dried stems were measured at the base with digital calipers to determine stem diameter.

Grain was separated from the flowers by hand and seed was cleaned using a column blower (Agriculex, Guelph, Ontario, Canada). Grain was air-dried to approximately 8-10% seed moisture, as determined by a GAC 500XT grain moisture tester (Dickey-John, Auburn, IL) and weighed to obtain grain yield. A subsample of 50 seeds was counted from each sample and a simple calculation was performed to extrapolate Thousand Seed Mass (TSM).

Biochemical Trait Analysis

Biochemical traits were analyzed from female flowers collected after plants were dried. Seeds were removed from the flowers by hand and composite samples were made with the flower chaff. Cannabinoid and terpene profiles were analyzed using high-performance liquid chromatography (HPLC) and gas chromatography with flame ionization detector (GC-FID) by ProVerde Labs (Milford, MA).

Sample Preparation for Cannabinoid Analysis

Sample preparation for the analysis of cannabinoid profiles was performed by extraction of the cannabinoids in organic solvent. Approximately 300 mg of homogenized plant material was extracted with 4 mL of isopropanol with sonication for 20 minutes. The resulting extract

was filtered with a syringe filter, and further diluted with 71% acetonitrile (ACN) to the appropriate concentration for LC analysis, and transferred to an auto-sampler vial.

Chromatographic Cannabinoid Analysis

The liquid chromatographic analyses were performed using a ultra-high pressure liquid chromatographic system (Waters UPLC) with Photo Diode Array, UV Detection (PDA), with an Cortecs C18 column (2.7 µm, 2.1 mm x 100 mm) (Waters Corporation, MA). Mobile phases were water (A) and acetonitrile (B), both acidified with 0.1% formic acid. Separation was achieved under gradient conditions of 59-100% mobile phase B over 2.5 min at a flow rate of 0.56 mL min^{-1} at 40°C. Samples were introduced with a 3.5 µL injection, with chromatographic data collected at 225 nm.

Cannabinoid certified reference standards (Cerilliant, Sigma-Aldrich and Cayman Chemicals) were used for peak identification and generation of calibration curves used for quantitation, and included: THC, THCA, CBD, CBDA, CBG, CBGA, CBC, CBN, THCV and CBDV. Data was recorded and processed using Empower Software (Version 3, Waters Corporation).

Sample Preparation for Terpene Analysis

Analysis of terpene profiles was performed using Full Evaporative Technique GC-FID Chromatography (FET-GC-FID). The Full Evaporative Technique is a form of head-space sampling, for which standards or samples are placed and sealed directly in a head space vial. The sealed vial is equilibrated at elevated temperatures to vaporize volatile compounds for head-space sampling. For these evaluations, samples were homogenized and sealed directly in to the

head-space vials, then equilibrated for 30 minutes at 140°C prior to injection using a Hewlett Packard head-space autosampler (HP G1290A).

Chromatographic Terpene Analysis

The GC analyses were performed using Shimadzu GC-2014 gas chromatograph with Flame Ionization Detection (FID), with an Rxi-624Sil MS column (30 m x 0.25 mm x 1.4 µm) (Restek, Bellefonte, PA). Samples were introduced directly from the head-space auto sampler via a transfer line held at 160°C to prevent condensation of sample vapors prior to injection.

Nitrogen was used as the GC carrier gas at a flow rate of ~80 mL min^{-1}. Hydrogen and compressed air were used as the combustion gases. The following instrument parameters were employed: air, 50 psi; hydrogen, 70 psi; nitrogen, 60 psi; linear velocity flow control, 33 cm s^{-1}; split ratio, 20:1; injector temperature, 250°C; detector temperature, 320°C; oven program, 75°C (hold 0.4 min) to 160°C at 8°C min^{-1}; ramped to 250°C at 20°C min^{-1}; ramped to 300°C at 12.5°C min^{-1} (hold 3 min); run time, 22.2 min.

Terpene certified reference materials (Restek CRMs #34095 and 34096) were used for peak identification and generation of calibration curves used for quantitation. Data was recorded and processed using Clarity Software (Version 5.0.4.158).

Statistical Analysis

Sequencing and Linkage Mapping

Whole genome sequencing (2x150bp paired-end reads) was performed on 375 samples, including two parent plants, one F_1 plant, and 372 F_2 progenies, using the Illumina Nextera

library preparation system. Sequencing efforts aimed for approximately 7x to 10x coverage for progeny lines, 20x coverage for the F_1 sample and 30x for the two parents. A total of 3,962,882 SNP sites (call rate ≥ 0.6) were characterized.

The reference sequence was a draft haploid genome of "Purple Kush" (van Bakel et al., 2011). Trimmomatic (Bolger et al., 2014) was applied to trim partial adapter and low quality sequences. The Burrows-Wheeler alignment with maximal exact matches (BWA-MEM) algorithm (Li, 2013) was used to align reads. Sources of erroneous SNP calling include misalignment due to an incomplete reference genome, gene duplication, and regions of low-complexity. To perform high-quality variant calling and genotyping, we used SAMtools (Li et al., 2009) to keep properly paired reads and alignment with mapping quality ≥ 10. With alignment BAM files, bcftools (Li et al., 2009) was used to call variants to produce a single variant call format (vcf) table for all the samples.

The following steps were included in linkage mapping the F_2 hemp population: (1) To check contaminants, an identity-by-state (IBS)-based distance matrix calculated by TASSEL (Bradbury et al., 2007) was used to detect outliers. (2) SNPs were selected if genotype in parents are differential homozygotes with genotype quality (GQ) ≥ 98, and genotype in F_1 is heterozygous with GQ ≥ 98. (3) In F_2 progeny, SNPs with alternate allele frequency (AAF) ≥ 0.4 and ≤ 0.6 were retained for the next step; (4) To cluster markers, an adjacency matrix with Pearson's correlation (rho) were derived from the remaining SNPs. (5) To remove redundant markers, we kept one marker with the most call rate, given a list of markers having rho = 1. A total of 9342 markers remained to create a genetic map. (6) On the basis of the adjacency matrix, the Louvain method (Blondel et al., 2008) implemented in NetworkX (http://networkx.github.io/) was applied to detect communities (clusters) in a network which is formed by markers (vertices)

and pairwise correlations (edges). The partition of highest modularity was able to generate 10 linkage groups, corresponding to 10 pairs of chromosomes in *Cannabis*. Modularity clustering approaches have been successfully applied to construct genetic maps in multiple species (Lu et al., 2013; Zhang et al., 2017). (7) The R package TSPmap (Monroe et al., 2017) was used to determine the marker order in each of the clusters based on recombination frequency generated by R/qtl (Broman et al., 2003).

QTL Detection and Model Selection

QTL were detected using the add-on library Rqtl for the statistical analysis software R, version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria). Data were quantile normalized before analysis and a penalized likelihood approach was used on QTL and pairwise interactions to control false positive rates (Broman et al., 2003; Manichaikul et al., 2009). Logarithm of Odds (LOD) thresholds were determined for each trait by permutation analysis, using a 5% type I error rate and 1000 permutations. Model selection for QTL was performed using Haley-Knott regression (Haley and Knott, 1992) with a maximum QTL threshold of five unique QTL. Multiple QTL terms detected during Haley-Knott regression provided evidence of multiple QTL affecting the trait, while interaction terms were indicative of epistasis. Each locus was tested for interaction with pairwise comparisons across the marker set. Pearson's product moment correlation analysis of traits with significant QTL was performed on simple trait means using the corrplot package in R (Wei, 2013).

RESULTS

Agronomic Traits

Summary statistics for all traits are presented for comparison of population ranges with parental average values (Table 4.1). Grain yield and biomass is reported for individual plants, rather than in kg ha^{-1}. The experimental design maximized interplant differences and was not representative of a cropping environment, so these metrics were more appropriate to describe the variation observed in the study. Forty-two QTL were detected for 10 agronomic traits. A list of these QTL with their associated loci, level of significance, percent variance explained, additive effects, and best fit models is shown below (Table 4.2.1).

Eight QTL for phenology related traits were found on four linkage groups. A four-QTL model for days to flower (DTF) explained 30.25% of trait variance (LOD = 22.99) and, similarly, a four-QTL model for days to maturity (DTM) explained 48.31% variance for that trait (LOD = 41.98). One QTL colocalized on linkage group (LG) 5 at position 19.464 for DTF and DTM. QTL detected on LGs 3 and 4 for each of the traits are within 10 cM of each other, which may indicate colocalization or close linkage for two other QTL as well.

The F_2 mapping population was created by selfing a single plant that exhibited approximately 50% male flowers and 50% female flowers, however the F2 generation was overwhelmingly female. The mean proportion of flower sex was 99.7% female across the population (Table 4.1). Despite the low number of individuals with male flowers, a relatively complex QTL model, comprised of five QTL on two LGs and two epistatic interactions, explained 18.75% of the variation in this trait. Due to many ties in rank and a highly non-normal distribution, however, this result should be interpreted with caution.

A wide range of values were observed for grain yield. Due to stress, some of the plants produced no seed while other individual plants produced up to 141.58 g of seed, with a population mean of 23.6 g plant^{-1}. This trait exhibited substantial transgressive segregation from the two parents which averaged 18.6 g plant^{-1} and 78.5 g plant^{-1} for 'Uso 31' and 'Carmagnola', respectively. Five QTL explained 46.19% of variance in grain yield (LOD = 39.56), with a single locus of large effect on LG 5 explaining 28.59% variance for the trait. Thousand seed mass (TSM) was used to characterize seed size and weight and also ranged considerably from 3 g to 41.5 g in the population, with a mean TSM of 14.8 g. The F_2 population again showed a mid-parent mean, with 'Uso 31' and 'Carmagnola' averaging 10.8 g and 17.2 g TSM, respectively. A five-QTL model also best explained variance in TSM (44.13%, LOD = 37.67), which also had a large effect QTL on LG 5 (19.04% variance, LOD = 18.98). The large effect QTL for grain yield and TSM were 6 cM apart, indicating possible colocalization for these QTL.

