

DISSERTATION

PHENOTYPE TO GENOTYPE AND BACK IN EMERGING AND ESTABLISHED CROP
SPECIES

Submitted by

Patrick O'Neal David Woods

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2023

Doctoral Committee:

Advisor: John McKay

Ruth Hufbauer

Chris Funk

Dan Sloan

Copyright by Patrick O'Neal David Woods 2023

All Rights Reserved

ABSTRACT

PHENOTYPE TO GENOTYPE AND BACK IN EMERGING AND ESTABLISHED CROP SPECIES

Understanding the relationship between the phenotype and genotype is a fundamental goal of genetics. Through the years, two primary approaches have been developed for studying the phenotype-genotype relationship: forward genetic and reverse genetics. Forward genetics enables the potential discovery of numerous candidate genes controlling a phenotype while reverse genetics allows for the mechanistic validation of a single gene's role in controlling a phenotype. Applying these two approaches to crops enables the discovery of genetic targets that can be used for crop improvement through breeding. In this dissertation, I focused on understanding the phenotype-genotype relationship in both the emerging crop *Cannabis sativa* and the established crop *Maize*. In Chapter 1, I used both a forward a reverse genetics approach to identify and validate candidate genes controlling agriculturally important traits (agronomic and biochemical) in *Cannabis sativa*. In Chapter 2, I used a reverse population genetics approach to identify the genetics underlying local adaptation in feral and domesticated populations of *Cannabis sativa*. In Chapter 3, I used a forward genetics approach to identify candidate genes controlling variation in root system architecture in *Maize*. Collectively, this work demonstrates how modern genomic techniques can be applied to both new and old crop systems to identify genetic targets for use in crop innovation through breeding.

TABLE OF CONTENTS

ABSTRACT.....	ii
CHAPTER 1.....	1
INTRODUCTION.....	1
SUMMARY OF SUBSEQUENT DISSERTATION CHAPTERS.....	5
REFERENCES.....	7
CHAPTER 2.....	8
QUANTITATIVE TRAIT LOCI CONTROLLING AGRONOMIC AND BIOCHEMICAL TRAITS IN CANNABIS SATIVA.....	8
SUMMARY.....	8
INTRODUCTION.....	9
MATERIALS AND METHODS.....	12
RESULTS.....	21
DISCUSSION.....	33
DATA AVAILABILITY.....	39
REFERENCES.....	40
CHAPTER 3.....	47
GENOME WIDE POLYMORPHISM AND GENIC SELECTION IN FERAL AND DOMESTICATED LINEAGES OF CANNABIS SATIVA.....	47
SUMMARY.....	47
INTRODUCTION.....	48
MATERIALS AND METHODS.....	50
RESULTS.....	56
DISCUSSION.....	64
CONCLUSION.....	70
DATA AVAILABILITY.....	71
REFERENCES.....	72
CHAPTER 4.....	78
ROOT PULLING FORCE ACROSS DROUGHT IN <i>MAIZE</i> REVEALS GENOTYPE BY ENVIRONMENT INTERACTIONS AND CANDIDATE GENES.....	78
SUMMARY.....	78
INTRODUCTION.....	79
MATERIALS AND METHODS.....	83
RESULTS.....	87
DISCUSSION.....	99
CONCLUSION.....	104
DATA AVAILABILITY.....	104
REFERENCES.....	105
APPENDIX I.....	110
APPENEIX II.....	131
APPENDIX III.....	135

CHAPTER 1

INTRODUCTION

Background

The genotype to phenotype relationship and its cyclical nature of study

Understanding the genetics that influence the phenotype (P, some observable trait) of an individual is a fundamental goal of genetics. It is well understood however that the genotype (the genetic constitution) of an individual is not the only parameter that influences the phenotype. In addition to the genotype (G), it has been well documented that the environment (E) and the interaction between the genotype and environment (G x E) significantly influence the phenotype (P) as well. As such, the basic phenotypic equation is usually written as: $P = G + E + (G \times E)$ (Lynch & Walsh 1998). Despite these additional parameters, geneticists place their attention on understanding the genotype parameter of the phenotype because this is the only parameter that can be explicitly controlled, allowing for specific manipulation of the phenotype (i.e., through selective breeding).

The genotype parameter of the phenotype is typically studied in two complimentary approaches which ultimately form a cyclical relationship of study. The first of these two approaches is often referred to as “forward genetics”. A forward genetics approach study begins with a known phenotype measurement (i.e., plant height) and seeks to identify the yet unknown genes associated with measurable variation in the phenotype (Gurumurthy et al. 2016). Forward genetic studies are typically conducted using a mapping population (bi-parental or diversity panel) that has been specifically designed to encompass high variation for the phenotype of interest. Once the phenotype of interest has been measured, associations between both the mapping population’s

phenotypic and genotypic variation are conducted. Genes found to possess highly significant associations to the phenotype are then denoted as “candidate genes” that are assumed to comprise the genotype parameter for that phenotype.

The second of the two approaches for studying the genotype parameter of the phenotype is often referred to as “reverse genetics”. A reverse genetics approach study begins with either a known gene or genotype information (i.e., alleles at a gene or across the genome) and seeks to either mechanistically understand a gene’s effect on a particular phenotype or predict how a particular combination of alleles may affect a particular phenotype (Gurumurthy et al. 2016). If a particular gene has been identified (and its sequence available) with no known effect on the phenotype of interest, the reverse genetics approach study typically proceeds by incorporating genetic perturbations (targeted mutations) into that gene’s sequence to mechanistically understand how altering the gene’s sequence affects the phenotype. If genotype information is available and no specific gene has been identified (i.e., genome wide polymorphism data), the reverse genetics approach study may proceed by identifying significant associations between allele frequencies and distinct genetic groups (i.e., populations) of individuals to infer the fitness of a particular allele.

