Title

Investigation into the Genetic Basis of Day Neutrality in Cannabis sativa L. through a Genome Wide Association Study

Author

Spivak, Steven

Publication Date

2022

Peer reviewed|Thesis/dissertation

Investigation into the Genetic Basis of Day Neutrality in *Cannabis sativa* L. through a Genome Wide Association Study

By

STEVEN SPIVAK
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Horticulture and Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

E. Charles Brummer, Chair

_____

Allen Van Deynze

_____

Daniel H. Putnam

Committee in Charge

2022

Acknowledgments

Thanks to all the help from my lab peers who were there to help me when I needed them: Daniel Pap, Allison-Krill Brown, Oon Ha Shin, Scott Newell, Matt Francis, Kreingkrai "Moo" Nonkum, and Sheriff-Seedy Phaal.

Thanks to the UC Davis professors and their teaching assistants whose classes and discussions helped me form the thoughts I have written down in this thesis: Edward Charles Brummer, Daniel H. Putnam, Allen Van Deynze, Patrick H. Brown, Steve Knapp, Michael Miller, Daniel Runcie.

And lastly, special thanks to Jenna Bess, who was with me even at my lowest points.

Abstract

Hemp (*Cannabis sativa* L.), which can be legally grown in the U.S. since 2018, is a multi-use crop producing fiber, seed, oil, and/or essential oils. A better understanding of the genetic basis of major traits will enable plant breeders to improve the crop more effectively. One trait of importance to hemp producers is crop maturity. Hemp is naturally a short-day plant, flowering in autumn when photoperiod declines below 12-13 hours. Day neutral types of hemp will flower at roughly the same time regardless of photoperiod, making this trait desirable for many production reasons. In this experiment, we hybridized day neutral and day sensitive hemp germplasm to locate the day neutral trait in the hemp genome. We conducted a multi-year, multi-environment experiment in which we scored individual plants on a '0' to '5' scale based on plant maturity and computed the average growing degree days at which each plant scored '3,' the beginning of flowering. We also obtained the day length at this time. In parallel, we obtained over 50,000 single nucleotide polymorphism (SNP) from genome sequencing data on 1507 individual plants. We used the phenotype and genotype data to conduct a genome wide association study using a mixed linear model. From our analysis we identified 54 SNPs were associated with our two traits, which shared the five most significant SNPs on Chromosome 2. Within a 21Mb region of our shared significant SNP hits we identified a gene: flowering locus K homology domain (FLK), a known down regulator of CONSTANS. Most plants we evaluated derived from paired crosses between divergent germplasm, resulting in large disequilibrium which resulted in a broad chromosomal region in which the day neutral flowering regulatory gene likely lies. While more work is needed, this experiment identifies the likely region, which controls day neutral flowering in hemp, a first step toward developing genetic markers to facilitate the incorporation of this trait into other germplasm.

Investigation into the Genetic Basis of Day Neutrality in Cannabis Sativa L. through a Genome Wide Association Study

Literature Review

Industrial hemp (*Cannabis sativa* L.)., has been an agronomically important crop since 2700 BC in China. Today, it serves a purpose in a variety of different industries, such as pharmaceuticals, nutraceuticals, textiles, composite materials, biofuels, foods, cosmetics, and hygiene products (Andre et al., 2006; Crini et al., 2020; Rupasinghe et al., 2020). Hemp is one of humanity's earliest domesticated plants going back to the Neolithic times in parts of East Asia (Ren et al., 2021). Hemp is the non-psychoactive form of *Cannabis*, differentiated from marijuana only by having less than 0.3% tetrahydrocannabinol (THC) concentration in dry mass (FDA, 2019). In 1970, the Controlled Substances Act was passed in the United States, which stated that all *Cannabis sativa*, psychoactive or not, was a Schedule 1 drug with "high abuse potential with no accepted medical use; medications within this schedule may not be prescribed, dispensed, or administered" (Gabay, 2013). The passage of the Controlled Substances Act forbade any individual from researching or growing *Cannabis* in any form, including hemp, and it was not until forty-eight years later with the passage of the 2018 Farm Bill (S.2667 - Hemp Farming Act of 2018) that researchers and growers could again study and grow hemp. With 48 years of absence from the scientific literature, the renewed interest in hemp as a crop with high agronomic value has stimulated significant research activity.

Abbreviations: THC, Tetrahydrocannabinol; CBD, Cannabidiol; CBC, Cannabichromene; CBG, Cannabigerol; THCA, Tetrahydrocannabinolic acid; CBDA, Cannabidiolic acid; CBCA, Cannabichromenic acid; CBGA, Cannabigerolic acid; QTL, Quantitative trait locus; GWAS, Genome-Wide Association Study; GDD, Growth degree days; TOF, Time of Flower

Taxonomy

All *Cannabis* germplasm is completely interfertile, suggesting that the genus consists of a single species, with two subspecies (subsp. *sativa* (hemp) and subsp. *indica* (marijuana) (Small and Cronquist, 1976). Through DNA analysis, evidence pointed towards *C. sativa, C. indica*, *C. ruderalis* populations of *Cannabis* as being distinct subspecies of *Cannabis sativa* (McPartland 2018). Common vernacular terms, *sativa* and *indica*, have been used to describe different types of marijuana, causing much confusion for consumers who believe that these different forms of *Cannabis* happen to function in different ways. Generally, *sativa* types are plants with tall and slender morphology, narrow leaflets, and late maturation while *indica* types have shortened stature, broad leaflets, and early maturation (de Meijer and van Soest 1991). Whatever true population differences there may have been between indica and sativa types of *C. sativa subsp. indica* have become lost over generations of repeated hybridization events (Small 2017), although clearer population differentiation between hemp and marijuana still exist (Sawler et al., 2018; Dufresnes et al., 2017) . Thus, industrial hemp is simply classified taxonomically as *Cannabis sativa* L. with low amounts of THC (Formato et al., 2020).

Chemotypes

One way to classify genotypes or populations of *C. sativa* is through their chemotypes, or chemical phenotypes, based on the predominant cannabinoids in the plant, in particular, cannabidiol (CBD), cannabigerol (CBG), and tetrahydrocannabinol (THC) (Pacifico et al., 2008; Aizpurua-Olaizola et al. 2016; Welling et al. 2016; Richins et al. 2018; Jin et al. 2021). *C. sativa* is divided into five different chemotypes: THC dominant, referred to as the drug-type, CBD dominant, referred to as fiber-type, and one which THC and CBD are present in equal proportion. Two other chemotypes, CBG dominant and those which contain low concentrations

of cannabinoids, are less frequently used in scientific literature (Richins et al. 2018). The classification of *C. sativa* germplasm with chemotypes enables scientists to easily classify individuals for breeding for certain uses like pharmacology, fiber, or seed (Welling et al. 2008, Stack et al. 2021).

Hemp Morphology and Life Cycle

Cannabis sativa L. (2n=20) is a dioecious plant with individuals from the male and female sexes having distinct morphological differences (Clarke, 1981; Van der Werf and Van den Berg, 1995; Clarke and Merlin, 2016). Male individuals are described as slenderer in stature than their female counterparts and have less reproductive biomass compared to a females' dense inflorescence (Bócsa and Karus, 1998; Struik et al., 2000; Petit et al., 2020). Differences in morphology, specifically between the sexes of hemp plants, only become apparent after the seedling stage (Campbell et al., 2021). The genetic basis for sex in *Cannabis* is determined by the inheritance of either an X or Y chromosome from the male parent.

Female plants have a sticky inflorescence that captures wind-dispersed pollen from male plants (Clarke 1981; Clarke and Merlin 2016). Male flowers, within their hanging panicles, have a perianth of five sepals that surround the androecium; the anthers at maturity split lengthwise, releasing the pollen grains (Moliterni et al., 2004). Female flowers develop as thick clusters called racemes, and receive the pollen grains through insect, wind, or mechanical dispersion onto the pistils. In production settings, formation of seeds is undesirable if the use of hemp is the extraction of essential oils or the sale of the flowers themselves; consequently, most hemp producers prefer to grow only female individuals and avoid fertilization from male plants (Nackley et al., 2020).

Most accessions of hemp are daylength sensitive, with flowering occurring when the critical photoperiod threshold of between 12-14 hours occurs. In contrast, day neutral (so-called "autoflower") accessions will flower regardless of day length depending on accumulated growth degree days (GDD) (Lisson et al., 2000; Struik, 2000; Petit et al., 2020).

Cannabinoid Functionality within Hemp

Grandular trichomes, a form of sessile trichomes, cover the surface of female flowering tissues and produce cannabinoid oils (Tannet et al., 2021). The secreted oils which burst from these trichome sacs coat the surface with a sticky resin, which results in flowers that are waxy in texture (Garrett and Hunt 1974, Mauseth 2006). Cannabinoids likely serve multiple purposes within the *Cannabis* plant, such as a defense response against herbivory from insects and the dissipation of heat stress in the environment. Cannabinoids are produced in substantially higher quantities when exposed to UV-B radiation (280-315 nm), suggesting they act as a barrier against the damaging effects of UV-B radiation (Bilodeau et al. 2019). The range of cannabinoid concentrations among hemp plants varies in response to heat stress (Toth et al., 2021)

Cannabinoid Synthesis

There are over 180 different cannabinoids present in *C. sativa*, with the primary cannabinoids being THC, also called Δ9-THC, CBD, and CBC (Tahi et al., 2021). These cannabinoids coexist along with their acid-precursors: tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), and cannabichromenic acid (CBCA). The acid forms of these cannabinoids change into their decarboxylated forms primarily from the application of light- and/or heat onto the harvested crop, but the decarboxylated forms still exist at certain levels within the flowering tissue before harvesting (Formato et al., 2020). All acid forms of

cannabinoids come from a primary precursor phytocannabinoid called cannabigerolic acid (CBGA, Taura et al., 2007; Formato et al., 2020; Tahi et al., 2021). CBGA is a product of two metabolites, olivetolic acid and geranyl diphosphate, which are formed from the polyketide pathway and plastidial deoxyxylulose phosphate/methyl-erythritol phosphate pathway (DOXP/MEP pathway), respectively (Jin et al., 2020).

Economic Value of Hemp

Cannabinoids have been used as a component of human therapeutic medicine for thousands of years (Russo et al 2007; National Academies of Sciences 2017). Specific cannabinoids have been used to minimize chronic pain, improve sleep quality, and treat a wide variety of other ailments. *C. sativa* has obtained more attention over the last 30 years as a source of medicine in America after California passed the Compassionate Use Act of 1996, a bill which allowed the state to provide patients with access to medical marijuana "in the treatment of cancer, anorexia, AIDS, chronic pain, spasticity, glaucoma, arthritis, migraine, or any other illness for which marijuana provides relief" (CA GOV 1996). Since then, there has been increasing interest in the use of cannabinoids as a way of minimizing pain, a malady affecting 1 in 5 Americans on a daily basis (CDC 2018). CBD is of particular interest because it can provide pain relief without the psychoactive effects that come with other cannabinoids like THC (Piomelli and Russo, 2016). The global value of CBD products is 2.8 billion dollars as of 2020 and is expected to increase on the order of 20-23% year-to-year over the next five years (Grand View Research, 2021).

In America, *Cannabis* was a widely grown crop for many years during the 18th and 19th centuries before being replaced by cotton as the predominant crop used for textiles (Hopkins, 2014). With increased interest in the crop medically for its oils, it has also renewed interest in its

fiber (hurd and bast), oil, and seed (Ahmed et al., 2022). Hemp bast, the long fibers from the outer stem of hemp, can be used to make carpets, shoes, diapers, insulation, yarn, composite materials, and plastics (Johnson, 2018). The hurd, or inner core fibers of the hemp plant, can be used to create hempcrete, animal bedding, potting mix, and soil amendments (Ahmed et al., 2022). With a wide variety of uses, industrial hemp fibers, as of 2019, have a current market value of $4.46 billion with an expected compound annual growth rate of 33% through 2027 to total $43.8 billion (Verified Market Research, 2020). With the value of industrial hemp between its fiber and CBD products currently being valued at over $7.2 billion, there is increasing interest in agronomic improvements for the crop.