Total plant biomass and its component, stem biomass, exhibited highly similar genetic architecture in this population. Five- and four-QTL models were fit for total biomass and stem biomass, respectively. Three of the four QTL for stem biomass directly colocalized with QTL for total biomass and the fourth was less than 2 cM away from one of the other two QTL. Both models explained a substantial amount of trait variation. The five-QTL model for total biomass explained 46.17% of trait variance (LOD = 39.67) and the four-QTL model for stem biomass explained 52.01% variance (LOD = 47.03). Similar to grain yield and TSM, both total biomass and stem biomass also had a large effect QTL on LG 5 in the same region that explained a substantial proportion of the model variance, 26.20% for total biomass (LOD = 25.41) and 32.89% for stem biomass (LOD = 33.44). Stem diameter shared several of the same QTL as well, with three of its four QTL at or near the same loci as the models for the biomass traits on

LGs 3, 4, and 5. The fourth QTL not directly shared was on LG 10 like the other traits, but at a different position. The four-QTL model for stem diameter explained 44.92% of trait variance (LOD = 38.20).

Contrasting phenotypes were also observed for the leaf water content (LWC) of the parental lines, as well as transgressive segregation in the F_2 population. 'Uso 31' had a mean LWC of 0.50 g and 'Carmagnola' averaged 1.98 g. The population mean was 1.10 g and ranged from 0.10 g to 3.54 g. Two QTL on LGs 3 and 5, explained 33.02% of variance in LWC with LOD = 25.93.

Plants in the F_2 population tended to be shorter, on average, than either parent with a population mean of 90.4 cm. 'Uso 31' had a mean height of 94.3 cm while 'Carmagnola' had a mean height of 164.3 cm. The genetic architecture for plant height was very similar to other agronomic traits; a four-QTL model, with loci on LGs 3,4,5, and 10, that explained 44.82% variance in height with LOD = 35.25. Although these loci did not share exact loci with the other traits they were in close proximity with QTL from many of the agronomic traits measured and colocalization or linkage with QTL for other yield related traits is likely.

Biochemical Traits

QTL associated with the production of six cannabinoids and 22 terpenes were detected in this mapping population (Tables 2.2 and 2.3). For the purpose of brevity, the cannabinoid and two terpenes that were present in the highest quantities will be described in the results, along with THCA because of its economic and regulatory importance.

Since the parents of the population were traditional industrial hemp varieties, they contained low levels of THC and both varieties regularly tested below the legal threshold of

0.3%. ‘Uso 31’ had a mean THCA content of 0.003% and ‘Carmagnola’ had a mean content of 0.070%. The mapping population ranged from 0-0.060%, with a mean THCA content of 0.004%. However, quantitative variation was observed in the population and we were able to detect two QTL for the production of THCA that explained 45.39% of the variance in the trait (LOD = 29.56). The two QTL contributed similar amounts to trait variance, with a QTL on LG 6 explaining 17.24% of trait variance (LOD = 13.41) and a QTL on LG 9 explaining 22.92% of variance in THCA production (LOD = 17.13).

The most abundant cannabinoid in the mapping population was CBDA. Since the population was developed from seed and fiber cultivars of hemp, CBD production was modest compared to cultivars that have been specifically bred for cannabinoid production, but, like THC, there was measurable variation in the set. The highest levels of CBD were produced in ‘Carmagnola’, with a mean CBDA content of 1.917%. ‘Uso 31’ had comparatively low CBDA content at 0.205%. The F_2 population had a mean of 0.164% CBDA and a range from 0-1.890% A QTL model was able to be developed from these data that explained 60.42% of the variance in CBDA production (LOD = 45.28). This was a relatively complex model with five QTL and four interactive terms. Two of these loci on LGs 6 and 9 were near the two loci for THCA production. The QTL on LG 6 had the strongest influence on the trait, responsible for 15.41% of variance in CBDA content (LOD = 16.07).

The terpene with the highest levels throughout the study was α-Pinene. ‘Uso 31’ had a mean α-Pinene content of 28.65 ppm and ‘Carmagnola’ had a mean content of 413.34 ppm. The mean of the F_2 population was 84.92 ppm, with a minimum value of 2.6 ppm and a maximum of 1081.86 ppm. A four-QTL model with one interactive term was the best fit for this trait with a model LOD of 29.69 and 47.22% variance explained. The QTL were on LGs 6,8, and 9. The

QTL on LGs 6 and 9 may colocalize with QTL found on those LGs for both THCA and CBDA production. The four QTL for α-Pinene were all of similar, modest effect, ranging from 3.55%-6.86% of variance in the production of α-Pinene.

The second most common terpene in the study was β-Caryophyllene. As with the other cannabinoids and terpenes described, 'Uso 31' produced less of the compound than 'Carmagnola' with mean values of 17.37 ppm and 131.25 ppm, respectively. The F_2 population mean was 32.83 ppm and ranged from 0-649.88 ppm. A five-QTL model explained 65.06% of variance in β-Caryophyllene levels with a LOD score of 48.87. The QTL in this model were each on separate LGs and included LGs 2,6,8,9 and 10. A QTL on LG5 is on the same scaffold as the QTL for THCA and CBDA production, showing evidence for a shared genetic basis for cannabinoid and terpene production.

Correlations

To further explore the relationships between these traits and their corresponding QTL, Pearson's product moment correlations were calculated using trait mean values. Many significant correlations were detected in this suite of traits (Figure 4.1; Supplementary Tables 4.2.1, 4.2.2, and 4.2.3). Clear trends can be seen, with the majority of the agronomic traits showing significant, positive correlations with each other and the biochemical traits showing significant, positive correlations amongst that group (Figure 4.1). These correlations show that plants that were larger, i.e. taller, thick-stemmed, and with more biomass, tended to yield more and larger grain. DTM was also correlated positively with these traits, but not as strongly as the relationships between yield and yield components. The biochemical trait correlations are slightly more variable but tend to be positively correlated with each other with the exception of Geraniol, which showed weak significant negative correlations with several of the biochemical traits,

ranging from r-values of -0.140 to -0.213. CBG also shows almost no correlations with other traits, with a few minor exceptions. One result to point out is that THCA and CBDA production appear to be highly positively correlated in this population with an r-value of 0.918 and p-value of 0 (the Pearson's correlation analysis in R rounds any p-values of more than 16 decimal points without a number above 0 to read as 0).

DISCUSSION

Overall, we found 121 significant QTL associated with 38 agronomic and biochemical traits, representing a significant addition to the body of knowledge surrounding genetic architecture of quantitative traits in industrial hemp. A range of models were observed with some traits strongly influenced by a small number of QTL, like leaf water content or THCA production, and other traits that were highly polygenic and exhibited epistatic interactions, like α-Pinene production. Of note was the detection of significant QTL for production of both THCA and CBDA which are important for the crop both legally and commercially. The complex model for CBDA production explained over 60% of the variance observed in the trait which supports the idea that CBDA content is a polygenic trait, but with one or more QTL of major effect that may enable the trait to be bred as an oligogenic trait. One potential challenge to breeding hemp for CBD content is its strong positive correlation with THC content, indicating that as levels of one increase the other compound will increase as well. This supports previous findings that total CBD and THC levels are positively correlated and have a substantial element of genetic control (Campbell et al., 2019). Since people have only recently begun breeding hemp to have high levels of CBD and low levels of THC, these mechanisms need to be more fully characterized. Particularly, since both of these compounds are highly relevant to the success or failure of a crop,

it is important to assess these traits and this correlation in a wider range of environments as well as other genetic backgrounds with more substantial levels of cannabinoids. Breaking this problematic correlation may be able to be achieved via recombination or genome editing to break close linkage or there may be underlying biological and/or biochemical explanations, which warrants further study.

Some of the correlations detected fell in line with expectations, where plants with greater biomass were taller, larger-stemmed plants that yielded more grain. The positive correlations of these traits with DTF and DTM imply that phenology plays a role, where plants that took longer to flower and mature fared better in terms of biomass and seed production. A previous study noted that this relationship is not necessarily linear, however, and cultivars that take too long to flower and mature can suffer in performance (Campbell et al., 2019). The plants in this population seem to have not reached the point where substantial losses would occur from flowering too late. This is possibly due to including 'Uso 31' as a parent which is a very early flowering cultivar, or simply that the particular growing season did not experience an early hard frost. Leaf water content was also correlated with these agronomic traits, which may imply that water use efficiency plays a role in achieving high-yielding phenotypes in plants that experience drought stress. It has been demonstrated that LWC can be associated with drought tolerant phenotypes and can be reliably mapped in QTL populations (McKay et al., 2008). However, since yield-related traits are positively correlated and share QTL with each other and yield, selection for optimal allelic combinations should be able to be reasonably achieved with selection for optimal phenology as a prerequisite for local adaptation.

Because this study is the first of its kind and the study was limited to a single environment, it is important to reiterate that these QTL are a starting point for exploring the

genetic architecture of quantitative traits and need to go through substantial validation, e.g. fine mapping, verification of QTL effects in different genetic backgrounds, development of near-isogenic lines, etc., before they are able to be implemented in breeding programs or used to identify causal genes. However, the detection of many statistically significant QTL for a range of traits supports the idea that using a diverse F_2 population to study multiple traits in tandem is a valid approach to understanding economically relevant traits in industrial hemp. As genetic resources for hemp become more developed and more readily available, the utility of field-based data collection will only continue to increase. It is our hope that this work contributes to an expanding body of knowledge that will enable future studies in functional genetics, as well as breeding activities to develop hemp cultivars for a wide range of end uses and target environments.

REFERENCES

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of Cannabis sativa. *Genome Biol* 12: R102

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 10008: 6

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890

Campbell, B., Berrada, A.F., Hudalla, C., Amaducci, S., McKay, J. 2019. Genotype-by-Environment Interactions of Industrial Hemp (*Cannabis sativa, L.*) Cultivars Highlight Diverse Responses to Environmental Factors. *Agrosystems, Geosciences, and Environment* (in press)

Gallily, R., Yekhtin, Z. and Hanuš, L.O., 2015. Overcoming the bell-shaped dose-response of cannabidiol by using cannabis extract enriched in cannabidiol. *Pharmacol Pharm*, *6*(2), pp.75-85.