The purpose of studying the genotype to phenotype relationship in crops

Plants in general are a very effective system for studying the genotype to phenotype relationship because of the possibility for extensive experimental design, replication, and control of environments (Speed & Balding 2012). Within plants, one of the main applications of studying the genotype to phenotype relationship is in crop species because of the potential to identify breeding targets to improve industrial varieties through introgression of advantageous genetic variation (alleles, gene copy number, etc.) (Kumar et al. 2017). In addition to identifying breeding targets for crop improvement, studying the genotype to phenotype relationship in crops can

elucidate how genetic factors such as dominance, additivity, epistasis, and pleiotropy affect agriculturally important phenotypes which may impact how introgression proceeds (Lynch & Walsh 1998).

Studying the genotype to phenotype relationship in both emerging and established crops

Depending on the crop species being studied, the viability of using forward versus reverse genetic approaches differs immensely due to the great variation in the volume and quality of genetic resources available across crop species. In this dissertation, I studied the genotype to phenotype relationship in both “emerging” and “established” crop species. I define “emerging” as a crop species that is relatively new to the formal agriculture market and consequently has a narrow volume of genetic literature and possesses limited genetic resources. The emerging crop species I studied the genotype to phenotype relationship of in this dissertation is *Cannabis sativa*, a species with a clandestine history that was prohibited in much of the world until recently (Kovalchuk et al. 2020). As time has progressed, regulations around the world surrounding the legality of *Cannabis sativa* have increasingly relaxed, allowing this crop to be grown commercially for human consumption (U.S. Govt., 2014 and U.S. Govt., 2018). This species is highly versatile for human use and possess numerous agriculturally important phenotypes that have applications in the medicinal, textile and food industries. With the increasing relaxation of growing *Cannabis sativa* commercially, scientists now have the exciting opportunity to investigate fundamental questions regarding this species’ evolution, local adaptation, and genetic architecture of agriculturally important phenotypes. In recent years, scientists have produced studies which provide insight on each of the aforementioned subject areas. While informative, each study on the genotype to phenotype relationship in *Cannabis sativa* has had to conform to the limitations of studying a non-model (emerging) species which include draft reference assemblies, limited publicly available

whole genome sequence data, no availability of mutants with gene specific perturbations, and poorly understood molecular mechanisms influencing agriculturally important traits (Kovalchuk et al. 2020). Because of these limitations, additional investigations regarding the fundamental aforementioned areas remain necessary. It is important however that when designing a study to understand the genotype to phenotype relationship in an emerging crop species, one must consider approaches to cope with the limitations to ensure a viable study.

In contrast to emerging crops, I define “established” crops as crop species that are staples to the formal agriculture market that also have a dense volume of literature and higher-quality genetic resources. The established crop species I studied the genotype to phenotype relationship of in this dissertation is *Zea mays*, known commonly as maize or corn. Maize is one of the major cereal crops commercially grown throughout the world, making up more than 95% of the grain feed acreage in the United States alone. Given its prominence in the agriculture industry, much priority has been given to understanding the genetics of evolution, local adaptation, and genetic architecture controlling agriculturally important phenotypes of maize. This has afforded maize with a wealth of genetic literature and extensive genetic resources such as high-quality reference genomes, numerous available gene specific mutants, and a library of publicly available whole genome sequence data (Lawrence et al. 2005). When studying the genotype to phenotype relationship in an established crop species such as maize, a main challenge can be identifying an aspect of the species which has not received as much scientific attention. Typically, new studies on maize genetics can focus on either describing the genetic architecture of historically ignored phenotypes or mechanistically understanding how a previously described candidate affects a target phenotype. An example of a less well studied aspect of maize that I investigated in this dissertation is the phenotypic plasticity and genetic control of maize root system architecture.

Summary of subsequent dissertation chapters

Chapter 2: Quantitative Trait Loci Controlling Agronomic and Biochemical Traits in Cannabis sativa

In this chapter, I used both forward and reverse genetic approaches to elucidate the quantitative genetic architecture controlling agriculturally important phenotypes in *Cannabis sativa*. Specifically, I used an F₂ mapping population bred from the industrial hemp parents Carmagnola and USO31 to perform quantitative trait locus mapping to identify candidate genes controlling agronomic and biochemical phenotypes. The parental alleles for one candidate gene were functionally validated by expressing the alleles in yeast. The resulting paper was published in *GENETICS* (Woods et al. 2021).

Chapter 3: Genome wide polymorphism and genic selection in feral and domesticated lineages of Cannabis sativa

In this chapter I used a reverse genetics approach to understanding the genetics of evolution and local adaptation in both feral and domesticated lineages of *Cannabis sativa*. Using whole genome sequencing data, I identified evidence supporting the Asian ancestry hypothesis of this species and found derived alleles in genes potentially contributing to local adaptation in feral and domesticated populations of this species. The resulting paper was published in *G3: Genes | Genomes | Genetics* (Woods et al. 2022a).

Chapter 4: Root pulling force across drought in Maize reveals genotype by environment interactions and candidate genes

In this chapter, I used a forward genetic approach to describe the quantitative genetic architecture of root system architecture in maize. Using a diversity panel comprising of more than 350 inbred maize lines, genome wide associations were performed to identify candidate genes

associated with measurable variation in numerous root phenotypes. Genotype by environment interactions for maize root traits were also estimated by performing the experiment in both irrigated and drought stressed environments. The resulting paper was published in *Frontiers in Plant Science* (Woods et al. 2022b).