Hemp Production

Hemp is typically grown in field settings but can be grown in greenhouses when growers are focusing on greater pest control, year-round growth, and control of lighting regiment. In field settings, hemp generally is grown on well-drained soil which is high in fertility (Cherney and Small, 2016; Rupasinghe et al., 2020). Planting density varies among varieties and the grower's intended use for the crop. Hemp grown for oil extraction requires wider spacing to promote branching and flower development; planting densities for the purpose of harvesting its fiber is typically double that of oilseed varieties, however, in general the architecture of the plant itself is strongly associated with the density of planting, nutritionally availability within the soil, and the length of day that the plant is exposed too during its life cycle (Campiglia and Radicetti and Mancinelli, 2017). Hemp is either planted by direct seeding or through the transplantation of seedlings or clones (Cherney and Small, 2016; Adesina et al., 2020; Rupasinghe et al., 2020).

When hemp is grown for medicinal oil extraction, the field typically consists of only female plants. Usually this is accomplished by using "feminized" seed, which is produced from

female plants that have been induced to produce male gametes and seed using either chemical or environmental stress which results in seeds which will produce seeds which result mostly in female plants (Cherney and Small, 2016; Punja and Homes, 2020). This process of masculizing female hemp plants on an industrial scale is done through foliar applications of silver thiosulfate (Lubell and Brand, 2018; Kurtz et al., 2020). The rationale for producing feminized seeds, outside of maintaining genetically identical inbred lines for commercial sale, is to minimize the production of seeds within the flowering tissue of female plants due to consumer, grower, and processor preferences.

Historical Breeding Practices

Hemp improvement can be done through the use of phenotypic recurrent selection, aka mass-selection, by selecting the best individual plants based on field performance and using their seed for the next evaluation cycle (Salentijn et al., 2019; Barcaccia et al., 2020). Once an elite variety has been developed, backcrossing can be used to incorporate new traits from undesirable germplasm (Zargar et al. 2015). For traits such as fiber quantity and quality, with high trait heritability, mass-selection can work well (Hennick et al., 1994). Other breeding practices have been employed for the purpose of either increasing variation, specifically when crossbreeding individuals, or for fixing a trait through inbreeding to produce inbred lines and/or to capitalize on heterosis of F1 hybrid cultivars (Ranalli, 2004; Salentijn et al., 2019).

Advances in Sequencing for Breeding Improvements

Advances in biotechnology can accelerate the incorporation of a trait into an existing population with high accuracy and speed using next generation sequencing (NGS) and marker assisted selection (MAS). Marker assisted selection works to track the phenotype of an

individual plant by associating genetic polymorphisms with trait variation, enabling selection on seedling plants without having to grow the individual to maturity (Collard and Mackill, 2007; Xu and Crouch 2008; Zargar et al., 2015). Recent advances in sequencing technology, specifically with the development of next-generation sequencing (NGS), has reduced the price of sequencing whole genomes of individual plants (Gasparyan 2019; Zargar et al. 2015), making marker identification and use more tractable.

Genetic Basis for Day-Neutrality in Hemp and Other Crops

Day neutrality is a trait which is present in many agricultural crops such as soybeans, wheat, barley, rice, tomatoes, strawberries, and alfalfa (Polson, 1972; Mizoguchi et al., 2007; Rowley et al., 2011; Turner et al., 2013; Adhikari et al., 2019). Several genes involved light sensing contribute to differences among genotypes in time of flower (TOF) and in day neutrality; the major genes appear to be conserved across species. For instance, day-neutrality is controlled by CONSTANS (CO), a gene which encodes a transcription factor involved in the transduction of light signals, promoting the expression of other genes downstream (Morris et al., 2010; Denoyes et al., 2020).

The genetic basis for the change from vegetative growth to flowering within hemp cultivars is mostly unknown. Petit et al. (2020), identified six QTLs related to genes which control the perception and transduction of light and their associated transcription factors). However, hemp germplasm has a wide range of flowering times, affected by both genetics and the environment (Petit et al. 2020; Dowling et al. 2021).

The objective of this experiment was to identify the genomic location of gene(s) associated with flowering time in hemp as a prelude to making genetic markers that could be routinely used to screen germplasm in a breeding program.

Investigation into the Genetic Basis of Day Neutrality in Cannabis Sativa L. through a Genome Wide Association Study

Introduction

Hemp, *Cannabis sativa* L. is a valuable medicinal, fiber, seed, and oil crop. Understanding and manipulating the flowering time of hemp could facilitate cultivar development for diverse environments and cropping systems. The time at which hemp transitions from vegetative growth to flowering is critical for the development and quality of the final harvested product. Most hemp germplasm requires short days to begin flowering and producing seed, oil, and cannabinoid products (Lisson et al., 2000; Jung and Müller, 2009; Salentijn et al., 2015; Blümel et al., 2015; Petit et al., 2020; Dowling et al., 2021). Before the transition to flowering can occur, a period of vegetative growth stage controlled by the accumulation of thermal time in the environment is necessary (Gordon and Bootsma, 1993; Lisson et al., 2000; Salentijn et al., 2019; Petit et al., 2020). Vegetative growth for hemp is optimal around 30^{o}C and continues to a maximum temperature about 42^{o}C. Around 300 to 600 units of cumulated thermal time over 1^{o}C must occur for the plant to be able to initiate flowering if the critical daylength is reached (Petit et al., 2020; Salentijn et al., 2019; Janes et al., 2017).

While typically a short-day plant, hemp germplasm has considerable variation in TOF (Spitzer-Rimon, 2019). Some hemp germplasm has day-neutral flowering, which means that regardless of photoperiod, they will flower after a certain amount of thermal time has accumulated (Castro et al., 2015; Punja, 2017; Petit et al., 2020). Day-neutral *Cannabis* varieties can go from seed to flowering in a soon as four weeks, and some varieties can be harvested within 100 days after seeding. The trait for day-neutrality originated from the *ruderalis* variety of *Cannabis sativa* in parts of Southern Russia. These plants are described as being short and stalky

in nature, while producing small amounts of flower with low concentrations of THC (Gloss, 2015; Ivanova, 2019). Despite its undesirable attributes, *ruderalis* has been used to introgress day neutrality (also called "autoflower") into high yielding flower and oil varieties. The day neutral trait can provide significant agronomic benefits by standardizing harvest time and potentially enabling growers to have more harvest cycles in one year compared to those growing day-sensitive varieties of hemp which takes much longer to flower in most environments.

The genetic basis for day neutrality, specifically in hemp, is currently unknown. We hypothesized that at least one major locus controls day neutral flowering with multiple minor loci affecting the variation in time to flower within day neutral or daylength sensitive germplasm. To test this hypothesis, we hybridized day neutral and daylength sensitive accessions of hemp and evaluated the progeny for flowering time across three field trials over two years in order and genotyped them using next generation sequencing technology to identify genetic loci involved in flowering time.

Materials and Methods

Germplasm

Several hemp germplasms used in the UC Davis hemp breeding program were included in the experiment (Supplementary Table 4). Only female genotypes were used for this experiment, with pollen produced following treatment with silver thiosulfate. Crosses between day neutral ('KG01', 'KG02', 'NUT') and daylength sensitive genotypes were conducted in the greenhouse during the winters of 2019-2020 and 2020-2021. Crosses were made in a three-way crossing fashion, with chemical-type day neutral cultivars being crossed with chemical-type daylength sensititive cultivars and hybrids being crossed together or to a seed-type hemp cultivar. This was done in order to create segregation of our phenotypes within families.

Seedling Preparation and Transplanting

Hemp seeds were placed into 72-cell flats containing potting soil mix, composed of mostly perlite, and were then covered in vermiculite and watered daily. Seedlings were grown for two weeks in a greenhouse with temperatures ranging between 22-35^{o} C following a 12-hour lighting regiment emitted by high-pressure sodium lightbulbs. Fertilizer, Nu-Crop (8-24-6), was incorporated in the irrigation water. After two weeks, seedlings were moved outside for three days to harden them to field conditions before transplanting. Seedlings were transplanted into the field at a depth of 5cm, covering the cotyledons.

Field Experimental Designs

We grew three trials, designated as "environments," planted in July 2020, May 2021, and July 2021 at the UC Davis Plant Sciences Farm in Davis, CA. The 2020 trial was designed as a randomized complete block design with two replications. The trial contained 132 entries, with

three entries acting as controls, two of which were day neutral and early flowering ('KG01' and 'KG02') and one daylength sensitive and late flowering ('Rogue'). Each entry was represented by five plants in a plot. Some entries were included as additional replications within blocks. Plots were transplanted on July 6, 2020, at the UC Davis Plant Sciences Farm in Davis CA. Seedlings were transplanted into the middle of raised soil bed, with beds spaced 1.5m on center. Each plot was 4.5 m long within a bed with five plants per plot spaced 90 cm apart, with no additional end-to-end space between plots. The trial plots were bordered by two rows of hemp plants on either side of the plot area and maize (*Zea mays* L.) was planted as border surrounding the entire perimeter of the field.

The May 2021 trial was designed as a randomized complete block design with four replicates with 64 plots per replication. The trial contained 88 entries because some entries were present in fewer than four replications. Entries included 14 cloned genotypes which were propagated from plants selected from the 2020 field, 33 families previously grown in the 2020 trial, 38 new families, and the same three check cultivars. Seedlings were transplanted on May 12, 2021 at the UC Davis Plant Sciences farm. The trial was planted similarly to the June 2020 trial. Two rows of hemp plants were transplanted to act as borders and sorghum (*Sorghum bicolor* L. Moench) surrounded the perimeter of the field.

The July 2021 trial was designed as a randomized complete block design with four replications each with 63 plots. This trial contained 69 entries, with the same three control entries, 34 previously grown in the May 2021 field, and 32 new families. Seedlings were transplanted on July 7, 2021 and each plot contained six plants grown in two rows 90 cm apart on 1.5m beds. Plants were spaced 120cm apart within rows and offset 60cm between rows. No additional space was added between plots on a bed.

Crop Management

The 2020 trial site was cover cropped with wheat in the fall. Prior to transplanting, field crews added a total of 168 kg/ha using a mixture of Starter 8-24-6, UAN-32, and CAN-17. Irrigation was applied for 6 hours daily Monday through Friday providing approximately 4 cm of water a week to the field. Coragen and Macho insecticides were applied through the drip irrigation. Coragen was readministered a month into the trial due to continued presence of corn ear worm (*Helicoverpa zea*). Before planting in May 2021, 28lbs N/acre was applied to the field. An additional 10lbs of N/acre using the same starter fertilizer, Nu-Crop (8-24-6), was applied 2 weeks after transplanting. UAN-32 fertilizer was applied at 28 kg/ha along with Coragen insecticide through the irrigation on June 30 and July 30, 2021. For the July 2021 field, fertilizer was applied at transplant at a rate of 200 lbs N/acre using a mixture of Starter 8-24-6, UAN-32, and CAN-17. Two weeks after transplanting, 10 lbs of N/acre using the same starter fertilizer (8-24-6) was applied 2 weeks after transplanting; the irrigation scheduling was the same as the previous year. For all fields, seedlings were watered using a surface-drip system for a month before being switched to a sub-surface drip system to minimize weeds.

Phenotyping

Plants were scored for maturity twice a week, on Mondays and Thursdays, on a 0 to 5 scale, with 0 = vegetative and 5 = mature cola ready for harvest (Table 1; Figure 1).

Table 1. The descriptive definitions of the developmental stages of *Cannabis sativa* L. throughout its lifespan from seedling to harvest ready.

Score	Developmental Stage Description
0	• From the primordia on the nodes only vegetative growth emerges. New tissue emerging from the apical shoot tip meristem is distinctly different from the flowering tissue.
1	• The meristem at the shoot tip may begin to differentiate from the vegetative tissue. The emergence of few immature pistillate flowers is the first sign of pre-flowering • Some plants stay in this stage for long period of time depending on their underlying genetics
2	• Multiple calyxes appear on the apical meristem, along with a few on the nodes. Fresh white, light yellow, or pink pistils, extend from these pistillates.
3	• The cola begins to develop as more calyx emerge on the nodes. The internode length continues to decrease. Calyxes appear on most nodes on the apical meristem and on branches. • Mono and trifoliate leaves join with each set of calyxes, these are called sugar leaves
4	• The cola starts to fill up and the internodes become harder to see. The secretion of resin is high. The bracts and sugar leaves on the cola are sticky to the touch. • Pistils still maintain their original color.
5	• Half of all pistils begin to turn to a brown color. • Resin production has peaked at this point; plant is ready to harvest

Figure 1. The scoring scale used to phenotype individuals for maturity in our field experiments. Descriptions of each stage are seen in Table 1.