Haley, C. S., and Knott S.A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315–324

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:*1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genet* 9: e1003215

Manichaikul A., Moon J.Y., Sen S., Yandell B.S., Broman K.W. 2009. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* 181:1077-1086.

McKay, John K., et al. (2008) Genetics of drought adaptation in Arabidopsis thaliana II. QTL analysis of a new mapping population, KAS-1× TSU-1. *Evolution: International Journal of Organic Evolution* 62.12: 3014-3026.

Monroe JG, Allen ZA, Tanger P, Mullen JL, Lovell JT, Moyers BT, Whitley D, McKay JK (2017) TSPmap, a tool making use of traveling salesperson problem solvers in the efficient and accurate construction of high-density genetic linkage maps. *BioData Min* 10: 1–15

Russo, Ethan B. 2011. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. British journal of pharmacology 163.7: 1344-1364.

Struik, P. C., et al. Agronomy of fibre hemp (Cannabis sativa L.) in Europe. *Industrial crops and products* 11.2-3 (2000): 107-118.

U.S. Govt. Print. Farm Bill – Section 7606 (2014) (enacted). Print

U.S. Govt. Print. Farm Bill – Sections 7125, 7401, 7415, 7501,7605, 10111, 10113, 11101, 11106, 11112, 11120, 11121, 12608, 12619 (2018) (enacted). Print

Van der Werf, H.M.G., Brouwer, K., Wijlhuizen, M., Witha- gen, J.C.M., 1995a. The effect of temperature on leaf appearance and canopy establishment in fibre hemp (*Cannabis sativa* L.). *Ann. Appl. Biol.* 126, 551–561.

Van der Werf, H.M.G., Van Geel, W.C.A., Van Gils, L.J.C., Haverkort, A.J., 1995c. Nitrogen fertilization and row width affect self-thinning and productivity of fibre hemp (*Cannabis sativa* L.). *Field Crops Res.* 42, 27–37.

Wei, Taiyun. 2013. corrplot: Visualization of a correlation matrix. R package version 0.73. http://CRAN.R-project.org/package=corrplot

Zhang D, Easterling KA, Pitra NJ, Coles MC, Buckler ES, Bass HW, Matthews PD (2017) Non-Mendelian Single-Nucleotide Polymorphism Inheritance and Atypical Meiotic Configurations are Prevalent in Hop. *Plant Genome* 10: 0

FIGURES

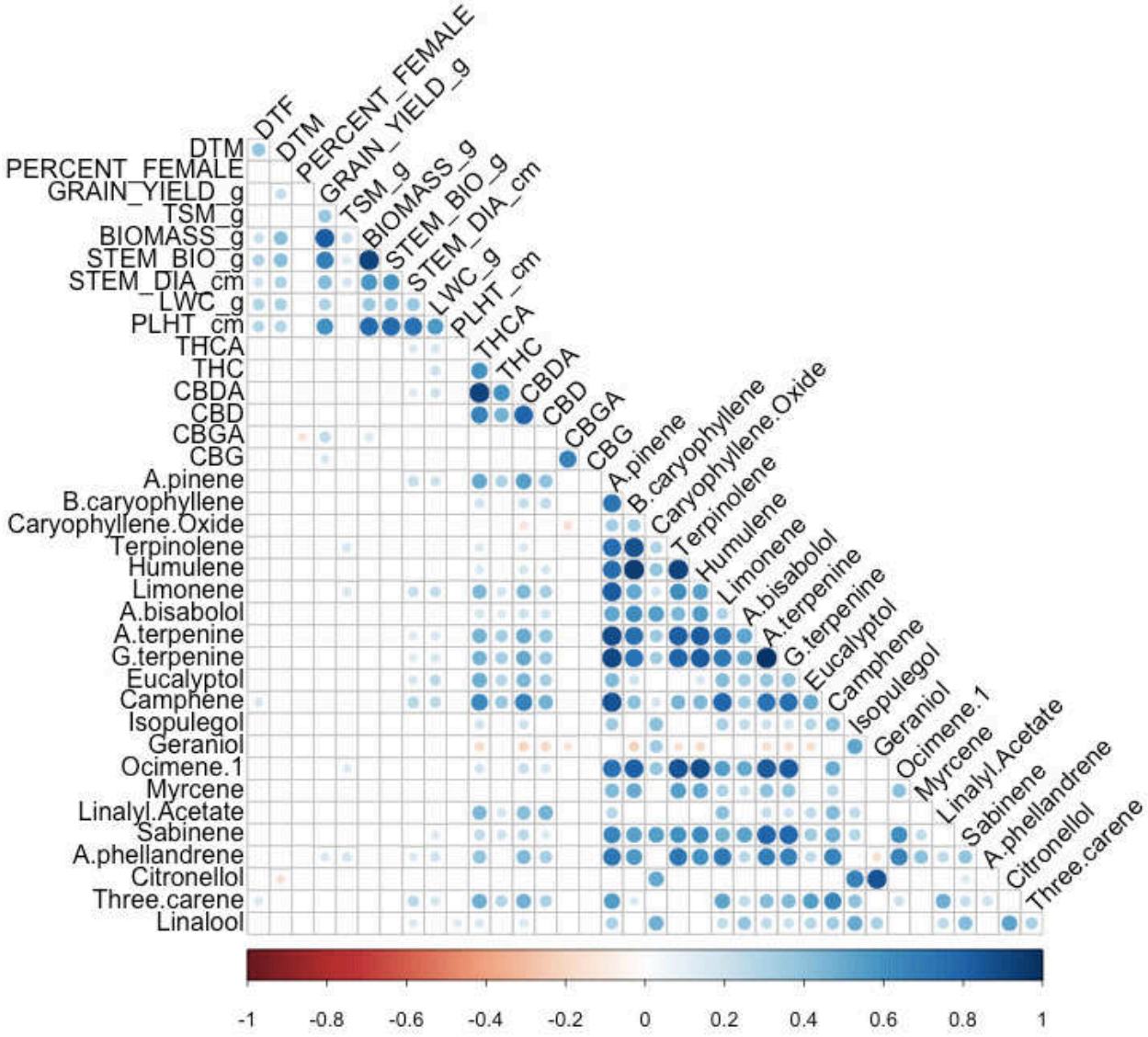


Figure 4.1. Trait correlation heat map. Only statistically significant correlations are shown, with the size and color of the circles representing the strength and direction of the correlations. Abbreviations: DTF = Days to Flower, DTM = Days to maturity, Percent Female = Proportion of female to male flowers, TSM = Thousand Seed Mass, Stem Bio = Stem Biomass, Stem Dia = Stem Diameter, LWC = Leaf Water Content, PLHT = Plant Height, THCA = Tetrahydrocannabinolic acid, THC = Δ9-Tetrahydrocannabinol, CBDA = Cannabidiolic Acid, CBD = Cannbidiol, CBGA = Cannbigerolic acid, CBG = Cannabigerol

TABLES

Table 4.1: F_2 Population Phenotype Ranges and Means Compared with Parental Values.

Trait	Uso 31 Mean	Carmagnola Mean	F_1 Mean	F_2 Population Mean	F_2 Population Minimum	F_2 Population Maximum
Days to Flower (Female)	39	51	38	38	21	72
Days to Maturity	89	118	95	100	82	128
Proportion of Female Flowers (%)	50	100	60	99.7	40	100
Grain Yield (g)	18.6	78.5	43.8	23.6	0	141.6
Thousand Seed Mass (g)	10.8	17.2	17.2	14.8	3	41.5
Total Biomass (g)	67.9	350.0	121.9	102.2	2.5	452.1
Stem Biomass (g)	19.7	107.7	21.4	24.2	0.4	142.0
Stem Diameter (cm)	0.7	2.1	1.0	1.0	0.2	6
Leaf Water Content (g)	0.50	1.98	0.68	1.10	0.10	3.54
Plant Height (cm)	94.3	164.3	76.1	90.4	13.5	190.5
THCA (%)	0.003	0.070	0.005	0.004	0	0.06
Δ9-THC (%)	0	0.007	0	0.000	0	0.015
CBDA (%)	0.205	1.917	0.118	0.164	0	1.890
CBD (%)	0.014	0.052	0.007	0.008	0	0.178
CBGA (%)	0.008	0.037	0.003	0.024	0	0.550
CBG (%)	0.003	0	0	0.001	0	0.024
α-Pinene (ppm)	28.65	413.34	574.13	84.92	2.6	1081.86
β-Caryophyllene (ppm)	17.37	131.25	347.98	32.83	0	649.88
Caryophyllene Oxide (ppm)	17.69	6.70	110.20	13.98	0	209.61
Terpinolene (ppm)	1.22	34.88	116.87	13.47	0	398.71
Humulene (ppm)	7.63	49.89	161.55	13.19	0	322.14
Ocimene-2 (ppm)	2.42	1.72	119.81	9.07	0	377.57
Limonene (ppm)	6.78	23.51	17.75	4.45	0	46.28
α-Bisabolol (ppm)	0	7.28	25.89	2.95	0	50.33
α-Terpenine (ppm)	0.84	8.25	1.30	2.41	0	51.12
ϒ-Terpenine (ppm)	1.03	8.44	51.94	2.09	0	101.23
Eucalyptol (ppm)	0	5.99	0	2.28	0	26.75
Camphene (ppm)	0.81	12.66	0.66	1.79	0	20.91
Isopulegol (ppm)	2.33	15.05	0	1.75	0	35.15
Geraniol (ppm)	0	0	0.68	1.62	0	22.74
Ocimene-1 (ppm)	1.42	0.32	1.91	1.31	0	41.23
Myrcene (ppm)	20.77	201.35	0.60	1.30	0	238.33
Linalyl Acetate (ppm)	1.96	3.90	1.655	1.29	0	10.24
Sabinene (ppm)	0.55	1.16	43.73	1.12	0	22.97
α-Phellandrene (ppm)	1.20	4.68	0.23	0.95	0	15.19
Guaiol (ppm)	1.06	2.53	0.83	0.89	0	13.2
3-Carene (ppm)	0	3.90	0	0.58	0	22.35
Linalool (ppm)	2.36	12.35	0	0.57	0	23.75

Table 4.2.1: QTL Results for Agronomic Traits.