REFERENCES

- Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., R. Ashnest, J., ... Sharbel, T. F. (2020). The Genomics of Cannabis and Its Close Relatives . *Annual Review of Plant Biology*. <https://doi.org/10.1146/annurev-arplant-081519-040203>
- Kumar, J., Gupta, D. Sen, Gupta, S., Dubey, S., Gupta, P., & Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*. <https://doi.org/10.1007/s00299-017-2127-y>
- Lawrence, C. J., Seigfried, T. E., & Brendel, V. (2005). The Maize Genetics and Genomics Database. The community resource for access to diverse maize data. *Plant Physiology*. <https://doi.org/10.1104/pp.104.059196>
- Lynch M, Walsh B. 1998. Genetics and Analysis of Quantitative Traits. Sunderland: Sinauer Associates.
- Gurumurthy, C. B., Grati, M., Ohtsuka, M., Schilit, S. L. P., Quadros, R. M., & Liu, X. Z. (2016). CRISPR: a versatile tool for both forward and reverse genetics research. *Human Genetics*. <https://doi.org/10.1007/s00439-016-1704-4>
- Speed, D., & Balding, D. J. (2012). Understanding complex traits: from farmers to pharma. *Genome Medicine*. <https://doi.org/10.1186/gm360>
- 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
- Woods, P., Campbell, B. J., Nicodemus, T. J., Cahoon, E. B., Mullen, J. L., & McKay, J. K. (2021). Quantitative trait loci controlling agronomic and biochemical traits in Cannabis sativa . *Genetics*. <https://doi.org/10.1093/genetics/iyab099>
- Woods, P., Price, N., Matthews, P., & McKay, J. K. (2022a). Genome-wide polymorphism and genic selection in feral and domesticated lineages of Cannabis sativa. *G3 Genes|Genomes|Genetics*, jkac209. <https://doi.org/10.1093/g3journal/jkac209>
- Woods, P., Lehner, K. R., Hein, K., Mullen, J. L., & McKay, J. K. (2022b). Root Pulling Force Across Drought in Maize Reveals Genotype by Environment Interactions and Candidate Genes. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.883209>

CHAPTER 2

QUANTITATIVE TRAIT LOCI CONTROLLING AGRONOMIC AND BIOCHEMICAL TRAITS IN *CANNABIS SATIVA*

Summary

Understanding the genetic basis of complex traits is a fundamental goal of evolutionary genetics. Yet, the genetics controlling complex traits in many important species such as hemp (*Cannabis sativa*) remain poorly investigated. Since hemp's change in legal status with the 2014 and 2018 U.S. Federal Farm Bills, interest in the genetics controlling its numerous agriculturally important traits has steadily increased. To better understand the genetics of agriculturally important traits in hemp, we developed an F₂ population by crossing two phenotypically distinct hemp cultivars (Carmagnola and USO31). Using whole genome sequencing, we mapped quantitative trait loci (QTL) associated with variation in numerous agronomic and biochemical traits. A total of 69 loci associated with agronomic (34) and biochemical (35) trait variation were identified. We found that most QTL co-localized, suggesting that the phenotypic distinctions between Carmagnola and USO31 are largely controlled by a small number of loci. We identified *TINY* and olivetol synthase as candidate genes underlying co-localized QTL clusters for agronomic and biochemical traits respectively. We functionally validated the olivetol synthase candidate by expressing the alleles in yeast. Gas chromatography-mass spectrometry assays of extracts from these yeast colonies suggest that the USO31 olivetol synthase is functionally less active and potentially explains why USO31 produces lower cannabinoids compared to Carmagnola. Overall, our results help modernize the genomic understanding of complex traits in hemp.

Introduction

A long-term goal of genetics and evolutionary genetics is to understand the genetic basis of complex traits. In the past, studies have approached this goal by using molecular markers to investigate fundamental questions such as: for any given trait, how many loci control variation; are these loci dominant; and is variation additive, or do epistatic interactions explain a large proportion of phenotypic variance (Hill, 2010)? Despite nearly 30 years of using molecular markers, our understanding of complex trait genetics remains incomplete because of the limited capacity for high resolution mapping of loci (MacKay et al. 2009). Now in the genomics era, with the ease of sequencing whole genomes, this long-term goal is more feasible since studies have an improved ability to dissect the genetic architecture of complex traits (Mackay et al. 2009). As a result, it is becoming increasingly common for studies to combine whole genome sequencing (WGS) with bi-parental mapping populations to identify quantitative trait loci (QTL) controlling variation in complex traits in numerous species (Mojica et al. 2016, Yang et al. 2017, Burga et al. 2019).

Historically, plants have been widely used to study fundamental questions related to the genetics of complex traits because of the possibility for extensive experimental design and control of environments (Speed & Balding 2012). In many staple crop species such as *Maize* and rice, the genetic understanding of complex traits has improved steadily in recent years because of the well-established genetic resources and dense volume of literature. For less studied crop species such as industrial hemp (*Cannabis sativa*), the availability of genetic resources and literature is narrow, which limits the capacity to understand complex traits. To help establish a basic quantitative genetic understanding in crops such as industrial hemp, studies that investigate fundamental questions regarding the genetics of these crop's complex traits are needed. Hemp is a scientifically

interesting plant and valuable crop, producing a high yield of plant biomass including stalks, bast fibers (used in building materials, composites and textiles) and a high protein and lipid grain with unique nutritional properties. *Cannabis sativa* is also the sister species of hops (*Humulus lupulus* Kovalchuk et al. 2020) and similarly produces an array of secondary metabolites that have numerous potential uses. Additionally, hemp is interesting because it is in a clade where it evolved dioecy and an annual habit from progenitors which were monoecious and perennial (Kovalchuk et al. 2020).