Data were collected using the Android application, Field Book (Rife and Poland, 2014). For each day of data collection, the accumulated growing degree days (AGDD) and the day length were recorded. AGDD was calculated using the formula:

$$AGDD = \sum_{i=0}^{n} \left(\frac{T_{max} - T_{min}}{2} \right) - T_b$$

where T_{max}, T_{min}, and T_b refer to the maximum temperature (°C), the minimum temperature (°C), and the base temperature, respectively (Akyüz and Ransom, 2015). AGDD data were obtained from the website GreenCast (Syngenta; https://www.greencastonline.com/growing-degree-days/home), using a base temperature of 0°C. We recorded the AGDD for each individual plant based on the date on which they scored "3" on the maturity scale. In addition, individuals within families were categorized by their accumulated GDD at flowering as either early flower (EF) or late flower (LF) relative to the family and the trial they were in. Chi-square tests were conducted for all families that were present in at least two of the three trials using the EF/LF classification of individual genotypes to assess if the proportion of plants in each category within families was similar across trials.

Genotyping of GWAS Panel

Leaf samples were collected from 5000 individuals from all trials one month after the transplant date of that field. We removed a single leaflet from the top three inches of each plant and placed it in a 96 well plate, which was subsequently sealed and frozen in a -20ºC freezer for at least one week before DNA was extracted. The extraction protocol follows that of Clark et al. (2022). Following DNA extraction, quality and concentration were checked and assessed using a nanodrop spectrophotometer. Reduced representation libraries were created for genotyping-by-sequencing (GBS) following the protocol of Elshire et al. (2011), with the exception that a *Hind*III restriction enzyme was used. 1500 individuals from families which were segregating for day neutrality were used in our GBS. Libraries were size selected and sequenced on an ABI Prism® 3730 Genetic Analyzer at the UC Davis Sequencing Facility.

Statistical analysis

The phenotypic data of all individuals, whether we genotyped them or not, were analyzed using linear mixed models (LMMs) with the R package, *lme4* (Bates et al., 2015):

$$Y \sim N\ (\mathbf{X}\beta + \boldsymbol{o},\ \sigma^2 \boldsymbol{W}^{-1}),$$

where Y was the observed phenotype for a given individual, N was the dimension of the response vector, W was a diagonal matrix of known prior weights, β was a p-dimensional coefficient vector accounting for random effects of Replication, Environment, Family, and the Family × Environment interaction (G × E), X was an n × p model matrix, and o was a vector of known prior offset terms. We also ran a separate model without family or family × environment interaction effects in order to adjust individual plant data for spatial and environmental variation.

Heritability

Variance components were computed for each trait across all environments and within each environment to calculate broad sense heritability (Table 5). We computed heritability as follows,

$$H^2 = \frac{\boldsymbol{\sigma_g^2}}{\boldsymbol{\sigma_g^2} + \frac{\boldsymbol{\sigma_r^2} + \boldsymbol{\sigma_e^2}}{\mathbf{r}}}$$

Where σ^2_g is the family variance, σ^2_e is the residual variance, σ^2_r is the replicate variance, and r is the number of replicates in the field.

GWAS

We used the statistical software TASSEL 5.0 (Bradbury et al., 2007) for our GWAS. Raw sequence data were processed using the TASSEL GBS pipeline. We used the *C. sativa* variety 'Purple Kush' reference genome (Bakel et al., 2011) to align our GBS sequences and identify SNP marker loci. We ran a k-nearest neighbor SNP imputation method (LD-kNNi) to fill in missing SNP information among our taxa (Money et al., 2015). After imputation, we removed SNP loci that had a read depth of less than 2, a minor allele frequency across all genotypes of less than 0.02, and a major allele frequency greater than 0.98 (Petit, et al., 2020).

Our GWAS was run using the mixed linear model

$$Y = \mu + \mathrm{X}_i\beta_i + Population\ structure + Kinship + \epsilon$$

where Y is the phenotypic observation on an individual plant as adjusted for spatial and environmental effects, μ is the population mean, $\mathrm{X}_i\beta_i$ is the fixed effect of the SNP, population structure as described by principle components from a principal components analysis (PCA)) of the marker data, kinship was assessed using a centered identity-by-state (IBS) method, and ϵ is the error. We also evaluated a second model without the kinship matrix. The GWAS was run using all individuals across all environments in one analysis with environment effects as random in our model.

A Bonferroni-correction was applied to determine the significance threshold for p-values assigned to each SNP. The Bonferroni correction is considered one for the most stringent ways to set a threshold for significant SNPs (Kuo, 2017). The threshold is set by the -log10 of the quotient of the significance threshold (0.05) divided by the total number of SNPs used in the

GWAS (50,907), which sets a significance threshold of 6.01. For all statistical analyses, significance was assessed at the 5% probability level unless noted otherwise.

Results

Phenotypic Trait Statistics and Model Effects

Across all environments, family, environment, replicate, and family × environment effects were present for AGDD and Daylength at Flower (Table 2). Because of significant family × environment effects were present; we analyzed each environment independently (Table 3).

Table 2: ANOVA results using all data across three fields with Family, Field, and Replicate terms as main effects and the Family:Field term as an interaction effect for AGDD at Flower and Daylength at Flower.

	Main Effects			**Interactions**
Trait	Family	Environment	Replicate	Family x Environment
AGDD at Flower	***	***	***	***
Daylength at Flower	***	***	***	***

* P < 0.05; ** = P < 0.01; *** = P< 0.001; n.s. = non-significant

Family effects were present for all trials for both traits, but block effects were only present in 2020, possibly because it was a larger trial. Segregating families grown in two or three environments were analyzed using Chi-square tests to determine if the proportions of EF/LF individual plants within the family changed across environments. No family tested showed a difference except KNR2C (Supplementary Table 1).

Table 3: ANOVA results using each field: July 2020, May 2021, and July 2021 as independent datasets. 'Family', and 'Replicate' terms were random effects in each fields individual model.

	July 2020		
Trait	Family	Replicate	Family x Replicate
AGDD at Flower	***	***	n.s.
Daylength at Flower	***	**	n.s.
	May 2021		
AGDD at Flower	***	n.s.	n.s.
Daylength at Flower	***	n.s.	n.s.
	July 2021		
AGDD at Flower	***	n.s.	***
Daylength at Flower	***	n.s.	***

* $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$; n.s. = non-significant; n.a. = not applicable

With the AGDD and Daylength at Flower traits, we ran a Pearson's Product-Moment Correlation to see if there was any association between the measurement of the two traits. Across each of environments, the two traits were strongly statistically significant inversely correlated, with the range of the correlation being -0.92 to -0.99 (Table 4). Further evidence of the inverse relation between the two traits can be seen in the histogram and boxplot distributions for the two traits (Supplementary Figures 1-4).

Table 4. Pearson correlation between our two traits, AGDD and Daylength at Flower within each individual environment.

Environments	Pearson's Product-Moment Correlation (95% CI)
2020	-0.99***
Early 2021	-0.92***
Late 2021	-0.99***

* P < 0.05; ** = P < 0.01; *** = P< 0.001; n.s. = non-significant

Heritability

Broad-sense heritability within fields for AGDD at Flower ranged from 0.60-0.73 and for Daylength at Flower, from 0.60-0.82; across all fields, AGDD and Daylength at Flower were 0.55 and 0.44 respectively (Table 5).

Table 5. Variance components for each trait and location combination and their heritability's

	Variance Components			
Trait:Field	**σ^2_g**	**σ^2_e**	**σ^2_r**	**H^2**
AGDD:July 2020	239.06	221.42	80.627	0.61
AGDD:May 2021	434.67	388.561	18.751	0.81
AGDD:July 2021	208.581	189.841	81.987	0.75
DL:July 2020	0.3	0.29	0.1	0.60
DL:May 2021	0.365	0.31	0.01	0.82
DL:July 2021	0.324	0.276	0.12	0.77
AGDD:All Fields	293.8	274.9	192.6	0.55
DL:All Fields	0.33	0.28	0.54	0.44

σ^2_g: Family; σ^2_e: Residual; σ^2_r : Replicate; H^2: Broad-sense heritability

Genetic markers

We obtained a total of 129,823 sequencing reads for the 1507 plants used across all three fields. We then filtered the remaining reads to include those with an average read depth per individual of 2 and that were present in 80% of individuals. This left a set of 50,907 markers used for this analysis.

Population structure

Within our three environments, when running a principle component analysis, we observed our individuals clustering into three groups (Figure 2; Figure 3). The way that our groupings could be described is that the day neutral individuals fall into one cluster, and day sensitive crosses falling into the two other clusters. Individuals which were mixed in between had genetics comprised that like those of the three clustered groupings.

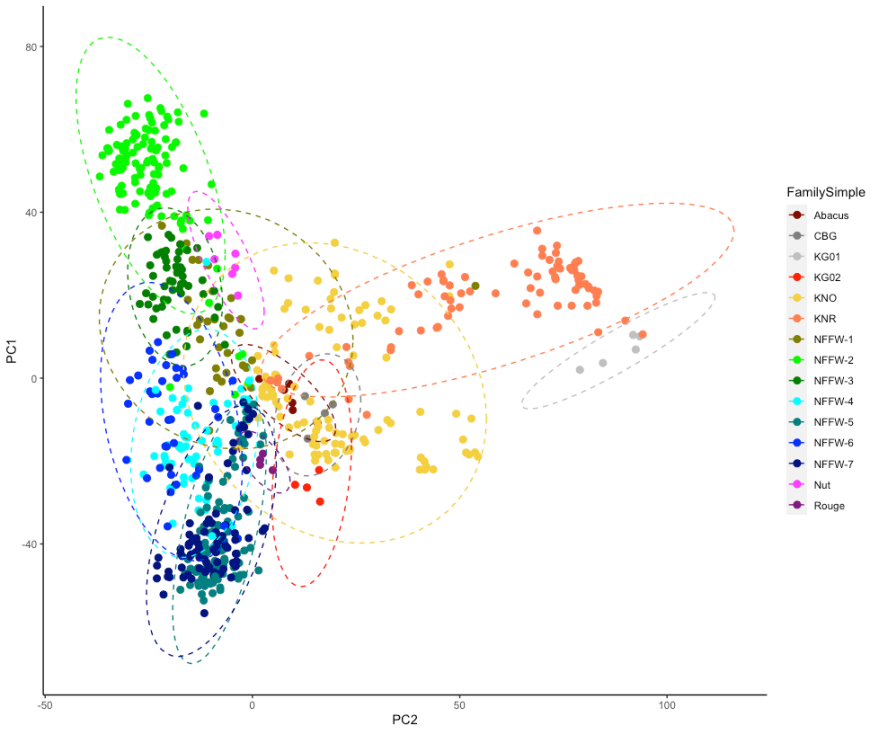


Figure 2. PCA plot for families included in the 2020 GWAS trial.