Trait	Marker	Linkage Group	Position	Additive Effects	Std. Error	LOD	% Variance
Days to Flower	29696686	3	79.682	-0.304	0.074	4.04	4.56
	33015451	4	37.493	-0.347	0.074	4.93	5.59
	70297735	4	54.607	-0.364	0.076	4.69	5.31
	16737372	5	19.464	-0.456	0.456	10.57	12.56
Model	y ~ Q1 + Q2 + Q3 + Q4					22.99	30.25
Days to Maturity	c3.loc64	3	64.000	-0.460	0.079	6.37	5.44
	85593559	4	62.654	-0.275	0.079	5.72	4.86
	16737372	5	19.464	-0.888	0.077	30.40	31.67
	7725327	9	45.576	0.165	0.086	3.35	2.80
Model	y ~ Q1 + Q2 + Q3 + Q4					41.98	48.31
Proportion of Female Flowers (%)	2949655	5	17.199	-0.139	0.045	3.35	4.38
	46499187	5	27.755	-0.084	0.047	1.40	1.79
	60401583	10	13.689	-0.187	0.037	5.60	7.44
	c10.loc59	10	59.000	-0.138	0.041	2.14	2.78
	30207402	10	70.392	-0.074	0.044	1.98	2.55
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q4 + Q2:Q5					13.25	18.75
Grain Yield (g)	87143848	2	77.000	0.315	0.078	7.68	6.88
	85717545	3	52.576	-0.208	0.082	5.27	4.63
	82057588	4	55.560	-0.215	0.084	4.22	3.68
	c5.loc18	5	18.000	-0.706	0.085	27.21	28.59
	c10.loc55	10	55.000	0.012	0.086	4.77	4.18
Model	- Q1 + Q2 + Q3 + Q4 + Q5					39.56	46.19
Thousand Seed Mass (g)	86013226	2	81.983	0.170	0.080	3.36	2.98
	80469429	3	59.187	-0.364	0.079	9.69	9.02
	c5.loc24	5	24.000	-0.220	0.084	18.98	19.04
	14987238	8	19.865	-0.410	0.082	6.94	6.32
	c10.loc9	10	9.000	-0.304	0.080	7.44	6.80
Model	- Q1 + Q2 + Q3 + Q4 + Q5					37.67	44.13
Total Biomass (g)	51181046	3	62.533	-0.408	0.080	6.64	5.88
	85998317	4	64.965	-0.279	0.078	6.39	5.65
	c5.loc19	5	19.000	-0.795	0.082	25.41	26.20
	c9.loc16	9	16.000	-0.736	0.082	3.91	3.39
	c10.loc50	10	50.000	-0.038	0.088	3.78	3.27
Model	- Q1 + Q2 + Q3 + Q4 + Q5					39.67	46.17
Stem Biomass (g)	41478776	3	60.917	-0.497	0.077	9.80	7.93
	85998317	4	64.965	-0.277	0.078	7.61	6.06
	c5.loc19	5	19.000	-0.900	0.079	33.44	32.89
	c10.loc50	10	50.000	-0.153	0.088	5.91	4.63
Model	y ~ Q1 + Q2 + Q3 + Q4					47.03	52.01
Stem Diameter (cm)	c3.loc58	3	58.000	-0.538	0.076	12.36	11.73
	85998317	4	64.965	-0.210	0.078	4.74	4.23
	30072537	5	23.425	-0.781	0.082	23.12	23.94
	2787586	10	30.213	-0.239	0.087	4.29	3.82
Model	y ~ Q1 + Q2 + Q3 + Q4					38.20	44.92
Leaf Water Content (g)	c3.loc60	3	60.000	-0.591	0.073	12.95	14.84
	c5.loc16	5	16.000	-0.612	0.083	13.11	15.04
Model	y ~ Q1 + Q2					25.93	33.02
Plant Height (cm)	59281665	3	58.696	-0.632	0.074	18.24	19.88
	c4.loc73	4	73.000	-0.185	0.090	3.05	2.91
	16737372	5	19.464	-0.755	0.104	16.94	18.25
	24772745	10	47.177	-0.132	0.090	4.44	4.29
Model	y ~ Q1 + Q2 + Q3 + Q4					35.25	44.82

Table 4.2.2: QTL Results for Cannabinoid Traits.

Trait	Marker	Linkage Group	Position	Additive Effects	Std. Error	LOD	% Variance
THCA (%)	c6.loc25	6	25.000	-0.568	0.075	13.41	17.24
	1824910	9	35.259	-0.695	0.075	17.13	22.92
Model	y ~ Q1 + Q2 + Q1:Q2					29.56	45.39
Δ9-THC (%)	31036998	4	34.187	0.042	0.041	4.54	7.57
	30232633	4	35.820	0.074	0.039	4.93	8.25
	24162086	4	53.536	0.075	0.040	0.39	0.62
	71062763	6	0.000	-0.052	0.037	0.68	1.09
	c9.loc29	9	29.000	-0.144	0.039	3.94	6.53
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q2:Q5 + Q1:Q5 + Q4:Q5 + Q3:Q4					12.35	22.33
CBDA (%)	23542577	6	33.349	0.264	0.083	16.07	15.41
	c7.loc47	7	47.000	-0.157	0.089	0.03	0.03
	2452161	9	29.237	-0.847	0.073	2.44	2.03
	c9.loc35	9	35.000	-0.913	0.075	1.98	1.64
	5092276	9	43.217	-0.727	0.079	0.46	0.37
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q4 + Q4:Q5 + Q1:Q2					45.28	60.42
CBD (%)	49183951	6	25.400	-0.595	0.074	15.78	19.08
	1824910	9	35.259	-0.729	0.073	20.56	26.18
Model	y ~ Q1 + Q2 + Q1:Q2					33.82	49.95
CBGA (%)	c6.loc25	6	25.000	0.449	0.086	12.61	17.63
	c9.loc35	9	35.000	-0.654	0.082	17.98	26.64
Model	y ~ Q1 + Q2 + Q1:Q2					25.07	40.13
CBG (%)	c6.loc17	6	17.000	0.372	0.055	14.88	23.46
	c9.loc8	9	8.000	-0.152	0.062	0.07	0.10
	2216196	9	28.480	-0.244	0.063	2.62	3.62
	2087676	9	33.012	-0.222	0.063	2.09	2.88
	c9.loc110	9	110.000	-0.048	0.065	0.03	0.03
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q4 + Q4:Q5 + Q1:Q3 + Q1:Q2					20.38	34.11

Table 4.2.3: QTL Results for Terpene Traits.

Trait	Marker	Linkage Group	Position	Additive Effects	Std. Error	LOD	% Variance
α-Pinene (ppm)	45651507	6	24.955	-0.454	0.091	5.52	6.66
	58688884	8	32.754	-0.342	0.098	5.68	6.86
	c9.loc30	9	30.000	-0.662	0.081	3.80	4.50
	c9.loc41	9	41.000	-0.736	0.086	3.02	3.55
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q1:Q3					29.69	47.22
β-Caryophyllene (ppm)	c2.loc20	2	20.000	-0.381	0.090	8.31	6.84
	23434860	6	8.092	-0.436	0.112	4.49	3.54
	70613512	8	32.888	-0.314	0.097	6.59	5.33
	c9.loc35	9	35.000	-0.651	0.073	33.54	36.97
	34520874	10	126.550	-0.160	0.089	6.06	4.86
Model	y ~ Q1 + Q2 + Q3 + Q4 + Q5					48.87	65.06
Caryophyllene Oxide (ppm)	2456216	9	32.980	0.468	0.093	5.56	9.89
	c10.loc115	10	115.000	0.454	0.086	5.63	10.02
Model	y ~ Q1 + Q2					11.61	22.11
Terpinolene (ppm)	23434860	6	8.092	-0.420	0.104	6.77	9.42
	c9.loc32	9	32.000	-0.524	0.076	18.43	29.26
Model	y ~ Q1 + Q2					23.66	39.89
Humulene (ppm)	1443158	2	19.789	-0.269	0.090	3.78	3.30
	69076640	3	59.755	-0.091	0.090	2.74	2.37
	69052559	8	21.588	-0.328	0.090	6.55	5.91
	c9.loc35	9	35.000	-0.671	0.873	31.30	37.49
	34520874	10	126.550	-0.148	0.088	5.05	4.48
Model	- Q1 + Q2 + Q3 + Q4 + Q5 + Q3:Q4					43.76	61.00
Ocimene-2 (ppm)	1084949	1	18.028	0.267	0.089	4.25	6.32
	65888416	6	19.600	-0.297	0.081	4.29	6.38
	c9.loc33	9	33.000	-0.366	0.080	13.56	22.33
Model	y ~ Q1 + Q2 + Q3					19.35	34.06
Limonene (ppm)	18614498	9	79.179	-0.341	0.086	3.89	8.03
Model	y ~ Q1					3.89	8.03
α-Bisabolol (ppm)	c1.loc60	1	60.000	0.273	0.083	4.16	5.67
	c3.loc51	3	51.000	-0.078	0.084	4.90	6.74
	58732028	8	30.288	-0.428	0.084	8.04	11.43
	2452161	9	29.237	-0.301	0.078	9.18	13.22
Model	y ~ Q1 + Q2 + Q3 + Q4					23.33	39.47
α-Terpenine (ppm)	65888416	6	19.600	-0.446	0.076	7.46	11.58
	c9.loc33	9	33.000	-0.586	0.078	11.88	19.36
Model	y ~ Q1 + Q2					19.00	33.57
ϒ-Terpenine (ppm)	49129733	6	40.551	-0.377	0.077	4.90	6.79
	27276368	8	28.417	-0.204	0.087	3.74	5.12
	1824910	9	35.259	-0.644	0.076	16.64	26.30
Model	y ~ Q1 + Q2 + Q3					22.90	38.90
Eucalyptol (ppm)	c1.loc17	1	17.000	-0.236	0.080	6.17	10.60
	49940261	1	47.480	0.147	0.076	5.44	9.28
	1824910	9	35.259	-0.382	0.075	6.59	11.39
Model	y ~ Q1 + Q2 + Q3					13.55	25.30