Since its initial change in legal status in the 2014 Federal Farm Bill and subsequent broadening of those rules in the 2018 Federal Farm Bill (U.S. Govt., 2014 and U.S. Govt., 2018), interest in cultivating and researching industrial hemp has steadily increased in the United States. In Canada, the European Union and the United States, hemp is legally defined as *C. sativa* plants with a total tetrahydrocannabinol (THC) content less than 0.3%. Cultivars of *C. sativa* with a total THC content above 0.3% are federally illegal and such plants are scheduled as a controlled substance known as marijuana. Given its importance for regulation, there have been a number of studies focused on trying to understand the biochemistry of THC synthesis (Sirikantaramas et al. 2004, Sirikantaramas et al. 2005 and Zirpel et al. 2018). THC is not the only biochemical trait that is important in hemp. There is a growing market for terpenes (e.g., alpha-pinene) and other cannabinoids (e.g., cannabidiol [CBD]) outside of THC, that have value for their medicinal and therapeutic properties, use as chemosensory additives, natural pesticides, and other potential uses (Russo, 2011 and Gallily et al. 2015). Agronomic traits like grain yield and plant biomass are also key breeding targets to make hemp a competitive grain and fiber crop. Despite a number of agronomic studies on industrial hemp (Van der Werf et al. 1995a, Van der Werf et al. 1995b, and Struik et al. 2000), literature that investigates the genetic factors that contribute to the variation in

these traits has only recently begun to emerge (Petit et al. 2020a, b, c). Petit et al. (2020a, b, c) used a diversity mapping panel of hemp cultivars to perform genome wide associations to identify additive loci and genotype by environment interactions associated with variation in numerous fiber quality and flowering traits.

To date, two studies have utilized bi-parental mapping populations to identify QTL associated with variation in biochemical traits in *C. sativa*. Both of these studies utilized the same mapping population derived from a cross between hemp and marijuana cultivars (Weiblen et al. 2015 and Grassa et al. 2021). Weiblen et al. 2015 identified a single large effect QTL associated with chemotype ratios for THC and CBD (Weiblen et al. 2015). In a follow-up study using the same mapping population, Grassa et al. 2021 mapped for loci associated with cannabinoid variation and were able to identify two candidate genes possibly linked to their QTL.

In this study, QTL contributing to the variation of important agronomic and of biochemical traits were characterized in *C. sativa* utilizing an F₂ mapping population derived from a cross between two foundational hemp cultivars bred for different markets as well as developed in different countries: Carmagnola and USO31. Carmagnola is a dioecious fiber cultivar developed in Italy and USO31 is a monoecious dual-purpose cultivar developed in Ukraine, bred for both grain and fiber production (Salentijn et al., 2015). Carmagnola produces late-flowering, tall plants that are typical of fiber cultivars, while USO31 is an early maturing, shorter-statured cultivar that is more suitable for grain cropping. We characterized several QTL by using WGS to identify segregating variants that span the *C. sativa* genome and phenotyping of numerous agronomic and biochemical traits. Our results identify numerous QTL of varying effect size, co-located QTL, and candidate genes underlying two co-located QTL clusters.

Materials and Methods

Mapping population creation

Seed of the cultivars of industrial hemp Carmagnola and USO31 were imported from Italy to Colorado in 2015 as part of a set of variety trials (Campbell et al. 2019). Seeds of Carmagnola and USO31 were sown in Promix potting soil (Premier Horticulture, Quakertown, PA, USA), in a Conviron E8 growth chamber at the Colorado State University greenhouse. Growth chambers were initially set to a 20:4 hour (light:dark) regimen in order to keep plants in the vegetative stage. During flowering, the light regimen was changed to a 12:12 hour regimen. Daytime and nighttime temperatures were kept at 23 degrees C and relative humidity was kept at 40%. Growth chamber light intensity was kept at $330 \mu\text{mol m}^{-2} \text{s}^{-1}$. Plants were watered with a full strength (1-1-1) vegetative nutrient solution (General Hydroponics Flora Series, Sebastopol, CA, USA). Healthy and representative plants of Carmagnola and USO31 were chosen as parents of a bi-parental QTL mapping population. Pollen from a monoecious USO31 plant was crossed to a female Carmagnola plant. F₁ seed was grown using the same methods as for the parent plants. A single monoecious healthy F₁ plant was then self-fertilized to produce the F₂ mapping population. Propagated clones were taken from the original parent and F₁ plants to use later in the field experiment by taking cuttings. These clonal cuttings were then dipped into cloning solution (Olivia's Solutions Cloning Solution, Calistoga, CA, USA), planted in Promix potting soil, kept under humidity domes and watered as needed with a full strength (1-1-1) vegetative nutrient solution (General Hydroponics Flora Series, Sebastopol, CA, USA). These clones were kept in their vegetive stage in the same growth chamber using the 20:4 regime. F₂ seed was germinated within rockwool plugs (Grodan, Roermond, the Netherlands) in the Colorado State University greenhouse in May 2017. No light supplementation occurred during the four weeks that F₂ seed were in the greenhouse. In order to

have replication of F₂ lines, clones were taken from each seedling so that each line could be replicated three times in the experiment. Clonal propagation of the F₂ plants was conducted using the same method as for the parent and F₁ plants. Once three clonal propagations of each F₂ plant were obtained, the parent, F₁ and F₂ plants were transplanted to a field located at the Colorado State University Agricultural Research and Education Center (ARDEC).