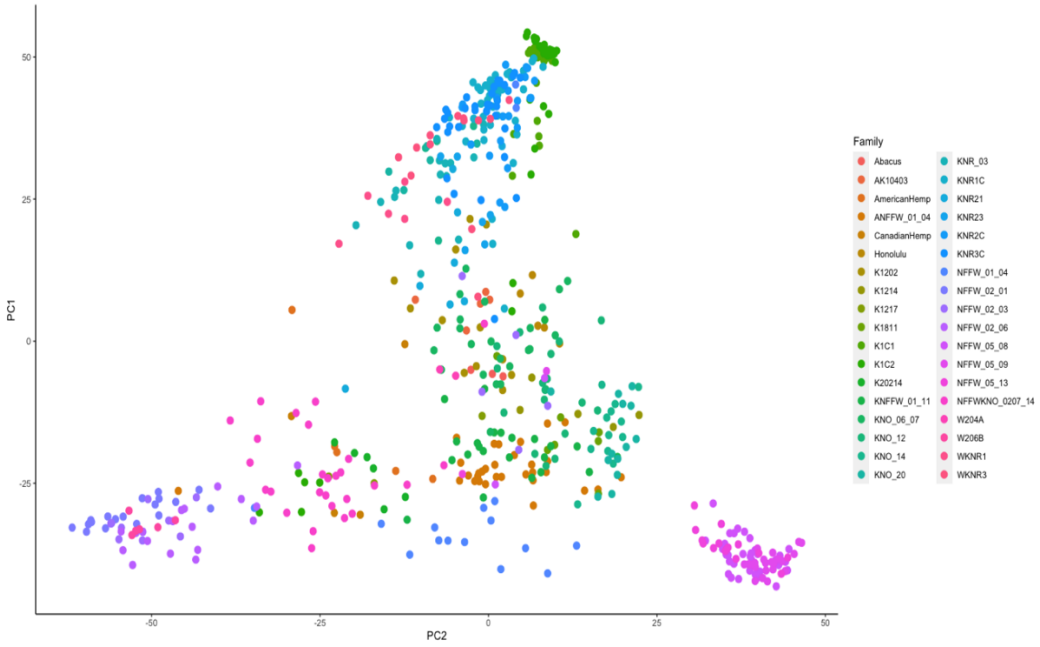


Figure 3. PCA plot for families included in both trials during the 2021 GWAS field trials.

Genome Wide Association Study

A genome wide association study was first done using a general linear model in TASSEL 5.0 for both AGDD and Day Length at Flower. For both traits, the QQ-plot (Figure 4), showed a large inflation of p-values, suggesting that many of our SNP markers could be considered as false-positive associations with the traits. Given that our GWAS population contained well-defined families of known relatedness, we expected that including kinship in the model would be necessary. Therefore, the GWAS was re-run using a mixed linear model (MLM), which incorporated the population structure using a Q matrix and integrated the relatedness of individuals with a kinship matrix. This deflated the magnitude of p-values and reduced the presence of false-positive results (Liu et al., 2016; Fig. 5).

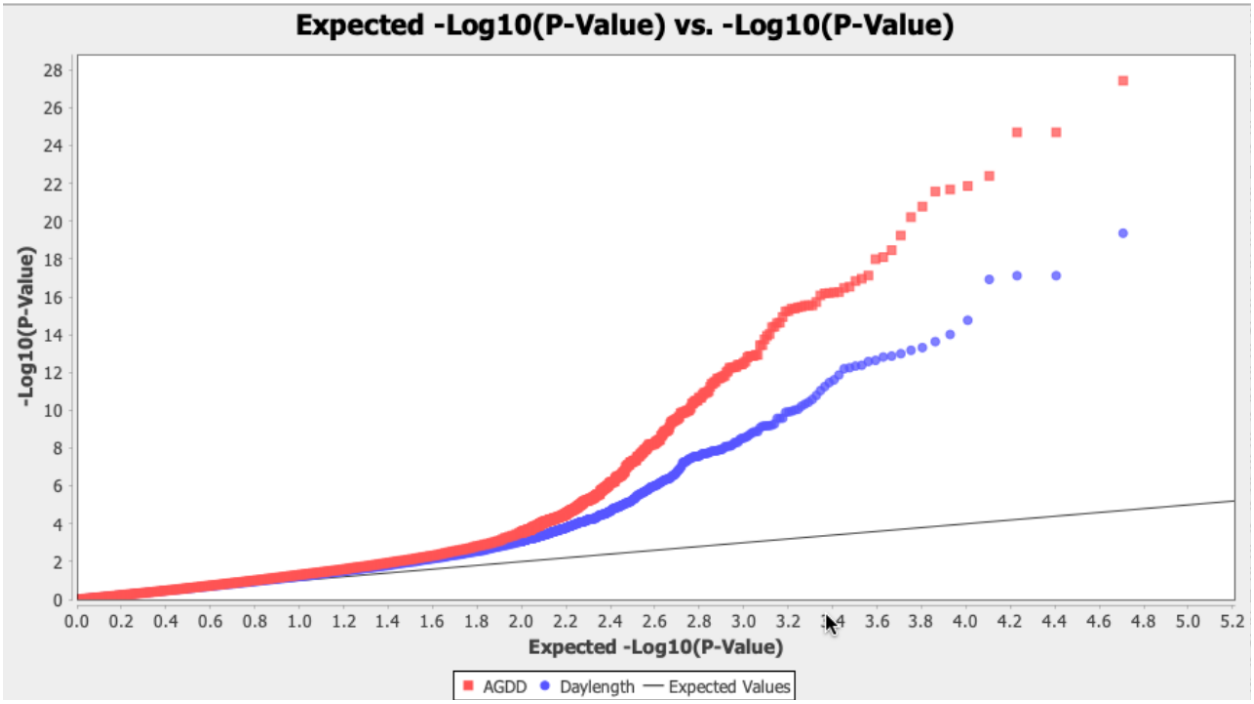


Figure 4. QQ plot of -$\log_{10}$(p-values) for accumulated growth degree days to flower and daylength at flower using a GLM method. The solid black line is the expected distribution of SNPs when comparing expected p-values versus actual p-values.

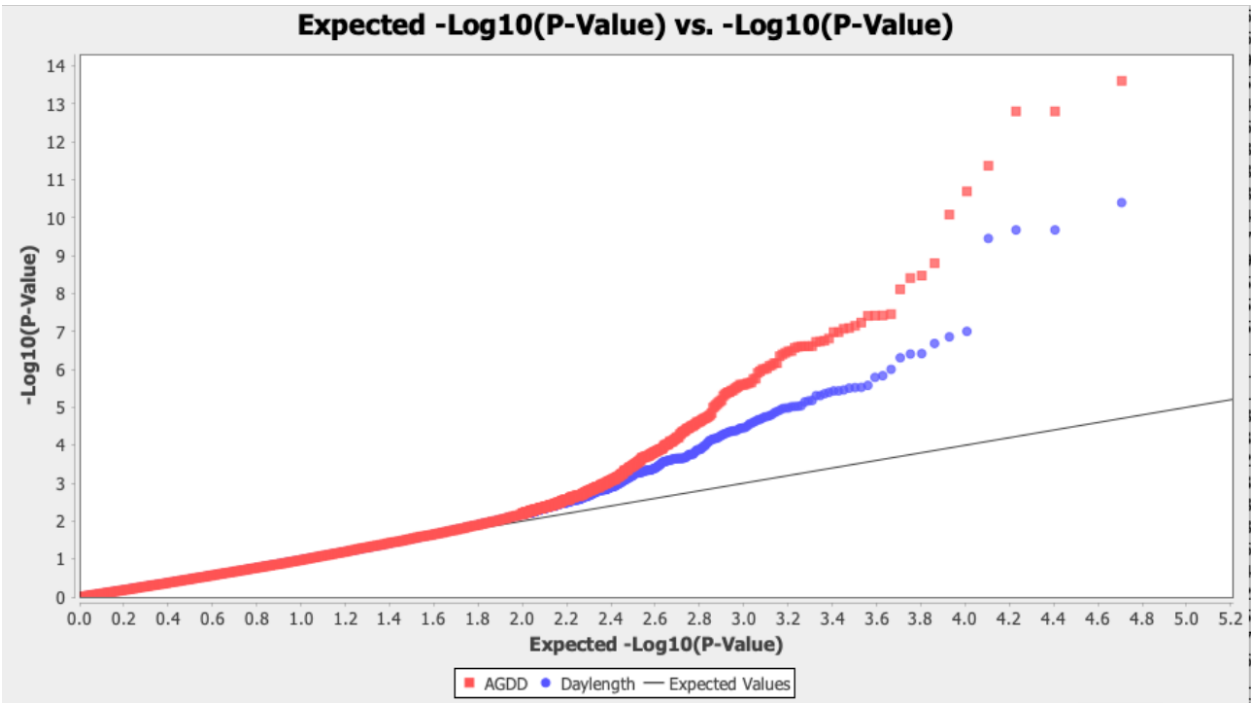


Figure 5. QQ plot of -$\log_{10}$(p-values) for accumulated growth degree days to flower and daylength at flower using a MLM method. The solid black line is the expected theoretical distribution of values; points which deviate further away from the expected line are significantly different from the expected distribution

The mixed linear Q+K model showed that that 54 SNP markers for both traits were present above the significance threshold set by a Bonferroni correction at 6.007 (Figure 6,7 and Supplementary Tables 2,3). These markers were mostly on Chromosome 2, although one marker on Chromosome 8 was also noted for AGDD.

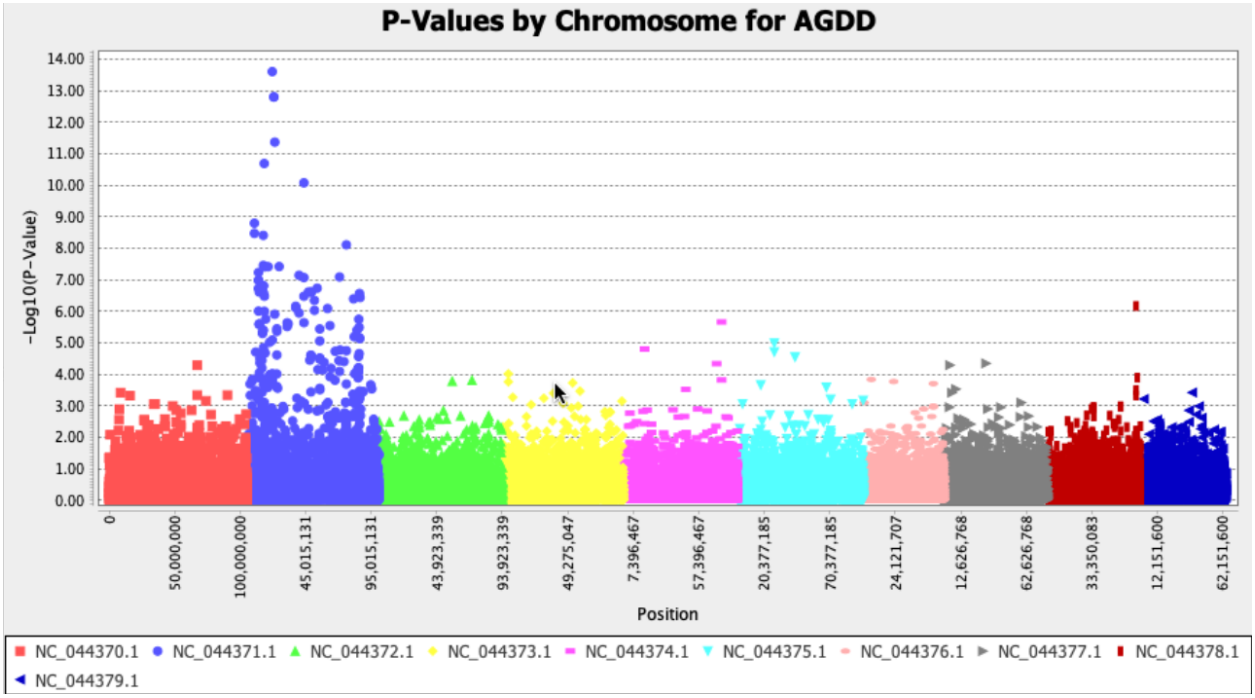


Figure 6. Manhattan plot for accumulated growth degrees to flower across all fields using a MLM method.