Trait	Marker	Linkage Group	Position	Additive Effects	Std. Error	LOD	% Variance
Camphene (ppm)	49183951	6	25.400	-0.411	0.080	6.67	10.64
	2087676	9	33.012	-0.515	0.082	7.68	12.39
	c10.loc81	10	81.000	0.270	0.086	3.59	5.53
Model	y ~ Q1 + Q2 + Q3					17.27	31.05
Isopulegol (ppm)	65864741	6	38.914	-0.293	0.066	4.80	9.80
Model	y ~ Q1					4.80	9.80
Geraniol (ppm)	c7.loc47	7	47.000	-0.203	0.067	3.46	4.20
	c9.loc35	9	35.000	0.574	0.057	26.36	41.48
Model	y ~ Q1 + Q2					28.35	45.67
Ocimene-1 (ppm)	c1.loc19	1	19.000	0.347	0.081	6.60	9.62
	19219780	6	34.430	-0.342	0.074	5.22	7.50
	12672160	8	16.951	-0.240	0.073	3.64	5.14
	c9.loc28	9	28.000	-0.296	0.075	8.16	12.10
Model	y ~ Q1 + Q2 + Q3 + Q4					21.43	36.94
Myrcene (ppm)	2452161	9	29.237	-0.295	0.072	6.71	13.45
Model	y ~ Q1					6.71	13.45
Linalyl Acetate (ppm)	13023661	6	19.883	-0.295	0.076	4.84	9.07
	74327192	9	7.660	-0.108	0.077	4.90	9.19
Model	y ~ Q1 + Q2					8.92	17.47
Sabinene (ppm)	91040066	2	81.606	0.076	0.076	5.85	10.68
	5685037	9	38.765	-0.332	0.078	5.35	9.72
Model	y ~ Q1 + Q2					10.56	20.33
α-Phellandrene (ppm)	13023661	6	19.883	-0.342	0.068	5.25	9.44
	2453522	9	33.832	-0.373	0.071	5.41	9.75
Model	y ~ Q1 + Q2					10.98	21.04
Citronellol (ppm)	38727658	1	45.656	-0.092	0.061	0.72	1.21
	1824910	9	35.259	0.356	0.056	11.43	21.56
Model	y ~ Q1 + Q2 + Q1:Q2					11.93	22.65
3-Carene (ppm)	c6.loc20	6	20.000	-0.258	0.056	6.06	11.10
	3435114	9	36.341	-0.257	0.059	4.49	8.09
Model	y ~ Q1 + Q2 + Q1:Q2					10.51	20.24
Linalool (ppm)	c6.loc1	6	1.000	-0.145	0.060	1.70	2.98
	30699989	7	35.163	-0.281	0.060	5.65	10.32
	30651759	10	80.991	0.245	0.065	4.15	7.45
Model	y ~ Q1 + Q2 + Q3 + Q1:Q2					10.45	20.14

APPENDIX:

SUPPLEMENTARY INFORMATION

Supplementary Table 2.1: Industrial Hemp Varieties grown at Fort Collins, CO and Yellow Jacket, CO during the 2016 growing season.

Variety	Latitude	Origin	Type	Supplier
CS	45	Italy	Dioecious	Consiglio per la Ricerca in Agricoltura (CREA)
Eletta Campana	45	Italy	Dioecious	Consiglio per la Ricerca in Agricoltura (CREA)
Carmaleonte	45	Italy	Monoecious	Consiglio per la Ricerca in Agricoltura (CREA)
Diana	45	Romania	Monoecious	Hempflax-Romania
Tiborszallasi	46	Hungary	Dioecious	Agromag Kft.
Monoica	46	Hungary	Monoecious	Agromag Kft.
Fédora 17	48	France	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)
Félina 32	48	France	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)
Férimon 12	48	France	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)
Futura 75	48	France	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)
Santhica 27	48	France	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)
Bialobrzeskie	52	Poland	Monoecious	Institute of Natural Fibres and Medicinal Plants
USO 31	53	Ukraine	Monoecious	Fédération Nationale des Producteurs de Chanvre (FNPC)

Supplementary Table 2.2: Procedure used for determining hemp flowering stage.

Growth stage	Procedure that was used
Beginning of female flowering	When the first flower bracts can be seen and pistils emerge.
Full female flowering	When flowers from the bottom to the top of the inflorescence (or the area of the inflorescence where female flowers exist) have reached at least the "beginning of female flowering" growth stage.
End of female flowering	When all pistils have turned red/brown or are no longer visible.
Beginning of male flowering	When the first male flower buds can be clearly identified.
Full male flowering	When approximately half of the male flowers open and release pollen.
End of male flowering	When all of the male flowers have released pollen.

Supplementary Table 2.3: Precipitation during the 2016 growing season compared with long-term averages at Fort Collins, CO and Yellow Jacket, CO.

Precipitation by Month						
Fort Collins, CO 2016 Growing Season						
	June	July	August	September	October	Season Total
Rainfall (mm)	14.5	1.3	21.8	6.9	10.7	55.1
Monthly avg. (mm)	55.1	42.9	40.9	34.0	29.0	201.9
% of avg.	26.3	3.0	53.4	20.1	36.8	27.3
Yellow Jacket, CO 2016 Growing Season						
	June	July	August	September	October	Season Total
Rainfall (mm)	1.5	65.0	41.9	26.4	1.5	136.4
Monthly avg. (mm)	15.0	39.1	41.9	39.1	50.0	185.2
% of avg.	10.2	166.2	100.0	67.5	3.0	73.6

Supplementary Table 2.4.1: Grain yield means and standard deviation by cultivar at Fort Collins, CO and Yellow Jacket, CO.

Cultivar	Environment	Mean Grain Yield (kg ha^{-1})	σ
Bialobrzeskie	FCDRY16	550	56
Bialobrzeskie	FCWET16	1471	177
Bialobrzeskie	YJWET16	519	242
Carmaleonte	FCDRY16	307	118
Carmaleonte	FCWET16	1363	179
Carmaleonte	YJWET16	796	268
CS	FCDRY16	27	24
CS	FCWET16	53	30
CS	YJWET16	774	135
Diana	FCDRY16	220	51
Diana	FCWET16	992	207
Diana	YJWET16	437	199
Eletta Campana	FCDRY16	61	5
Eletta Campana	FCWET16	138	23
Eletta Campana	YJWET16	1161	323
Fedora 17	FCDRY16	613	77
Fedora 17	FCWET16	1505	166
Fedora 17	YJWET16	633	451
Felina 32	FCDRY16	619	316
Felina 32	FCWET16	2366	775

Felina 32	YJWET16	1050	836
Ferimon 12	FCDRY16	552	173
Ferimon 12	FCWET16	1284	277
Ferimon 12	YJWET16	1040	524
Futura 75	FCDRY16	394	57
Futura 75	FCWET16	1096	329
Futura 75	YJWET16	988	285
Monoica	FCDRY16	349	47
Monoica	FCWET16	1230	134
Monoica	YJWET16	945	373
Santhica 27	FCDRY16	366	128
Santhica 27	FCWET16	1165	166
Santhica 27	YJWET16	269	135
Tiborszallasi	FCDRY16	426	109
Tiborszallasi	FCWET16	1136	108
Tiborszallasi	YJWET16	1167	423
USO 31	FCWET16	894	210
USO 31	YJWET16	382	257

Supplementary Table 2.4.2: Dry biomass means and standard deviation by cultivar at Fort Collins, CO and Yellow Jacket, CO.

Cultivar	Environment	Mean Dry Biomass (kg ha^{-1})	σ
Bialobrzeskie	FCDRY16	2827	211
Bialobrzeskie	FCWET16	6602	722
Bialobrzeskie	YJWET16	5086	664
Carmaleonte	FCDRY16	2329	255
Carmaleonte	FCWET16	7000	679
Carmaleonte	YJWET16	6871	662
CS	FCDRY16	2443	648
CS	FCWET16	5762	385
CS	YJWET16	8674	748
Diana	FCDRY16	1722	222
Diana	FCWET16	5199	663
Diana	YJWET16	6440	904
Eletta Campana	FCDRY16	2321	540
Eletta Campana	FCWET16	5960	541
Eletta Campana	YJWET16	9284	1273
Fedora 17	FCWET16	5942	365

Fedora 17	YJWET16	6144	721
Felina 32	FCWET16	7018	-
Felina 32	YJWET16	7257	1305
Ferimon 12	FCDRY16	2734	503
Ferimon 12	FCWET16	6114	471
Ferimon 12	YJWET16	6001	1125
Futura 75	FCDRY16	2881	383
Futura 75	FCWET16	7129	674
Futura 75	YJWET16	7113	1267
Monoica	FCDRY16	2608	195
Monoica	FCWET16	6607	594
Monoica	YJWET16	7831	882
Santhica 27	FCDRY16	2250	377
Santhica 27	FCWET16	5744	557
Santhica 27	YJWET16	5920	750
Tiborszallasi	FCDRY16	2702	744
Tiborszallasi	FCWET16	6412	539
Tiborszallasi	YJWET16	7535	1252
USO 31	YJWET16	4691	895

Supplementary Table 2.4.3: Days to maturity means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	Mean Days to Maturity	σ
Bialobrzeskie	FCDRY16	114	0
Bialobrzeskie	FCWET16	114	0
Carmaleonte	FCDRY16	114	0.6
Carmaleonte	FCWET16	113	0.5
CS	FCDRY16	133	0
CS	FCWET16	133	0
Diana	FCDRY16	119	0
Diana	FCWET16	119	0
Eletta Campana	FCDRY16	133	0
Eletta Campana	FCWET16	133	0
Fedora 17	FCDRY16	108	0
Fedora 17	FCWET16	110	2.3
Felina 32	FCDRY16	108	0
Felina 32	FCWET16	110	2.3
Ferimon 12	FCDRY16	114	0.6

Ferimon 12	FCWET16	113	0.5
Futura 75	FCDRY16	123	7
Futura 75	FCWET16	119	0
Monoica	FCDRY16	121	0
Monoica	FCWET16	121	0
Santhica 27	FCDRY16	114	0
Santhica 27	FCWET16	114	0
Tiborszallasi	FCDRY16	121	0
Tiborszallasi	FCWET16	121	0
USO 31	FCDRY16	93	0
USO 31	FCWET16	100	0

Supplementary Table 2.4.4: Plant height means and standard deviation by cultivar at Fort Collins, CO and Yellow Jacket, CO.