Field experiment

The experiment was conducted at ARDEC located in Fort Collins, Colorado. Prior to transplanting clones, glyphosate (RoundUp, Powermax, Monsanto) and dicamba (Sterling Blue, Winfield United) were applied (15.70 gallons per acre) to clear the field of any existing weeds. Clones representing 372 F₂ lines along with clones of the parents and F₁ were transplanted from the greenhouse into the field by hand in June 2017. Plants were spaced 1.5 m apart in both directions to avoid interplant competition. A Latinized row-column design was utilized to minimize spatial bias. The experiment was replicated three times, with one clone of each F₂ line and 3 clones of each parental and F₁ 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. Weed pressure was controlled manually and no pesticides were applied during the growing season. 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. Plots were hand watered with a hose due to breakage of the overhead linear irrigation sprinkler system.

The date that each plant reached initiation of 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 and were calculated. Plant maturity was considered as seed maturity, i.e., when bracts began

to dehisce and darkening of the seed coat was visible as described by Campbell et al. 2019. Mature plants were harvested within three days.

To measure leaf water content, one fully expanded and undamaged leaf was randomly selected from the middle of the primary stem of each plant at a single time point during the vegetative growth stage of the plant 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.

Before harvest, plant height was measured as the vertical distance from the soil surface to the tallest naturally occurring part of a plant.

Plants were cut at the soil surface and air-dried for a minimum of 30 days. Total plant biomass (dry biomass) was measured as the mass of the aboveground portion of the plant material. Stems were weighed separately after threshing to determine stem biomass. The dried stems were measured at the widest part of the base with digital calipers to determine stem diameter.

Grain was separated from inflorescences by hand and seed was cleaned using a column blower (Agricullex, 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). A subsample of 50 seeds was counted from each sample to extrapolate Thousand Seed Mass.

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 ultra-high-pressure liquid chromatography (Waters UPLC) and gas chromatography (Shimadzu GC-2014) with flame ionization detector (GC-FID) by ProVerde Labs (Milford, MA). 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.

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 standards (Cerilliant, Sigma-Aldrich and Cayman Chemicals) were used for peak identification and generation of calibration curves used for quantitation, and included: THC acid (THCa), CBD acid (CBDa), cannabigerolic acid (CBGa), and cannabichromene (CBC). Data was recorded and processed using Empower Software (Version 3, Waters Corporation).

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

Gas chromatography was performed using Shimadzu GC-2014 gas chromatograph with Flame Ionization Detection (FID), with a Rxi-624Sil MS column (30 m x 0.25 mm x 1.4 μ m) (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).

Whole-genome sequencing

DNA was extracted using the Qiagen DNeasy Plant Mini Kit (Valencia, CA) and then quantified using a Qubit Fluorometer (ThermoFisher Scientific). A total of 375 samples were whole genome sequenced (2x150bp paired-end reads) using Illumina Nextera library preparation system. Sequencing efforts aimed for 30x, 15x and 7x coverage of the parents, F₁ and F₂ progeny respectively. All samples were sequenced at the University of Colorado Anschutz Medical Campus using an Illumina NovaSeq.

Raw sequence data were evaluated with FastQC (Andrews S. 2010, version 0.11.8) to assess read quality and adapter contamination. Trimmomatic (Bolger et al. 2014, version 0.39) was then used with default parameters to remove low quality reads and any adapter contamination identified in the FastQC report. The trimmed sequence reads were then aligned to version 2 of the

Finola reference genome (Lavery et al. 2018, GenBank assembly accession ID = GCA_003417725.2) using BWA-MEM with the default settings (Li 2013, version 0.7.17). Samtools (Li et al. 2009, version 1.9) was then used to sort sequence alignment files and mark duplicate reads. BCFtools (Narasimhan et al. 2016, version 1.9) was then used with default parameters to identify genetic variants using both the “mpileup” and “call” functions to produce three separate variant call files (VCF) for the Carmagnola/USO31 parents, the F₁, and F₂ respectively.

BCFtools was used to filter the F₂ VCF to contain biallelic single nucleotide polymorphisms (SNPs) that possessed a genotyping rate of $\geq 75\%$ across individuals, quality of ≥ 30 , base quality bias of ≥ 0.8 , base position bias of ≥ 0.8 and mapping quality of ≥ 60 . VCFtools (Danecek et al. 2011, version 0.1.16) was then used to filter the F₂ VCF to contain loci with genotype frequencies resembling 1:2:1 Mendelian segregation ratios by incorporating an exact test with a p-value threshold of 0.05 followed by a minor allele frequency filter of 0.4. The BCFtools command “isec” was then used to extract the filtered F₂ VCF loci from the Carmagnola/USO31 parent and F₁ VCF files. The parent, F₁ and F₂ VCF files were then filtered again with BCFtools to contain only loci where Carmagnola and USO31 possessed alternate homozygous SNPs with quality ≥ 200 , read depth of ≥ 50 and phred-scaled genotype quality ≥ 99 for which the F₁ was also heterozygous. All three VCF files were then merged to contain a total of 1827 SNPs across all samples. Using the “VariantsToTable” command from the Genome Analysis Toolkit (McKenna et al. 2010, version 4.1.4.0), the merged VCF was then exported to a tab separated file format. This tab-separated file was read into excel and the F₂ genotypes were manually converted to the “a” (Carmagnola), “b” (USO31), and “h” (heterozygote) genetic linkage map format.