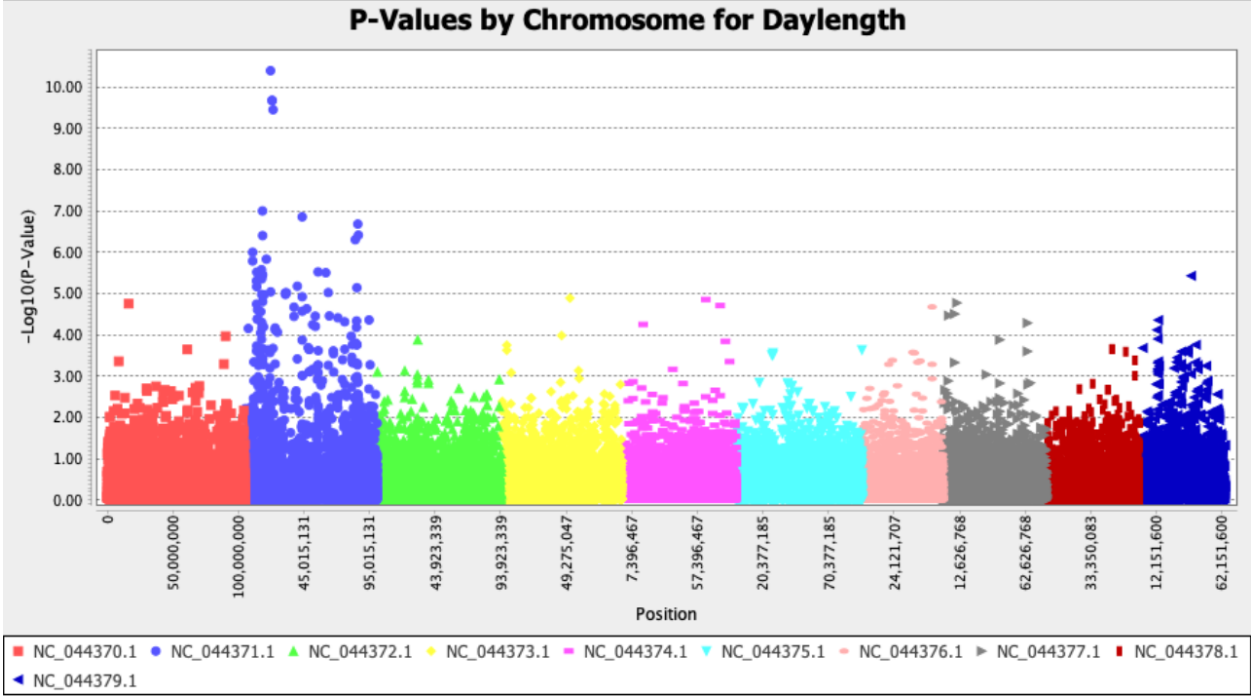


Figure 7. Manhattan plot for day length at flower across all fields using a MLM method.

Candidate Genes for Day Neutrality

Fifty-four combined SNPs between AGDD and Daylength at Flower were marked as hits by having a p-value above the significance threshold set by a Bonferroni correction, 6.007: AGDD and Daylength at flower, both of which had the same top 5 SNP hits, were located on Chromosome 2 in the hemp genome (Supplementary Table 2,3). The GWAS shows SNPs associated with both traits in several regions on Chromosome 2, possibly because of the manner in which our families were constructed and the large linkage disequilibrium that they contained.

Discussion

Phenotypic Trait Statistics and Model Effects

We expected families to differ for flowering time because they included both day neutral and daylength sensitive phenotypes. We also observed a significant family × environment interaction, suggesting that flowering differed across trials. This is likely to be related to the earlier planted May 2021 trial compared to the two July trials, given the differences in photoperiod. The two traits measured in our experiment had a strong and significant inverse correlation as expected. Due to hemp plants typically being short day plants which need to reach a critical photoperiod to initiate flowering, plants which need very short days would accumulate more AGDD; due to our planting dates in May to June, days only begin to shorten past the summer solstice, around June 21st, and only then, and sometimes even a few weeks after that, do hemp plants initiate flowering.

Segregation of Genotypes Across Fields

After categorizing individuals in families within two groups, early flowering (EF) and late flowering (LF), we saw no significant differences ($P< 0.05$) in the proportion of segregants in each family for all except for one, KNR2C. Regarding KNR2C, the family had a larger proportion of earlier flowering individuals in May 2021 than July 2021, where the proportion of early to late flowering individuals was reversed. There was significant G×E interactions when observing the dataset, and in some capacity, there may be a way in which the individuals within KNR2C are responding to changes in the environment; specifically, because of differences which arise in temperature and daylength between fields planted weeks apart. The May 2021 trial had an uneven distribution of nitrogen, with plots at the southern end of our field, showing

N deficiency and possibly inducing a larger proportion of KNR2C individuals to flower earlier compared to the same family planted in the July trial. The KNR2C family had few individuals and was not included in all replications, and hence the difference between fields could have been due to sampling. In addition, with any field trial, the individuals transplanted into the field could have been mislabeled or placed incorrectly.

Heritability of Phenotypic Traits

With both traits, across all fields, broad sense heritability ranged between 0.60-0.82 (Table 5). This falls just out of line with similar conclusions from Petit et al., 2020 which found broad sense heritability of flowering time in hemp to be 0.94 - 0.95 for beginning and full flowering, respectively. Broad sense heritability of a specific trait "is the proportion of phenotypic variation attributable to genetics" (Wang et al., 2013). The heritability of both traits was markedly lower in the July 2020 trial in which we had two replications vs. four for the 2021 trials. The moderate heritability of both traits in 2020 indicates that while variation for flowering time has a large genetic component, it can still be influenced by the environment, particularly if the experimental design is not robust. Across all three environments, heritability of AGDD was 0.55 and of Daylength at Flower, 0.44.

Genome Wide Association Study

Our parental germplasm was quite diverse, including different sources of high cannabinoid material as well as grain types with desirable architecture. Therefore, we controlled for both population structure and kinship in our association study. We identified numerous SNP located on Chromosome 2 for both traits. The QQ-plot for our model, while better than a general linear model that excluded population structure and kinship, still suggested that some of these

associations could be false positives, as seen for SNPs trending above the line between the expected and actual statistical significance (Figure 4).

Using mixed linear models to find associations between traits and SNPs have been used extensively, not just in plants in general, but when looking at flowering traits specifically. Papers identifying SNPs for flowering timing have been found for hemp (Petit et al., 2021), alfalfa (Adhikari et al., 2019), and cowpea (Paudel et al., 2021).

Candidate Genes for Day Neutrality

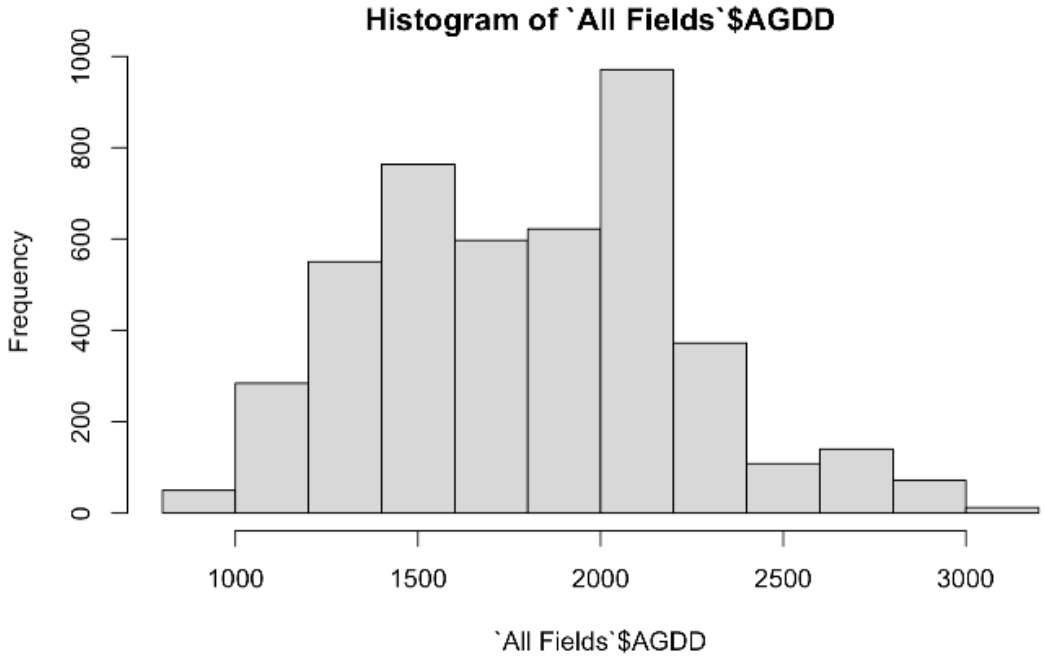
When a GWAS was run with all trials, we observed that the five most significant SNPs were shared between the two phenotypic traits (Supplementary Table 2,3). We looked at the *C. sativa* L. variety 'Purple Kush' on the GenBank website (Benson 2013) and tried to identify if there were any candidate genes which might have been within, or just downstream of our significant hits. We investigated a 31 Mb region of Chromosome 2 between our two most and distal SNPs in order to see if we could identify genes which were linked to day-neutrality much like CONSTANS (CO). What we have found was one gene, flowering locus K homology domain (FLK), which was located on position 28,634,491 of chromosome 2, 7Mb away from two significant hits shared between both traits. In Arabidopsis, the flowering locus K is a regulator of other flowering time genes. Arabidopsis plants with mutations to FLK were shown to effect relative expression levels of CONSTANS in the individual plants (Mockler, 2004). With this, it might be possible that FLK works in the same way protein DAY NEUTRAL FLOWERING (DNF) acts in Arabidopsis. DNF works in the same pathway as CO and works to maintain low levels of CO in short day settings (Morris, 2010). DNF when mutated, shows to initiate earlier flowering in individuals overexpressing the protein, and also in instances in mutant *dnf* individuals.

Other significant SNPs which weren't as near to the gene might still be connected through linkage disequilibrium (LD). However, it is just as likely that due to low sequencing depth, there is an artificial inflation of certain SNPs which may also have no association to any particular trait and are purely artifactual. To see if some of these peaks are true hits, perhaps rerunning the sequencing data for more stricter filtering can serve to increase the resolution of our most significant hits.

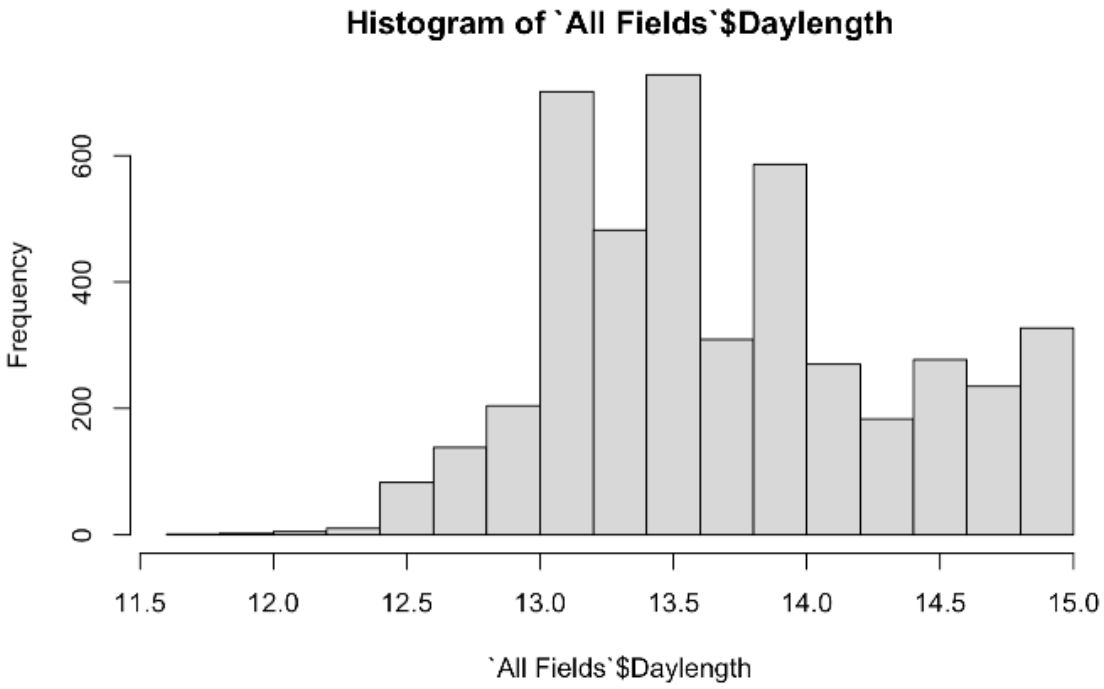
Conclusion

From our GWAS analysis of 1507 individual plants across two years and three different environments, we looked at over 50,000 SNPs and identified 54 total significant SNPs, of which the top five most significant were shared between our two phenotypic traits, AGDD and Daylength at Flower. We searched the SNP sequences against the reference genome *C. sativa* L. 'Purple Kush' in GenBank and found a proteins that were within a 21Mb region of our shared significant SNP hits: flowering locus K homology domain (FLK), a known down regulator of CONSTANS. It is also possible that associated SNPs which were not close to the gene locus might still be associated due to linkage disequilibrium. To see if this is the case, deeper sequencing is needed to resolve our significant hits and remove any other SNPs falsely associated with day neutrality.