Cultivar	Environment	Mean Plant Height (cm)	σ
Bialobrzeskie	FCDRY16	135	8
Bialobrzeskie	FCWET16	174	11
Bialobrzeskie	YJWET16	113	20
Carmaleonte	FCDRY16	126	13
Carmaleonte	FCWET16	178	9
Carmaleonte	YJWET16	139	11
CS	FCDRY16	155	20
CS	FCWET16	210	7
CS	YJWET16	198	14
Diana	FCDRY16	150	7
Diana	FCWET16	194	8
Diana	YJWET16	177	21
Eletta Campana	FCDRY16	147	15
Eletta Campana	FCWET16	204	5
Eletta Campana	YJWET16	192	25
Fedora 17	FCDRY16	127	2
Fedora 17	FCWET16	158	5
Fedora 17	YJWET16	127	25
Felina 32	FCDRY16	132	13
Felina 32	FCWET16	179	7
Felina 32	YJWET16	137	9
Ferimon 12	FCDRY16	121	9
Ferimon 12	FCWET16	159	3

Ferimon 12	YJWET16	123	22
Futura 75	FCDRY16	130	16
Futura 75	FCWET16	191	12
Futura 75	YJWET16	159	8
Monoica	FCDRY16	134	6
Monoica	FCWET16	199	6
Monoica	YJWET16	169	15
Santhica 27	FCDRY16	132	7
Santhica 27	FCWET16	170	12
Santhica 27	YJWET16	154	13
Tiborszallasi	FCDRY16	139	11
Tiborszallasi	FCWET16	193	9
Tiborszallasi	YJWET16	184	17
USO 31	FCDRY16	126	3
USO 31	FCWET16	151	4
USO 31	YJWET16	123	8

Supplementary Table 2.4.5: Stem diameter means and standard deviation by cultivar at Fort Collins, CO and Yellow Jacket, CO.

Cultivar	Environment	Mean Stem Diameter (mm)	σ
Bialobrzeskie	FCDRY16	6.1	0.5
Bialobrzeskie	FCWET16	5.8	0.7
Bialobrzeskie	YJWET16	4.6	0.7
Carmaleonte	FCDRY16	6.0	0.6
Carmaleonte	FCWET16	6.3	1.1
Carmaleonte	YJWET16	7.1	0.7
CS	FCDRY16	7.6	2.1
CS	FCWET16	7.4	2.2
CS	YJWET16	7.2	0.5
Diana	FCDRY16	7.6	0.8
Diana	FCWET16	11.7	8.1
Diana	YJWET16	8.0	1.1
Eletta Campana	FCDRY16	6.4	1.8
Eletta Campana	FCWET16	9.2	1.3
Eletta Campana	YJWET16	9.2	2.9
Fedora 17	FCDRY16	4.5	0.5
Fedora 17	FCWET16	4.9	0.2
Fedora 17	YJWET16	6.5	1.0

Felina 32	FCDRY16	4.9	0.4
Felina 32	FCWET16	7.8	4.0
Felina 32	YJWET16	5.1	1.0
Ferimon 12	FCDRY16	4.6	1.2
Ferimon 12	FCWET16	5.2	0.5
Ferimon 12	YJWET16	5.6	1.2
Futura 75	FCDRY16	5.4	0.3
Futura 75	FCWET16	7.4	0.8
Futura 75	YJWET16	6.7	1.1
Monoica	FCDRY16	5.8	0.8
Monoica	FCWET16	6.7	0.3
Monoica	YJWET16	7.4	0.9
Santhica 27	FCDRY16	5.4	0.5
Santhica 27	FCWET16	6.3	0.8
Santhica 27	YJWET16	6.2	0.6
Tiborszallasi	FCDRY16	5.0	1.0
Tiborszallasi	FCWET16	6.6	1.5
Tiborszallasi	YJWET16	6.6	2.1
USO 31	FCDRY16	5.4	1.0
USO 31	FCWET16	5.4	0.7
USO 31	YJWET16	5.8	0.7

Supplementary Table 2.4.6: Stand establishment means and standard deviation by cultivar at Fort Collins, CO and Yellow Jacket, CO.

Cultivar	Environment	Mean Stand Establishment (%)	σ
Bialobrzeskie	FCDRY16	10.3	2.0
Bialobrzeskie	FCWET16	15.9	2.4
Bialobrzeskie	YJWET16	38.1	6.5
Carmaleonte	FCDRY16	13.6	2.3
Carmaleonte	FCWET16	18.5	2.0
Carmaleonte	YJWET16	32.0	6.2
CS	FCDRY16	11.8	2.8
CS	FCWET16	16.3	2.6
CS	YJWET16	32.0	4.1
Diana	FCDRY16	4.6	0.9
Diana	FCWET16	7.7	2.5
Diana	YJWET16	16.0	4.6
Eletta Campana	FCDRY16	11.2	1.8
Eletta Campana	FCWET16	15.5	2.3

Eletta Campana	YJWET16	26.3	7.0
Fedora 17	FCDRY16	16.8	2.2
Fedora 17	FCWET16	21.7	1.8
Fedora 17	YJWET16	35.1	8.2
Felina 32	FCDRY16	17.2	3.1
Felina 32	FCWET16	21.8	2.4
Felina 32	YJWET16	26.7	3.2
Ferimon 12	FCDRY16	18.3	4.9
Ferimon 12	FCWET16	24.8	2.8
Ferimon 12	YJWET16	31.6	4.9
Futura 75	FCDRY16	16.2	4.9
Futura 75	FCWET16	21.8	3.0
Futura 75	YJWET16	35.1	8.2
Monoica	FCDRY16	15.0	2.1
Monoica	FCWET16	15.0	10.6
Monoica	YJWET16	28.2	6.8
Santhica 27	FCDRY16	14.4	2.4
Santhica 27	FCWET16	19.0	1.7
Santhica 27	YJWET16	27.8	4.9
Tiborszallasi	FCDRY16	17.0	3.9
Tiborszallasi	FCWET16	20.6	1.5
Tiborszallasi	YJWET16	25.9	5.7
USO 31	FCDRY16	10.4	2.0
USO 31	FCWET16	17.5	3.0
USO 31	YJWET16	20.6	2.6

Supplementary Table 2.4.7: Carbon isotope ratio means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	δ13C (μg mg^{-1})	σ
Bialobrzeskie	FCDRY16	-23.36	0.01
Bialobrzeskie	FCWET16	-26.04	0.04
Carmaleonte	FCDRY16	-25.34	0.81
Carmaleonte	FCWET16	-27.30	0.02
CS	FCDRY16	-23.06	0.03
CS	FCWET16	-24.80	0.04
Diana	FCDRY16	-23.71	0.01
Diana	FCWET16	-27.09	0.01

Eletta Campana	FCDRY16	-23.34	0.02
Eletta Campana	FCWET16	-26.91	0.08
Fedora 17	FCDRY16	-24.85	0.33
Fedora 17	FCWET16	-26.66	0.03
Felina 32	FCDRY16	-24.26	-
Felina 32	FCWET16	-26.38	0.04
Ferimon 12	FCDRY16	-23.83	0.10
Ferimon 12	FCWET16	-28.01	0.03
Futura 75	FCDRY16	-23.62	-
Futura 75	FCWET16	-26.68	0.02
Monoica	FCDRY16	-24.38	0.02
Monoica	FCWET16	-26.14	0.06
Santhica 27	FCDRY16	-25.00	0.02
Santhica 27	FCWET16	-26.50	0.00
Tiborszallasi	FCDRY16	-26.99	0.04
Tiborszallasi	FCWET16	-24.73	0.10
USO 31	FCDRY16	-24.37	0.03
USO 31	FCWET16	-26.49	0.06

Supplementary Table 2.4.8: Total Potential Δ9THC means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	Total Potential Δ9THC (%)	σ
Bialobrzeskie	FCDRY16	0.182	0.098
Bialobrzeskie	FCWET16	0.117	0.042
Carmaleonte	FCDRY16	0.060	0.015
Carmaleonte	FCWET16	0.096	0.012
CS	FCDRY16	0.199	0.031
CS	FCWET16	0.279	0.055
Diana	FCDRY16	0.011	0.008
Diana	FCWET16	0.019	0.015
Eletta Campana	FCDRY16	0.157	0.065
Eletta Campana	FCWET16	0.200	0.040
Fedora 17	FCDRY16	0.080	0.041
Fedora 17	FCWET16	0.078	0.015
Felina 32	FCDRY16	0.077	0.009
Felina 32	FCWET16	0.114	0.041
Ferimon 12	FCDRY16	0.069	0.022
Ferimon 12	FCWET16	0.092	0.010
Futura 75	FCDRY16	0.089	0.027
Futura 75	FCWET16	0.113	0.019
Monoica	FCDRY16	0.071	0.011
Monoica	FCWET16	0.131	0.014
Santhica 27	FCDRY16	0.002	0.003
Santhica 27	FCWET16	0.002	0.002
Tiborszallasi	FCDRY16	0.471	0.501
Tiborszallasi	FCWET16	0.798	0.802
USO 31	FCDRY16	0.017	0.013
USO 31	FCWET16	0.020	0.020

Supplementary Table 2.4.9: Total Potential CBD means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	Total Potential CBD (%)	σ
Bialobrzeskie	FCDRY16	1.101	0.295
Bialobrzeskie	FCWET16	1.839	0.795
Carmaleonte	FCDRY16	1.257	0.386
Carmaleonte	FCWET16	2.188	0.360
CS	FCDRY16	3.766	0.767