Quantitative trait loci mapping and trait correlations

A genetic linkage map was created using JoinMap 4 (Van Ooijen 2006), with markers assigned to linkage groups based on a recombination frequency threshold of 0.25. We identified 10 linkage groups, corresponding to the 10 chromosomes from Lavery et al. 2018, whose numbering convention we used. The markers on the 10 linkage groups were mapped using the regression mapping algorithm and Kosambi mapping function. A total of 10 duplicate markers were identified and removed, for a remaining total of 1817 markers in the genetic map. All QTL mapping was conducted in R/qtl (Broman et al. 2003, version 1.44 - 9). Recombination frequencies calculated from JoinMap and the R package “qtlTools” (Lovell 2019, version 1.2.0) were used to estimate the 1.5 log of odds (LOD) QTL location confidence intervals.

We calculated the simple means of each F₂ line’s phenotype to use for downstream analyses. Raw mean phenotype data were assessed for normality using the Shapiro-Wilk test base R function. Since no trait’s distribution passed the Shapiro-Wilk normality test, the raw phenotype data were quantile normalized to better fit assumptions of normality. For traits reported, there was no substantial differences in QTL between raw and normalized data. Multiple QTL models for normally distributed traits were selected using the STEPWISE.QTL(max.qtl = 6) function with penalties based on 1000 permutations. QTL models for traits that could not be adequately transformed were constructed using the significant peak locations based on 1000 permutations identified from the SCANONE output. QTL peak positions obtained from SCANONE were further refined using REFINEQTL. Significance and effect sizes of QTL in models were validated using FITQTL. QTL that did not explain a significant proportion of variance ($p > 0.05$) were removed from models.

To test for significant correlations ($p < 0.05$) among measured traits, we used the “rstatix” package (Kassambara 2020, version 0.6.0) in R (R Core Team 2019, version 3.6.0) to obtain Spearman’s rank correlation coefficient (P). Correlation coefficients were then organized into a matrix and plotted using ggplot2 (Wickham 2016, version 3.3.2).

QTL candidate gene identification

To identify candidate genes of agronomic traits, we focused on identifying genes underlying QTL with narrow 1.5 LOD intervals spanning ~ 50,000 base pairs or less. To investigate these narrow QTL intervals and their close surrounding regions, we extracted the reference sequences contained in these LOD intervals plus an extra 15,000 base pairs on each flanking end. Since there was no gene annotation available for version 2 of the Finola assembly, we used the AUGUSTUS (Stanke et al. 2004) output from BUSCO (Simão et al. 2015) to predict the models of potential genes located within these extracted regions. Basic local alignment search tool (BLAST) was then used to identify homologs to the predicted gene model sequences. Predicted gene model sequences that showed strong homolog matches (100% coverage and $\geq 95\%$ identity) were then further investigated for sequence variation between Carmagnola and USO31 in the VCF. The predicted gene model sequences were then annotated with identified SNPs using Geneious Prime (Kearse et al. 2012, version 2020.1.2).

To identify candidate genes of biochemical traits, we focused on identifying genes underlying the co-located QTL clusters. Using published coding sequences of genes involved in terpene and cannabinoid biosynthesis, we used BLAST to identify alignments within the biochemical trait QTL intervals. Gene sequences that had high coverage, identity, and functional relevance to traits included in QTL intervals were then investigated for genetic differences between Carmagnola and USO31 within the VCF. Sequence variation identified in the VCF were then

confirmed in Carmagnola and USO31 using Sanger sequencing. Gene sequences with confirmed genetic variation were then annotated with SNPs using Geneious Prime (Kearse et al. 2012, version 2020.1.2).

Olivetol synthase functional assay

Olivetol synthase coding sequences from Carmagnola and USO31 were synthesized by Twist Bioscience with codon optimization for *Saccharomyces cerevisiae* and addition of flanking *Bam*HI (5') and *Not*I (3') restriction sites. We used the published and functionally validated olivetol synthase coding sequence from Taura et al. 2009 as the basis for sites where Carmagnola and USO31 did not genetically differ. Only overlapping genetic variation between Carmagnola and USO31 from our Illumina and Sanger sequence data were incorporated into the sequences. Thus, we denoted these two olivetol synthase alleles as “Carmagnola-derived” and “USO31-derived”. At the amino acid scale, the USO31-derived sequence was identical to the Taura et al. 2009 OLS while the Carmagnola-derived OLS differed by 9 amino acids. Synthesized genes were cloned into the *Bam*HI/*Not*I sites of the pYES2 expression vector containing a *GAL*I promoter for galactose-inducible expression of the inserted genes. The resulting constructs and the empty pYES2 vector were introduced into *S. cerevisiae* BY4741 cells using lithium acetate/polyethylene glycol transformation with the Frozen-EZ Yeast Transformation II Kit (Zymo). Colonies selected on -uracil (-URA) dropout media were grown in 3 ml liquid cultures in media lacking uracil and containing raffinose as a non-inductive carbon source [0.17% (w/v) yeast nitrogen base without amino acids, 0.08% (w/v) CSM-URA, 0.5% (w/v) ammonium sulfate, 2% (w/v) raffinose]. For induction, raffinose was substituted with galactose [2% (w/v)], and cultures were initiated in 50 ml of media in 250 ml Erlenmeyer flasks at a density of 0.1 optical density. Cells were maintained at 25°C with shaking (130 rpm). Hexanoic acid (NuChek Prep) was added at a concentration of 1

mM to cultures after 24 h of growth. Cultures were sampled in 6 ml aliquots at two-day intervals over a six-day time course. Results provided are from an experiment with three independent cultures for each treatment, and experiments were repeated three times with similar trends.