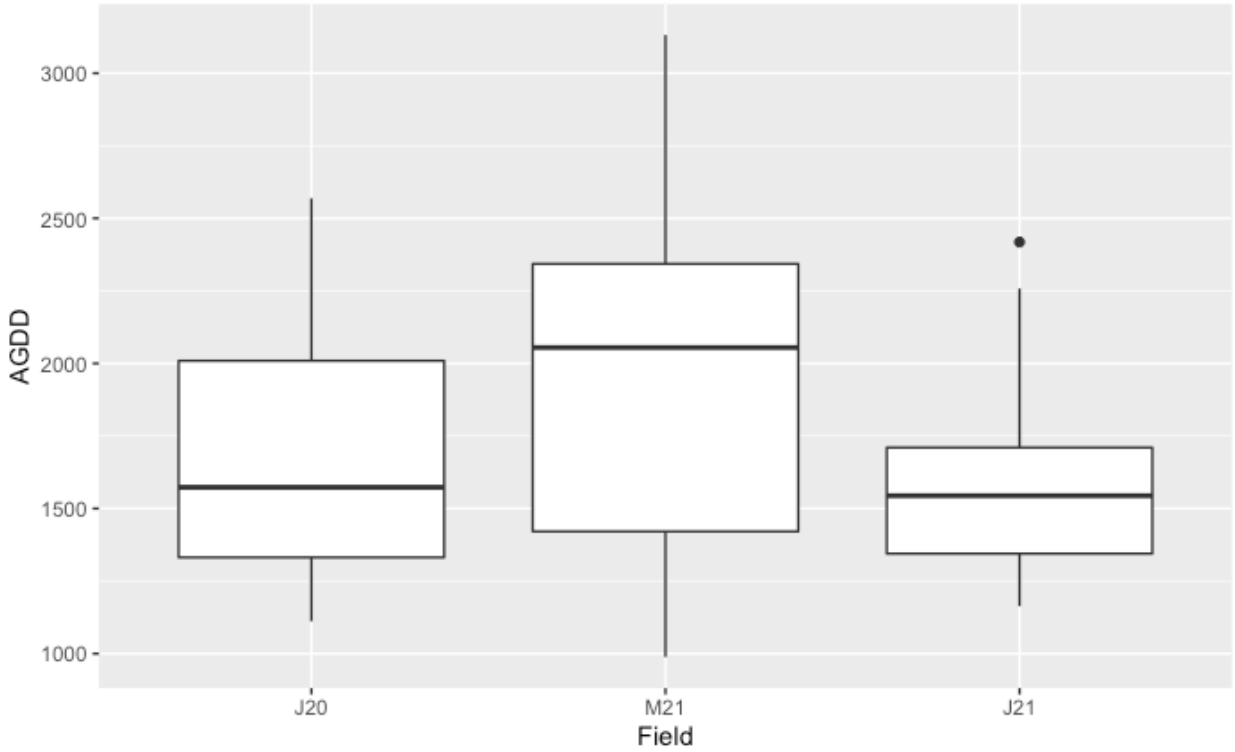
Appendix



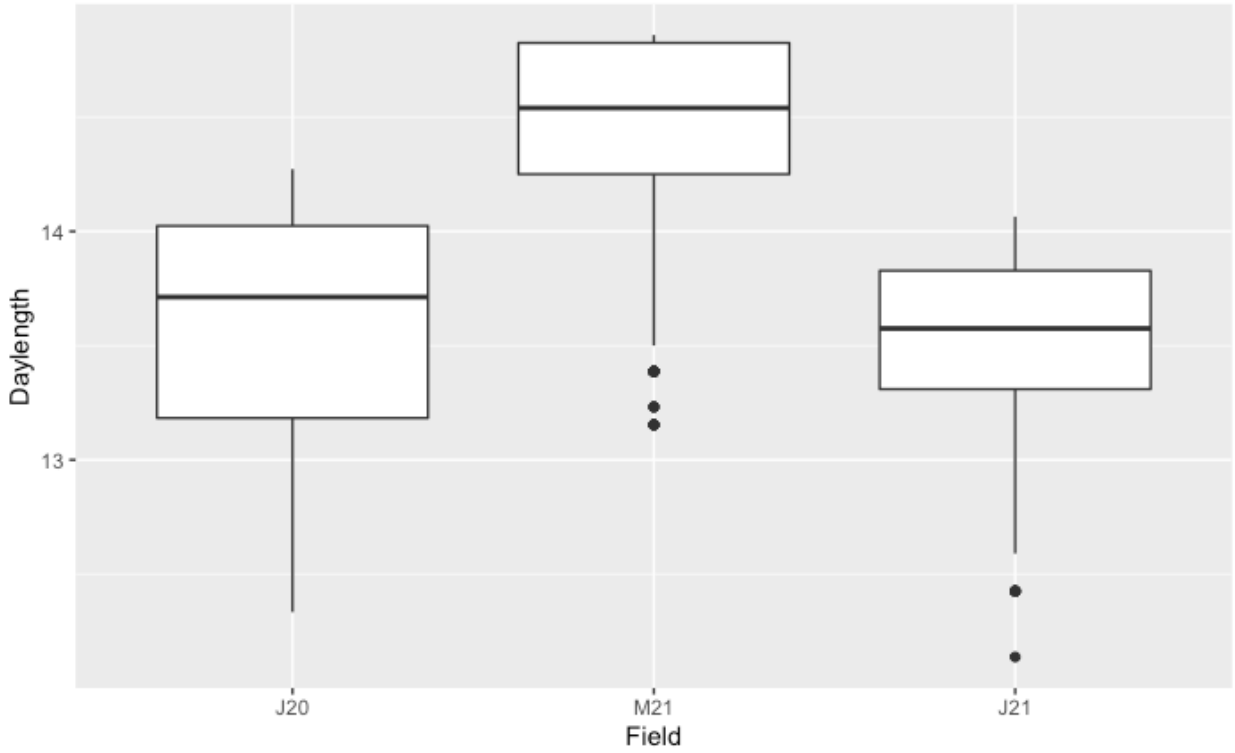
Supplementary Figure 1: Distribution of AGDD at Flower for all plants across all fields



Supplementary Figure 2: Distribution of Daylength at Flower for all plants across all fields



Supplementary Figure 3: The accumulated growth degree days at flower for July 2020 field (J20), May 2021 field (M21), and the July 2021 field (J21). The white region of the boxplot, known as the interquartile range (IQR), highlights the 25th (Q1) and 75th (Q3) percentile of the data, while the solid black line is the median of the data. The lines drawn outside of the IQR are the (Q1 – 1.5 * IQR) and (Q3 +1.5 * IQR) values of the data. Circles outside of the boxplot are outlier values



Supplementary Figure 4: The day length at flower for July 2020 field (J20), July 2021 field (J21), and May 2021 field (M21). The white region of the boxplot, known as the interquartile range (IQR), highlights the 25th (Q1) and 75th (Q3) percentile of the data, while the solid black line is the median of the data. The lines drawn outside of the IQR are the (Q1 – 1.5 * IQR) and (Q3 +1.5 * IQR) values of the data. Circles outside of the boxplot are outlier values

Supplementary Table 1. Chi-square scores and p-values for flowering phenotypes which families were present in at least two of the three experimental trials. Families were divided into two groups, early flower (EF) and later flower (LF).

Family	Chi Square Score	P-Value
ANFFW-01-04	1.14	0.28
Canadian	0.196	0.65
Chinese	0.1628	0.687
K1C1	2.59	0.107
K1C2	3.0952	0.0785
K2C	0.1939	0.659
KG01	2.5264	0.2827
KG02	3.0047	0.222
KKNO-07	0.9023	0.34
KNFFW-01-11	0.4532	0.5
KNO-06-07	0.0007	0.979
KNO-07	2.6729	0.102
KNO-07-SP	1.7943	0.18
KNO-12	3.1586	0.07553
KNO-14	5.1605	0.0758
KNO-15	3.9408	0.1394
KNO-15-07	0.0974	0.75494
KNR-01-01-SP	0.1309	0.7175
KNR-02	0.6332	0.7286
KNR-03	0.1757	0.675082
KNR-04	0.5397	0.462
KNR1C	0.8187	0.3656
KNR2C	5.1089	**0.02384***
KNR3C	1.0004	0.3172
KNR81C	0.7917	0.373
KNR83C	0.0075	0.9311
NFFW-1-01	0.1206	0.7284
NFFW-1-04	0.4158	0.519
NFFW-2-01	0.65	0.42
NFFW-2-03	0.3635	0.546
NFFW-2-03-SP	0.009	0.924
NFFW-2-04	4.0457	0.1323
NFFW-2-05	0.0297	0.863
NFFW-2-06	4.9608	0.08371
NFFW-3-02	4.603	0.100109
NFFW-3-04	1.2228	0.2688
NFFW-3-05	0.0767	0.782
NFFW-4-01	0.9637	0.617
NFFW-4-04	0.2205	0.64
NFFW-4-09	2.1407	0.3429
NFFW-5-02	3.915	0.1411
NFFW-5-04	0.001	0.974
NFFW-5-08	3.546	0.1698
NFFW-5-09	0.0638	0.968
NFFW-5-13	0.9423	0.331
NFFW-7-01	1.2416	0.5375
NFFW-7-03	0.3048	0.581
NFFW-7-07	0.808	0.3867
NFFWKNO-207-14	0.1902	0.6627
NFFWKNO-505-14	0.588	0.443
Nut SP	0.0794	0.778
Rogue	0.7962	0.671
W209C	1.329	0.249

* $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$

Supplementary Table 2. Top SNP hits for AGDD to flower, calculated from the MLM, with their associated location in the hemp chromosome, minor allele frequency (MAF), and their effect size.

Marker	Chromosome	Position	F-Score	P-Value	Additive Effect
SNC_044371.1_19585509	2	19585509	32.13409	2.48E-14	163.38397
SNC_044371.1_20770861	2	20770861	30.18504	1.58E-13	-1.45E+02
SNC_044371.1_20770895	2	20770895	30.18504	1.58E-13	-1.45E+02
SNC_044371.1_21521079	2	21521079	26.74198	4.27E-12	140.83962
SNC_044371.1_13444061	2	13444061	25.11266	2.04E-11	64.44883
SNC_044371.1_43735521	2	43735521	23.64122	8.35E-11	55.08206
SNC_044371.1_5943154	2	5943154	20.58576	1.60E-09	149.84704
SNC_044371.1_5943179	2	5943179	19.81641	3.36E-09	-1.47E+02
SNC_044371.1_12734992	2	12734992	19.67701	3.88E-09	-1.11E+02
SNC_044371.1_76126443	2	76126443	18.96636	7.74E-09	-1.12E+02

Supplementary Table 3. Top SNP hits for day length at flower, calculated from the MLM, with their associated location in the hemp chromosome, minor allele frequency (MAF), and their effect size.

Marker	Chromosome	Position	F-Score	P-Value	Additive Effect
SNC_044371.1_19585509	2	19585509	24.40511	4.02E-11	-1.48E-01
SNC_044371.1_20770861	2	20770861	22.68173	2.11E-10	0.13916
SNC_044371.1_20770895	2	20770895	22.68173	2.11E-10	0.13916
SNC_044371.1_21521079	2	21521079	22.14955	3.54E-10	-1.37E-01
SNC_044371.1_13444061	2	13444061	16.334	9.96E-08	-5.64E-02
SNC_044371.1_43735521	2	43735521	15.98973	1.39E-07	-4.37E-02
SNC_044371.1_86154459	2	86154459	15.57956	2.07E-07	0.07901
SNC_044371.1_86595207	2	86595207	14.94694	3.84E-07	0.12836
SNC_044371.1_13523517	2	13523517	14.92582	3.94E-07	0.07761
SNC_044371.1_84200442	2	84200442	14.69129	4.96E-07	0.05059

Supplementary Table 4. Crosses made and used in the 2020/2021 Hemp field trials along with the number of individuals submitted in our GBS panel for DNA sequencing.