CS	FCWET16	5.949	1.537
Diana	FCDRY16	0.110	0.168
Diana	FCWET16	0.318	0.250
Eletta Campana	FCDRY16	2.775	1.089
Eletta Campana	FCWET16	4.062	1.013
Fedora 17	FCDRY16	1.312	0.338
Fedora 17	FCWET16	1.680	0.394
Felina 32	FCDRY16	1.506	0.142
Felina 32	FCWET16	2.422	0.978
Ferimon 12	FCDRY16	1.189	0.137
Ferimon 12	FCWET16	2.126	0.363
Futura 75	FCDRY16	1.656	0.674
Futura 75	FCWET16	2.275	0.549
Monoica	FCDRY16	1.368	0.234
Monoica	FCWET16	2.684	0.395
Santhica 27	FCDRY16	0.003	0.004
Santhica 27	FCWET16	0.008	0.003
Tiborszallasi	FCDRY16	2.236	0.295
Tiborszallasi	FCWET16	3.218	0.622
USO 31	FCDRY16	0.325	0.235
USO 31	FCWET16	0.401	0.437

Supplementary Table 2.4.10: CBC means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	CBC (%)	σ
Bialobrzeskie	FCDRY16	0.013	0.007
Bialobrzeskie	FCWET16	0.006	0.004
Carmaleonte	FCDRY16	0.005	0.001
Carmaleonte	FCWET16	0.005	0.001
CS	FCDRY16	0.002	0.000
CS	FCWET16	0.003	0.001
Diana	FCWET16	0.000	-
Eletta Campana	FCDRY16	0.001	0.000
Eletta Campana	FCWET16	0.002	0.001
Fedora 17	FCDRY16	0.008	0.005
Fedora 17	FCWET16	0.003	0.002
Felina 32	FCDRY16	0.002	0.001
Felina 32	FCWET16	0.011	0.011
Ferimon 12	FCDRY16	0.006	0.005

Ferimon 12	FCWET16	0.007	0.002
Futura 75	FCDRY16	0.006	0.004
Futura 75	FCWET16	0.008	0.004
Monoica	FCDRY16	0.003	0.002
Monoica	FCWET16	0.000	0.000
Santhica 27	FCDRY16	0.007	0.007
Santhica 27	FCWET16	0.002	0.001
Tiborszallasi	FCDRY16	0.001	0.001
Tiborszallasi	FCWET16	0.001	0.001
USO 31	FCDRY16	0.000	-
USO 31	FCWET16	0.001	-

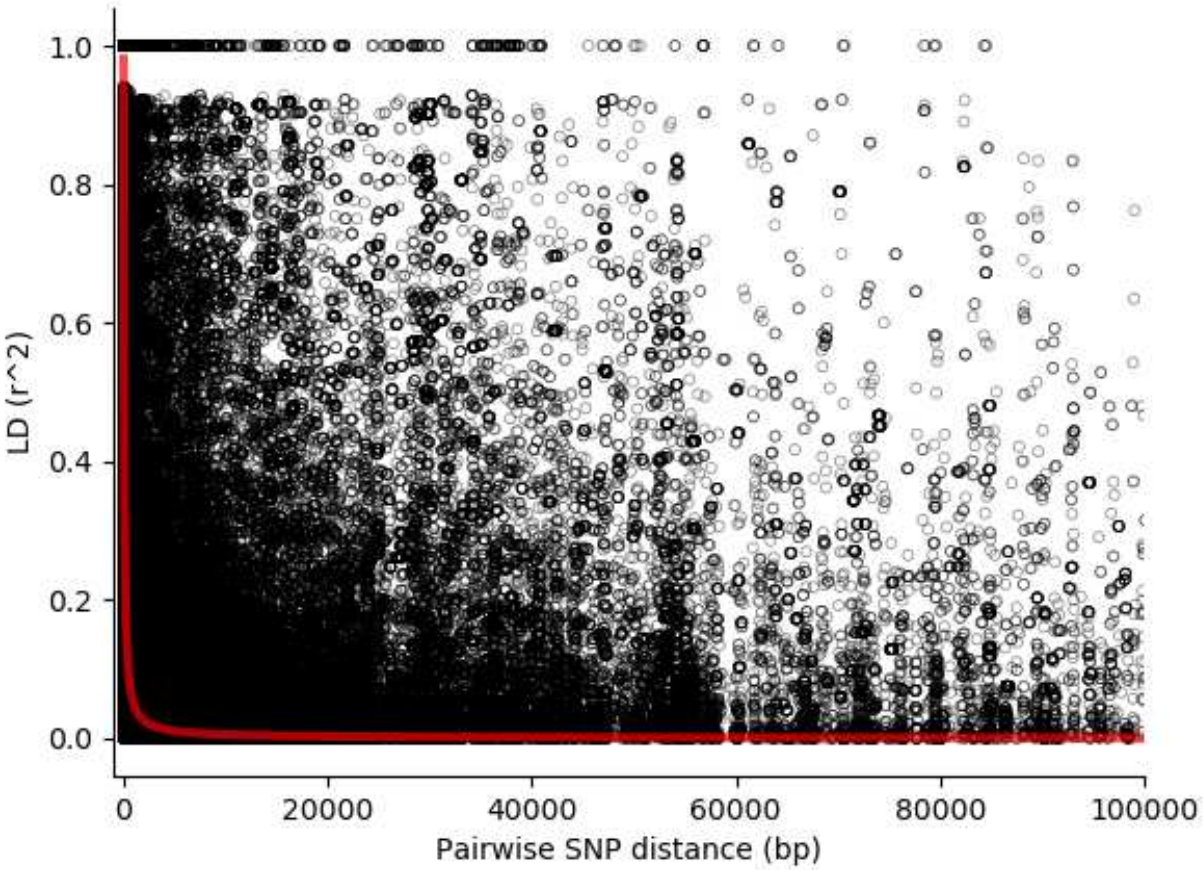
Supplementary Table 2.4.11: α-pinene means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	α-pinene (%)	σ
Bialobrzeskie	FCDRY16	0.030	0.017
Bialobrzeskie	FCWET16	0.008	0.006
Carmaleonte	FCDRY16	0.014	0.005
Carmaleonte	FCWET16	0.024	0.014
CS	FCDRY16	0.020	0.006
CS	FCWET16	0.028	0.013
Diana	FCDRY16	0.004	0.001
Diana	FCWET16	0.005	0.004
Eletta Campana	FCDRY16	0.026	0.011
Eletta Campana	FCWET16	0.020	0.004
Fedora 17	FCDRY16	0.016	0.006
Fedora 17	FCWET16	0.009	0.004
Felina 32	FCDRY16	0.021	0.015
Felina 32	FCWET16	0.009	0.004
Ferimon 12	FCDRY16	0.022	0.009
Ferimon 12	FCWET16	0.007	0.002
Futura 75	FCDRY16	0.023	0.011
Futura 75	FCWET16	0.030	0.019
Monoica	FCDRY16	0.017	0.006
Monoica	FCWET16	0.019	0.004
Santhica 27	FCDRY16	0.004	0.002
Santhica 27	FCWET16	0.003	0.002
Tiborszallasi	FCDRY16	0.029	0.006

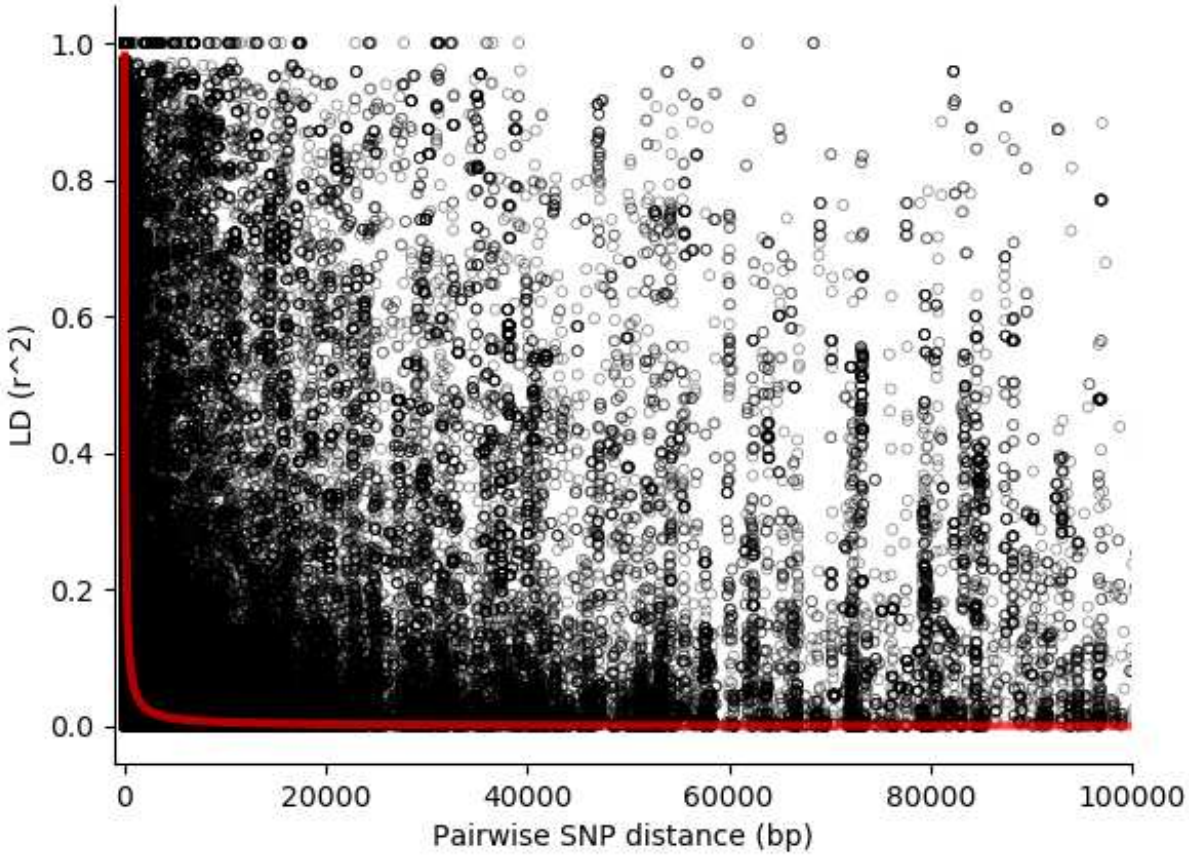
Tiborszallasi	FCWET16	0.033	0.010
USO 31	FCDRY16	0.012	0.005
USO 31	FCWET16	0.005	0.004

Supplementary Table 2.4.12: β-pinene means and standard deviation by cultivar at Fort Collins, CO.