For olivetol analyses, pelleted cells were extracted in 2 ml of chloroform with 30 min of incubation in a sonicating water bath (Branson 2800). Following sonication, tubes were centrifuged at 16,000 x g for 10 min. The solvent was transferred to a glass screw cap tube, dried under N₂, and dissolved in 100 µl of chloroform. Olivetol in extracts was identified and quantified by analysis on an Agilent 7890A gas chromatograph (GC) interfaced with an Agilent 5975C mass selective detector fitted with an Agilent HP-5 column (30 m length x 0.25 mm outer diameter, 0.25 µm film thickness). The inlet temperature was 250°C and a 9 ml/min flow rate of H₂ carrier with the oven programmed for 90°C for 1 min followed by a 30°C/min temperature ramp to 300°C. The olivetol product was identified by the 124 *m/z* diagnostic ion fragment and 180 *m/z* molecular ion and by retention time and mass spectrum identical to those of an authentic olivetol standard (Sigma Aldrich). Olivetol production was quantified using a standard curve derived from the olivetol standard.

Results

Trait values and correlations

In our field experiment, we measured a total of eight agronomic and seventeen biochemical traits. A summary of trait values for the parent, F₁, and F₂ plants can be found in Supplementary Table 1. For nearly all traits measured, Carmagnola exhibited higher trait values compared to USO31. Since the parents of the population we developed were traditional fiber and seed industrial hemp cultivars, their production of biochemical traits was modest compared to cultivars that have been specifically bred for cannabinoid and terpene production. We also note that pollination and

seed set may have also reduced production of biochemical traits (Mehmedic et al. 2010). F₂ population mean trait values were generally intermediate relative to the Carmagnola and USO31 parents. The range for most F₂ traits extended beyond the mean trait values of Carmagnola and USO31. Some biochemical traits, such as citronellol and geraniol produced no detectible quantities in the parents or F₁ but did show a distribution of detectible quantities among the F₂ population. As described in the methods, F₂ genotypes were replicated by vegetative propagation, and then clones were transplanted into an agricultural field, where survival was low. In total 256, 170 and 238 F₂ plants were phenotyped in replicate blocks 1, 2 and 3 respectively. Transplant death appeared to be random with respect to genotype and thus resulted in many F₂ lines having reduced replication or only a single observation. While this did not prevent us from detecting QTL and fitting polygenic models for most traits, our experiment's power was reduced and thus may have inhibited our ability to detect a larger number small effect QTL.

Significant ($p < 0.05$) and positive correlations were observed among all agronomic traits measured (Figure 1 A and Supplementary Table 2). Dry biomass and stem biomass exhibited the strongest correlation strength ($\rho = 0.95$). Correlations between agronomic and biochemical traits were low, with the strongest between days to maturity and CBC ($\rho = -0.23$, Figure 1 A and Supplementary Table 2).

Most of the 17 biochemical traits were positively correlated (Figure 1 A and Supplementary Table 2). Of the cannabinoids measured, CBDa and THCa were the most associated ($\rho = 0.85$, Figure 1 A). Correlations between terpenes were largely positive with gamma and alpha terpinene being the most associated ($\rho = 0.90$, Figure 1 A) among all biochemical traits. Citronellol, geraniol

and caryophyllene oxide, which were inversely correlated with most other biochemical traits, displayed significant positive correlations.