Family	Cross	Type	2020 GBS	May 2021 GBS	July 2021 GBS
NFFW-1-03	NFF-01.07 X NFF-01.23		8		
NFFW-1-04	NFF-01.08 X NFF-01.23		18	13	
NFFW-1-10	NFF-01.29 X NFF-01.23		7		
NFFW-2-01	NFF-02.02 X NFF-02.03		9	17	
NFFW-2-02	NFF-02.04 X NFF-02.03		9		
NFFW-2-03	NFF-02.05 X NFF-02.03		18	15	
NFFW-2-04	NFF-02.06 X NFF-02.03		15	16	14
NFFW-2-06	NFF-02.08 X NFF-02.03		40	12	18
NFFW-2-07	NFF-02.21 X NFF-02.03		10		
NFFW-2-08	NFF-02.22 X NFF-02.03		7		
NFFW-2-09	NFF-02.23 X NFF-02.03		9		
NFFW-2-11	NFF-02.25 X NFF-02.03		10		
NFFW-3-02	NFF-03.02 X NFF-03.28		17	20	8
NFFW-3-04	NFF-03.05 X NFF-03.28		10	19	
NFFW-3-05	NFF-03.06 X NFF-03.28		17	15	
NFFW-3-08	NFF-03.10 X NFF-03.28		16		
NFFW-4-01	NFF-04.03 X NFF-04.02		15	12	12
NFFW-4-04	NFF-04.07 X NFF-04.02		17	18	
NFFW-4-09	NFF-04.26 X NFF-04.02		14	14	14
NFFW-4-10	NFF-04.27 X NFF-04.02		18	17	
NFFW-5-02	NFF-05.02 X NFF-05.26		18	17	19
NFFW-5-04	NFF-05.04 X NFF-05.26		15	16	
NFFW-5-08	NFF-05.09 X NFF-05.26		20	21	2
NFFW-5-09	NFF-05.21 X NFF-05.26		40	20	13
NFFW-5-13	NFF-05.51 X NFF-05.26		20	15	
NFFW-5-14	NFF-05.52 X NFF-05.26		10		
NFFW-6-03	NFF-06.02 X NFF-06.04		10		
NFFW-6-15	NFF-06.24 X NFF-06.04		10		
NFFW-6-17	NFF-06.25 X NFF-06.04		10		
NFFW-6-35	NFF-06.×× X NFF-06.04		10		
NFFW-7-01	NFF-07.01 X NFF-07.07		19	20	7
NFFW-7-03	NFF-07.03 X NFF-07.07		18	19	
NFFW-7-04	NFF-07.06 X NFF-07.07		10		
NFFW-7-07	NFF-07.11 X NFF-07.07		20	18	
NFFW-7-15	NFF-07.31 X NFF-07.07		17		
KNO-05	KG-9201.05 X NFF-05.26		9		

KNO-07	KG-9201.44 X NFF-03.28		20	19	
KNO-08	KG-9201.45 X NFF-07.12		9		
KNO-09	KG-9202.01 X NFF-03.28		9		
KNO-12	NF-9201.01 X NFF-04.29		10	17	
KNO-13	NF-9201.01 X NFF-03.28		8		
KNO-14	NF-9201.01 X NFF-05.26		18	14	
KNO-17	NF-9201.28 X NFF-05.26		10		
KNO-20	NF-9201.46 X NFF-05.26		20	20	
KNR-01	KG-9201.01 X NF-9201.26		10		
KNR-02	KG-9201.61 X NF-9201.26		40		
KNR-03	KG-9201.62 X NF-9201.26		20	13	
KNR-08	KG-9202.61 X NF-9201.04		9		
NFF-01-23	NF.17 X C307.005	Parent	1		
NFF-02-03	NF.17 X CW.023	Parent	1		
NFF-03-28	NF.21 X CW.023	Parent	1		
NFF-04-02	NF.33 X BB.025	Parent	1		
NFF-04-29	NF.33 X BB.025	Parent	1		
NFF-05-26	NF.33 X C307.005	Parent	1		
NFF-06-04	NF.33 X CW.023	Parent	1		
NFF-06-21	NF.33 X CW.023	Parent	1		
NFF-07-07	NF.33 X CW.025	Parent	1		
NFF-07-12	NF.33 X CW.025	Parent	1		
NF9201-04	KG-9201-bulk X (NF.8 or NF.20)	Parent	1		
NF9201-26	KG-9201-bulk X (NF.8 or NF.20)	Parent	1		
NUT	Original seed X	Check	8		
ROGUE		Check	5		
CBG#1		Check	5		
KG01		Check	5		
KG02		Check	5		
KNR-02	KG-9201.61 X NF-9201.26			11	22
KKNO-07	KNO-07_GH01 X KG02_GH33			19	21
KNFFW-01-11	NFFW-01-11_GH01 X KG02_GH33			11	11
KNO-06-07	KNO-06_GH03 X KNO-07_GH02			12	9
KNO-W-07	KNO-07_GH03 X KNO-07_GH02			17	7
NFFWKNO-207-14	NFFW-2-07_GH02 X KNO-14_GH04			18	9
K1202	KNO-12_F20-1 X KNR-02_#8				13
W110A-TS	NFFW-1-10_L34-4 X NFFW-1-10_L34-4				10
W204A	NFFW-2-04_#2 X NFFW-2-04_#4				13
W206B	NFFW-2-06_#3 X NFFW-2-06_#1				15

WKNR1	KNR-01_#1 X KNR-01_#7				5
WKNR3	KNR-03_#5 X KNR-03_#6				11
KNR23	KNR-02_#5 X KNR-03_#6				12
K20214	NFFW-2-02_I17-1 X KNO-14_#2				19
K1217	KNO-12_F20-1 X KNO-17_K03-4				7
K1214	KNO-12_F20-1 X KNO-14_#2				7
K1811	KNO-18_B41-2 X KNFFW-1-11_#2				13
Abacus		Parent			5
Honolulu Haze		Future Parent			5
CBG#1		Parent			5

References

Abernethy, Amy. "Hemp Production and the 2018 Farm Bill - 07/25/2019." *FDA*, 25 July 2019, https://www.fda.gov/news-events/congressional-testimony/hemp-production-and-2018-farm-bill-07252019.

Adesina, Ifeoluwa, et al. "A Review on the Current State of Knowledge of Growing Conditions, Agronomic Soil Health Practices and Utilities of Hemp in the United States." *Agriculture*, vol. 10, no. 4, 4, Apr. 2020, p. 129. *www.mdpi.com*, https://doi.org/10.3390/agriculture10040129.

Adhikari, Laxman, et al. "QTL Mapping of Flowering Time and Biomass Yield in Tetraploid Alfalfa (Medicago Sativa L.)." *BMC Plant Biology*, vol. 19, no. 1, Aug. 2019, p. 359. *BioMed Central*, https://doi.org/10.1186/s12870-019-1946-0.

Andre, Christelle M., et al. "Cannabis Sativa: The Plant of the Thousand and One Molecules." *Frontiers in Plant Science*, vol. 7, Feb. 2016, p. 19. *PubMed Central*, https://doi.org/10.3389/fpls.2016.00019.

Barcaccia, Gianni, et al. "Potentials and Challenges of Genomics for Breeding Cannabis Cultivars." *Frontiers in Plant Science*, vol. 11, 2020. *Frontiers*, https://www.frontiersin.org/article/10.3389/fpls.2020.573299.

Bates, Douglas, et al. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software*, vol. 67, Oct. 2015, pp. 1–48. *www.jstatsoft.org*, https://doi.org/10.18637/jss.v067.i01.

Benson, Dennis A., et al. "GenBank." *Nucleic Acids Research*, vol. 41, no. Database issue, Jan. 2013, pp. D36-42. *PubMed*, https://doi.org/10.1093/nar/gks1195.

Bill Text - AB-223 Compassionate Use Act of 1996. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201120120AB223. Accessed 25 Jan. 2022.

Blümel, Martina, et al. "Flowering Time Regulation in Crops—What Did We Learn from Arabidopsis?" *Current Opinion in Biotechnology*, vol. 32, Apr. 2015, pp. 121–29. *PubMed*, https://doi.org/10.1016/j.copbio.2014.11.023.

Bradbury, Peter J., et al. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics*, vol. 23, no. 19, Oct. 2007, pp. 2633–35. *Silverchair*, https://doi.org/10.1093/bioinformatics/btm308.

Campiglia, Enio, et al. "Plant Density and Nitrogen Fertilization Affect Agronomic Performance of Industrial Hemp (Cannabis Sativa L.) in Mediterranean Environment." *Industrial Crops and Products*, vol. 100, June 2017, pp. 246–54. *ScienceDirect*, https://doi.org/10.1016/j.indcrop.2017.02.022.

Cannabis | A Complete Guide | Ernest Small | Taylor & Francis Group. https://www.taylorfrancis.com/books/mono/10.1201/9781315367583/cannabis-ernest-small. Accessed 15 Dec. 2021.

Castro, Patricia, et al. "Genetic Mapping of Day-Neutrality in Cultivated Strawberry." *Molecular Breeding*, vol. 35, Feb. 2015. *ResearchGate*, https://doi.org/10.1007/s11032-015-0250-4.

Cherney, Jerome H., and Ernest Small. "Industrial Hemp in North America: Production, Politics and Potential." *Agronomy*, vol. 6, no. 4, 4, Dec. 2016, p. 58. *www.mdpi.com*, https://doi.org/10.3390/agronomy6040058.

Clarke, Robert Connell. *Marijuana Botany: An Advanced Study: The Propagation and Breeding of Distinctive Cannabis*. Ronin Publishing, 1981.

Clarke, Robert, and Mark Merlin. *Cannabis: Evolution and Ethnobotany*. 2016.

Collard, Bertrand C. Y., and David J. Mackill. "Marker-Assisted Selection: An Approach for Precision Plant Breeding in the Twenty-First Century." *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, Feb. 2008, pp. 557–72. *PubMed Central*, https://doi.org/10.1098/rstb.2007.2170.

Dahlhamer, James. "Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults — United States, 2016." *MMWR. Morbidity and Mortality Weekly Report*, vol. 67, 2018. *www.cdc.gov*, https://doi.org/10.15585/mmwr.mm6736a2.

Darghan, Aquiles E., et al. "Effect of Experimental Blocking on the Suppression of Spatial Dependence Potentially Attributable to Physicochemical Properties of Soils." *Modelling and Simulation in Engineering*, vol. 2021, Sept. 2021, p. e3322074. *www.hindawi.com*, https://doi.org/10.1155/2021/3322074.

de Meijer, E. P. M., and L. J. M. van Soest. "The CPRO Cannabis Germplasm Collection." *Euphytica*, vol. 62, no. 3, Jan. 1992, pp. 201–11. *Springer Link*, https://doi.org/10.1007/BF00041754.

Dufresnes, Christophe, et al. "Broad-Scale Genetic Diversity of Cannabis for Forensic Applications." *PLOS ONE*, vol. 12, no. 1, Jan. 2017, p. e0170522. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0170522.

Ehret, Georg B. "Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension." *Current Hypertension Reports*, vol. 12, no. 1, Feb. 2010, pp. 17–25. *PubMed Central*, https://doi.org/10.1007/s11906-009-0086-6.

Eichhorn Bilodeau, Samuel, et al. "An Update on Plant Photobiology and Implications for Cannabis Production." *Frontiers in Plant Science*, vol. 10, 2019, p. 296. *Frontiers*, https://doi.org/10.3389/fpls.2019.00296.

Formato, Marialuisa, et al. "(–)-Cannabidiolic Acid, a Still Overlooked Bioactive Compound: An Introductory Review and Preliminary Research." *Molecules*, vol. 25, no. 11, June 2020, p. 2638. *PubMed Central*, https://doi.org/10.3390/molecules25112638.