Cultivar	Environment	β-pinene (%)	σ
Bialobrzeskie	FCDRY16	0.005	0.002
Bialobrzeskie	FCWET16	0.002	0.001
Carmaleonte	FCDRY16	0.003	0.001
Carmaleonte	FCWET16	0.006	0.003
CS	FCDRY16	0.006	0.002
CS	FCWET16	0.010	0.004
Diana	FCDRY16	0.001	0.001
Diana	FCWET16	0.001	0.001
Eletta Campana	FCDRY16	0.007	0.003
Eletta Campana	FCWET16	0.006	0.001
Fedora 17	FCDRY16	0.003	0.001
Fedora 17	FCWET16	0.002	0.001
Felina 32	FCDRY16	0.004	0.003
Felina 32	FCWET16	0.002	0.001
Ferimon 12	FCDRY16	0.004	0.002
Ferimon 12	FCWET16	0.002	0.000
Futura 75	FCDRY16	0.004	0.002
Futura 75	FCWET16	0.005	0.003
Monoica	FCDRY16	0.004	0.002
Monoica	FCWET16	0.005	0.001
Santhica 27	FCDRY16	0.000	0.000
Santhica 27	FCWET16	0.000	0.000
Tiborszallasi	FCDRY16	0.007	0.001
Tiborszallasi	FCWET16	0.008	0.002
USO 31	FCDRY16	0.003	0.001
USO 31	FCWET16	0.001	0.002



Supplemental Figure 3.1.1: Linkage disequilibrium decay in hemp samples. Pairwise disequilibrium coefficients (r^2) in the genome. Black circles denote the observed values. Red line denotes the curve resulting from least squares fitting.



Supplemental Figure 3.1.2: Linkage disequilibrium decay in high-THC *Cannabis* samples Pairwise disequilibrium coefficients (r^2) in the genome. Black circles denote the observed values. Red line denotes the curve resulting from least squares fitting.

Supplementary Table 4.1: Precipitation during the 2017 growing season compared with long-term averages at Fort Collins, CO.

Precipitation by Month						
Fort Collins, CO 2017 Growing Season						
	June	July	August	September	October	Total
Rainfall (mm)	3.0	45.0	54.4	42.9	32.0	177.3
Monthly avg. (mm)	55.1	42.9	40.9	34.0	29.0	201.9
% of avg.	5.5	104.7	132.9	126.1	110.5	87.8

Supplementary Table 4.2.1: Agronomic trait correlations (r-values) at Fort Collins, CO.

	Days to Flower	Days to Maturity	ortion of Female Flowers	Grain Yield (g)	Thousand Seed Mass (g)	Total Biomass (g)	Stem Biomass (g)	Stem Diameter (cm)	Leaf Water Content (g)	Plant Height (cm)
Days to Flower	1									
Days to Maturity	0.37 ***	1								
ortion of Female Flowers	0.021 NS	0.052 NS	1							
Grain Yield (g)	0.09 NS	0.23 ***	0.10 NS	1						
Thousand Seed Mass (g)	0.01 NS	0.09 NS	0.06 NS	0.37 ***	1					
Total Biomass (g)	0.21 ***	0.41 ***	0.08 NS	0.82 ***	0.23 ***	1				
Stem Biomass (g)	0.29 ***	0.41 ***	0.08 NS	0.67 ***	0.15 *	0.92 ***	1			
Stem Diameter (cm)	0.20 ***	0.30 ***	0.10 NS	0.42 ***	0.18 **	0.58 ***	0.58 ***	1		
Leaf Water Content (g)	0.29 ***	0.32 ***	0.10 NS	0.29 ***	0.08 NS	0.37 ***	0.37 ***	0.36 ***	1	
Plant Height (cm)	0.27 ***	0.29 ***	0.08 NS	0.60 ***	0.08 NS	0.77 ***	0.77 ***	0.74 ***	0.56 ***	1

NS non significant,* Significant at the 0.05 probability level; ** Significant at the 0.01 probability level; * Significant at the 0.001 probability level**

Supplementary Table 4.2.2: Cannabinoid trait correlations (r-values) at Fort Collins, CO.

	THCA	THC	CBDA	CBD	CBGA	CBG
THCA	1					
THC	0.60 ***	1				
CBDA	0.92 ***	0.59 ***	1			
CBD	0.66 ***	0.47 ***	0.80 ***	1		
CBGA	0.10 NS	0.12 NS	0.08 NS	0.04 NS	1	
CBG	0.12 NS	0.13 NS	0.09 NS	0.05 NS	0.67 ***	1

NS non significant,* Significant at the 0.05 probability level; ** Significant at the 0.01 probability level; * Significant at the 0.001 probability level**

Supplementary Table 4.2.3: Terpene trait correlations (r-values) at Fort Collins, CO.

	α-Pinene	β-Caryophyllene	Caryophyllene Oxide	Terpinolene	Humulene	Ocimene-2	Limonene	α-Bisabolol	α-Terpenine	ϒ-Terpenine	Eucalyptol	Camphene	Geraniol	Ocimene-1	Myrcene	inalyl Acetat	Sabinene	-Phellandren	Citronellol	3-Carene	Linalool
α-Pinene	1																				
β-Caryophyllene	0.73 ***	1																			
Caryophyllene Oxide	0.33 ***	0.34 ***	1																		
Terpinolene	0.75 ***	0.86 ***	0.30 ***	1																	
Humulene	0.75 ***	0.94 ***	0.38 ***	0.91 ***	1																
Ocimene-2	0.82 ***	0.51 ***	0.19 **	0.61 ***	0.54 ***	1															
Limonene	0.52 ***	0.61 ***	0.55 ***	0.44 ***	0.55 ***	0.29 ***	1														
α-Bisabolol	0.88 ***	0.75 ***	0.35 ***	0.82 ***	0.82 ***	0.69 ***	0.52 ***	1													
α-Terpenine	0.89 ***	0.74 ***	0.34 ***	0.80 ***	0.83 ***	0.71 ***	0.49 ***	0.97 ***	1												
ϒ-Terpenine	0.44 ***	0.19 **	0.12 NS	0.11 NS	0.16 *	0.43 ***	0.33 ***	0.38 ***	0.41 ***	1											
Eucalyptol	0.86 ***	0.41 ***	0.18 **	0.45 ***	0.44 ***	0.80 ***	0.35 ***	0.74 ***	0.74 ***	0.50 ***	1										
Camphene	0.33 ***	-0.03 NS	0.41 ***	-0.01 NS	-0.03 NS	0.31 ***	0.24 ***	0.20 **	0.20 **	0.27 ***	0.41 ***	1									
Geraniol	-0.11 NS	-0.21 **	0.35 ***	-0.14 *	-0.18 **	-0.06 NS	-0.08 NS	-0.15 *	-0.16 *	-0.15 *	-0.04 NS	0.51 ***	1								
Ocimene-1	0.73 ***	0.81 ***	0.37 ***	0.85 ***	0.88 ***	0.54 ***	0.50 ***	0.83 ***	0.82 ***	0.11 NS	0.48 ***	0.05 NS	-0.11 NS	1							
Myrcene	0.43 ***	0.50 ***	0.11 NS	0.54 ***	0.52 ***	0.31 ***	0.21 **	0.39 ***	0.38 ***	0.03 NS	0.27 ***	-0.04 NS	-0.12 NS	0.40 ***	1						
Linalyl Acetate	0.26 ***	-0.02 NS	0.01 NS	0.01 NS	-0.03 NS	0.40 ***	0.00 NS	0.22 **	0.19 **	0.25 ***	0.43 ***	0.24 ***	0.00 NS	0.12 NS	-0.07 NS	1					
Sabinene	0.65 ***	0.53 ***	0.55 ***	0.59 ***	0.62 ***	0.47 ***	0.52 ***	0.79 ***	0.78 ***	0.35 ***	0.48 ***	0.28 ***	-0.06 NS	0.61 ***	0.24 ***	0.06 NS	1				
α-Phellandrene	0.71 ***	0.57 ***	0.10 NS	0.71 ***	0.59 ***	0.70 ***	0.26 ***	0.67 ***	0.67 ***	0.25 ***	0.65 ***	0.06 NS	-0.15 *	0.66 ***	0.40 ***	0.25 ***	0.38 ***	1			
Citronellol	0.00 NS	-0.12 NS	0.50 ***	-0.09 NS	-0.11 NS	0.01 NS	0.12 NS	-0.02 NS	-0.05 NS	-0.03 NS	0.05 NS	0.67 ***	0.85 ***	-0.04 NS	-0.08 NS	0.07 NS	0.15 *	-0.10 NS	1		
3-Carene	0.55 ***	0.14 *	0.06 NS	0.12 NS	0.12 NS	0.53 ***	0.24 ***	0.42 ***	0.43 ***	0.54 ***	0.65 ***	0.39 ***	-0.09 NS	0.20 **	0.02 NS	0.49 ***	0.23 ***	0.21 **	-0.02 NS	1	
Linalool	0.31 ***	0.04 NS	0.48 ***	0.03 NS	0.02 NS	0.24 ***	0.38 ***	0.25 ***	0.21 **	0.28 ***	0.34 ***	0.49 ***	0.32 ***	0.11 NS	-0.01 NS	0.25 ***	0.43 ***	0.06 NS	0.53 ***	0.31 ***	1