Gabay, Michael. "The Federal Controlled Substances Act: Schedules and Pharmacy Registration." *Hospital Pharmacy*, vol. 48, no. 6, June 2013, pp. 473–74. *PubMed Central*, https://doi.org/10.1310/hpj4806-473.

Garrett, E. R., and C. A. Hunt. "Physiochemical Properties, Solubility, and Protein Binding of Delta9-Tetrahydrocannabinol." *Journal of Pharmaceutical Sciences*, vol. 63, no. 7, July 1974, pp. 1056–64. *PubMed*, https://doi.org/10.1002/jps.2600630705.

Gasparyan, Lusine, et al. "DNA Sequencing: Current State and Prospects of Development." *Open Journal of Biophysics*, vol. 9, no. 3, 3, May 2019, pp. 169–97. *www.scirp.org*, https://doi.org/10.4236/ojbiphy.2019.93013.

"Hemp Fiber Market Size | Share | Trends | Growth | Scope | Forecast." *Verified Market Research*, https://www.verifiedmarketresearch.com/product/hemp-fiber-market/. Accessed 9 Apr. 2022.

Hopkins, James F. *A History of the Hemp Industry in Kentucky*. University Press of Kentucky, 2014.

Janes, E. W., et al. *Cardinal Temperatures and Thermal Time for Seed Germination of Industrial Hemp (Cannabis Sativa)©a*. p. 6.

Jin, Dan, et al. "Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes." *Scientific Reports*, vol. 10, no. 1, 1, Feb. 2020, p. 3309. *www.nature.com*, https://doi.org/10.1038/s41598-020-60172-6.

Johnson, Renée. *Hemp as an Agricultural Commodity*. p. 48.

Jung, Christian, and Andreas E. Müller. "Flowering Time Control and Applications in Plant Breeding." *Trends in Plant Science*, vol. 14, no. 10, Oct. 2009, pp. 563–73. *ScienceDirect*, https://doi.org/10.1016/j.tplants.2009.07.005.

Lisson, S. N., et al. "Development of a Hemp (Cannabis Sativa L.) Simulation Model 2.The Flowering Response of Two Hemp Cultivars to Photoperiod." *Australian Journal of Experimental Agriculture*, vol. 40, no. 3, 2000, p. 413. *DOI.org (Crossref)*, https://doi.org/10.1071/EA99059.

Liu, Xiaolei, et al. "Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies." *PLOS Genetics*, vol. 12, no. 2, Feb. 2016, p. e1005767. *PLoS Journals*, https://doi.org/10.1371/journal.pgen.1005767.

Lubell, Jessica D., and Mark H. Brand. "Foliar Sprays of Silver Thiosulfate Produce Male Flowers on Female Hemp Plants." *HortTechnology*, vol. 28, no. 6, Dec. 2018, pp. 743–47. *journals.ashs.org*, https://doi.org/10.21273/HORTTECH04188-18.

Mauseth, James D. "Structure-Function Relationships in Highly Modified Shoots of Cactaceae." *Annals of Botany*, vol. 98, no. 5, Nov. 2006, pp. 901–26. *PubMed*, https://doi.org/10.1093/aob/mcl133.

McPartland, John M. "Cannabis Systematics at the Levels of Family, Genus, and Species." *Cannabis and Cannabinoid Research*, vol. 3, no. 1, Dec. 2018, pp. 203–12. *liebertpub.com (Atypon)*, https://doi.org/10.1089/can.2018.0039.

McPartland, John M., and Geoffrey W. Guy. "Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names." *The Botanical Review*, vol. 83, no. 4, Dec. 2017, pp. 327–81. *Springer Link*, https://doi.org/10.1007/s12229-017-9187-0.

Mizoguchi, Tsuyoshi, et al. "Day-Neutral Response of Photoperiodic Flowering in Tomatoes: Possible Implications Based on Recent Molecular Genetics of Arabidopsis and Rice." *Plant Biotechnology*, vol. 24, Dec. 2007. *ResearchGate*, https://doi.org/10.5511/plantbiotechnology.24.83.

Mockler, Todd C., et al. "Regulation of Flowering Time in Arabidopsis by K Homology Domain Proteins." *Proceedings of the National Academy of Sciences*, vol. 101, no. 34, Aug. 2004, pp. 12759–64. *pnas.org (Atypon)*, https://doi.org/10.1073/pnas.0404552101.

Money, Daniel, et al. "LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms." *G3: Genes, Genomes, Genetics*, vol. 5, no. 11, Nov. 2015, pp. 2383–90. *www.g3journal.org*, https://doi.org/10.1534/g3.115.021667.

Morris, Karl, et al. "DAY NEUTRAL FLOWERING Represses CONSTANS to Prevent Arabidopsis Flowering Early in Short Days." *The Plant Cell*, vol. 22, no. 4, Apr. 2010, pp. 1118–28. *Silverchair*, https://doi.org/10.1105/tpc.109.066605.

National Academies of Sciences, Engineering, et al. "Therapeutic Effects of Cannabis and Cannabinoids." *The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research*, National Academies Press (US), 2017. *www.ncbi.nlm.nih.gov*, https://www.ncbi.nlm.nih.gov/books/NBK425767/.

Petit, Jordi, et al. "Genetic Architecture of Flowering Time and Sex Determination in Hemp (Cannabis Sativa L.): A Genome-Wide Association Study." *Frontiers in Plant Science*, vol. 11, 2020, p. 1704. *Frontiers*, https://doi.org/10.3389/fpls.2020.569958.

Piomelli, Daniele, and Ethan B. Russo. "The Cannabis Sativa Versus Cannabis Indica Debate: An Interview with Ethan Russo, MD." *Cannabis and Cannabinoid Research*, vol. 1, no. 1, Jan. 2016, pp. 44–46. *PubMed Central*, https://doi.org/10.1089/can.2015.29003.ebr.

Polson, D. E. "Day-Neutrality in Soybeans1." *Crop Science*, vol. 12, no. 6, 1972, p. cropsci1972.0011183X001200060017x. *Wiley Online Library*, https://doi.org/10.2135/cropsci1972.0011183X001200060017x.

Pook, Torsten, et al. "Increasing Calling Accuracy, Coverage, and Read-Depth in Sequence Data by the Use of Haplotype Blocks." *PLOS Genetics*, vol. 17, no. 12, Dec. 2021, p. e1009944. *PLoS Journals*, https://doi.org/10.1371/journal.pgen.1009944.

Punja, Zamir K., et al. "Assessing Genetic Diversity in Cannabis Sativa Using Molecular Approaches." *Cannabis Sativa L. - Botany and Biotechnology*, edited by Suman Chandra et al., Springer International Publishing, 2017, pp. 395–418. *Springer Link*, https://doi.org/10.1007/978-3-319-54564-6_19.

Ranalli, Paolo. "Current Status and Future Scenarios of Hemp Breeding." *Euphytica*, vol. 140, no. 1, Jan. 2004, pp. 121–31. *Springer Link*, https://doi.org/10.1007/s10681-004-4760-0.

Ren, Guangpeng, et al. "Large-Scale Whole-Genome Resequencing Unravels the Domestication History of Cannabis Sativa." *Science Advances*, vol. 7, no. 29, p. eabg2286. *science.org (Atypon)*, https://doi.org/10.1126/sciadv.abg2286.

Rife, Trevor W., and Jesse A. Poland. "Field Book: An Open-Source Application for Field Data Collection on Android." *Crop Science*, vol. 54, no. 4, 2014, pp. 1624–27. *Wiley Online Library*, https://doi.org/10.2135/cropsci2013.08.0579.

Rowley, Daniel, et al. "Late-Season Strawberry Production Using Day-Neutral Cultivars in High-Elevation High Tunnels." *HortScience*, vol. 46, no. 11, Nov. 2011, pp. 1480–85. *journals.ashs.org*, https://doi.org/10.21273/HORTSCI.46.11.1480.

Rupasinghe, H. P. Vasantha, et al. "Industrial Hemp (Cannabis Sativa Subsp. Sativa) as an Emerging Source for Value-Added Functional Food Ingredients and Nutraceuticals." *Molecules*, vol. 25, no. 18, Sept. 2020, p. 4078. *PubMed Central*, https://doi.org/10.3390/molecules25184078.

Russo, Ethan B., et al. "Cannabis, Pain, and Sleep: Lessons from Therapeutic Clinical Trials of Sativex®, a Cannabis-Based Medicine." *Chemistry & Biodiversity*, vol. 4, no. 8, 2007, pp. 1729–43. *Wiley Online Library*, https://doi.org/10.1002/cbdv.200790150.

Salentijn, Elma M. J., Qingying Zhang, et al. “New Developments in Fiber Hemp (Cannabis Sativa L.) Breeding.” *Industrial Crops and Products*, vol. 68, June 2015, pp. 32–41. *ScienceDirect*, https://doi.org/10.1016/j.indcrop.2014.08.011.

Salentijn, Elma M. J., Jordi Petit, et al. “The Complex Interactions Between Flowering Behavior and Fiber Quality in Hemp.” *Frontiers in Plant Science*, vol. 10, 2019. *Frontiers*, https://www.frontiersin.org/article/10.3389/fpls.2019.00614.

Small, Ernest, and Arthur Cronquist. “A Practical and Natural Taxonomy for Cannabis.” *TAXON*, vol. 25, no. 4, 1976, pp. 405–35. *Wiley Online Library*, https://doi.org/10.2307/1220524.

Spitzer-Rimon, Ben, et al. “Architecture and Florogenesis in Female Cannabis Sativa Plants.” *Frontiers in Plant Science*, vol. 10, 2019. *Frontiers*, https://www.frontiersin.org/article/10.3389/fpls.2019.00350.

Struik, P. C., et al. “Agronomy of Fibre Hemp (Cannabis Sativa L.) in Europe.” *Industrial Crops and Products*, vol. 11, no. 2, Mar. 2000, pp. 107–18. *ScienceDirect*, https://doi.org/10.1016/S0926-6690(99)00048-5.

Takeno, Kiyotoshi. “Stress-Induced Flowering: The Third Category of Flowering Response.” *Journal of Experimental Botany*, vol. 67, no. 17, Sept. 2016, pp. 4925–34. *Silverchair*, https://doi.org/10.1093/jxb/erw272.

Toth, Jacob A., et al. “Limited Effect of Environmental Stress on Cannabinoid Profiles in High-Cannabidiol Hemp (Cannabis Sativa L.).” *GCB Bioenergy*, vol. 13, no. 10, 2021, pp. 1666–74. *Wiley Online Library*, https://doi.org/10.1111/gcbb.12880.

Turner, Adrian S., et al. “The Effect of Day-Neutral Mutations in Barley and Wheat on the Interaction between Photoperiod and Vernalization.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, vol. 126, no. 9, 2013, pp. 2267–77. *PubMed Central*, https://doi.org/10.1007/s00122-013-2133-6.

van Bakel, Harm, et al. “The Draft Genome and Transcriptome of Cannabis Sativa.” *Genome Biology*, vol. 12, no. 10, Oct. 2011, p. R102. *PubMed*, https://doi.org/10.1186/gb-2011-12-10-r102.

Van der Werf, H. M. G., and W. Van den Berg. “Nitrogen Fertilization and Sex Expression Affect Size Variability of Fibre Hemp (Cannabis Sativa L.).” *Oecologia*, vol. 103, no. 4, 1995, pp. 462–70.

Wang, Yunpeng, et al. “Effect of Regulatory Architecture on Broad versus Narrow Sense Heritability.” *PLOS Computational Biology*, vol. 9, no. 5, May 2013, p. e1003053. *PLoS Journals*, https://doi.org/10.1371/journal.pcbi.1003053.

Xu, Yunbi, and Jonathan H. Crouch. "Marker-Assisted Selection in Plant Breeding: From Publications to Practice." *Crop Science*, vol. 48, no. 2, 2008, pp. 391–407. *Wiley Online Library*, https://doi.org/10.2135/cropsci2007.04.0191.

Zargar, Sajad, et al. "Recent Advances in Molecular Marker Techniques: Insight into QTL Mapping, GWAS and Genomic Selection in Plants." *Journal of Crop Science and Biotechnology*, vol. 18, Dec. 2015, pp. 293–308. *ResearchGate*, https://doi.org/10.1007/s12892-015-0037-5.