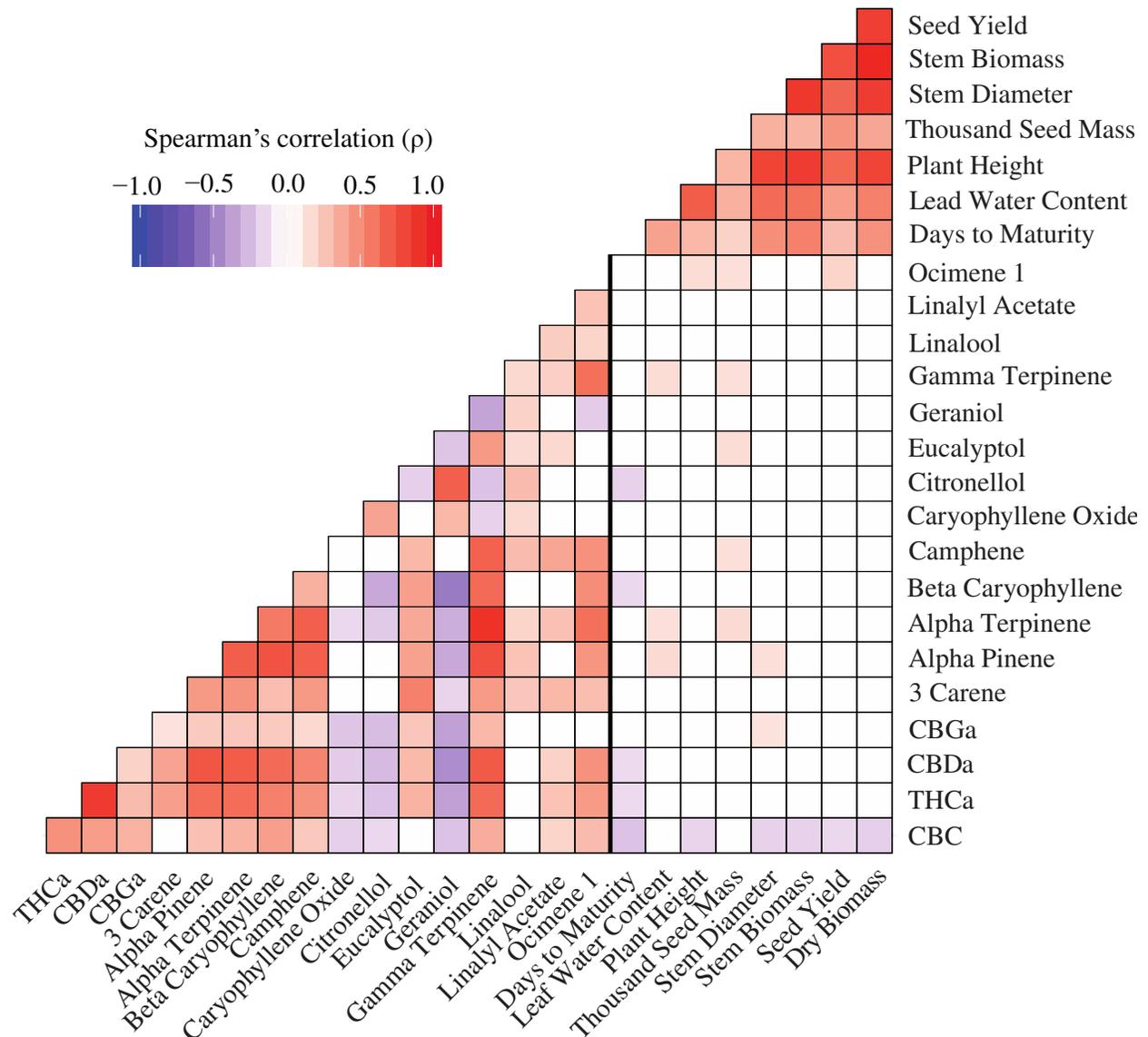


Figure 1: Trait correlations of measured phenotypes. Correlation plot depicting Spearman's rank correlation coefficient (ρ), between measured phenotypes in F₂ population. Red colors indicate positive correlations while blue colors indicate negative correlations. The vertical bolded black line separates biochemical from agronomic traits. Only significant correlations ($p < 0.05$) are shown.

QTL mapping

Our genetic map identified 10 linkage groups consistent with the 10 chromosomes identified by Lavery et al. 2018, which we used for naming our linkage groups. Mapping of QTL identified a total of 69 loci associated with measurable variation in agronomic (34) and biochemical traits (35). Of these 69 QTL identified, we found that numerous agronomic and biochemical QTL co-localized across our linkage map. In total, four agronomic and two biochemical QTL co-localized clusters (Figure 2 A and Figure 2 B) were identified. Henceforth, we refer to each of these co-located QTL clusters by their linkage group name followed by their average genetic position (e.g., LG3.60, Figure 2 A).

Individual QTL models for agronomic traits are shown in Table 1. Agronomic trait QTL models were largely additive with the exception of stem diameter which had the most complex model that included a significant interaction between QTL on linkage groups 3 and 9 (Supplementary Figure 1). Leaf water content and plant height exhibited the simplest models consisting of three QTL. For 31 of the 34 agronomic QTL identified, F₂ plants homozygous for the Carmagnola allele exhibited higher trait values (Supplementary Table 3). F₂ plants homozygous for the USO31 allele at two QTL (SY.1 and DTM.4) had higher trait values (Supplementary Figures 2 and 3). F₂ plants that were heterozygous at QTL generally exhibited either additivity or dominance among alleles at each agronomic QTL. The only exception to this pattern was a single QTL for thousand seed mass (TSM.2) which displayed genotype-phenotype patterns suggestive of overdominance (Supplementary Figure 4). Agronomic co-located QTL clusters explained ranges of 5.22% - 22.35% (LG3.60), 3.91% - 5.92% (LG4.50), 12.92% - 34.27% (LG5.05) and 3.81% - 5.65% (LG10.25) of variance across traits with a detected QTL in these clusters (Supplementary Table 4).

Individual QTL models for all biochemical traits are shown in Table 2. Biochemical trait QTL models were overall more complex with multiple significant QTL interactions identified. Alpha-pinene had the most complex model that consisted of 6 additive QTL and two separate interactions. CBC, citronellol, geraniol, linalool and linalyl acetate exhibited the simplest models consisting of a single QTL. Of the 35 biochemical QTL, 30 exhibited a pattern similar to agronomic QTL where F₂ plants homozygous for the Carmagnola allele produced greater trait values (Supplementary Table 5). Models for CBGa, citronellol, geraniol, ocimene 1 and caryophyllene oxide all contained at least one QTL for which F₂ plants homozygous for the USO31 allele exhibited greater trait values. F₂ plants that were heterozygous at the QTL exhibited evidence for either additivity or dominance among alleles at each biochemical QTL. Biochemical co-located QTL clusters explained ranges of 9.35% - 28.21% (LG6.35) and 10.23 - 44.79% (LG9.40) of variance across traits with a detected QTL in these clusters (Supplementary Table 4). These co-located QTL clusters were comprised of both cannabinoids and terpenes. For 8 of the 17 measured biochemical traits, we detected a significant interaction between QTL in LG6.35 and LG9.40 which explained a range of 3.89% - 15.11% of variance observed across these 8 phenotypes. Figure 3 shows this interaction between QTL within LG6.35 and LG9.40 for CBDa production.

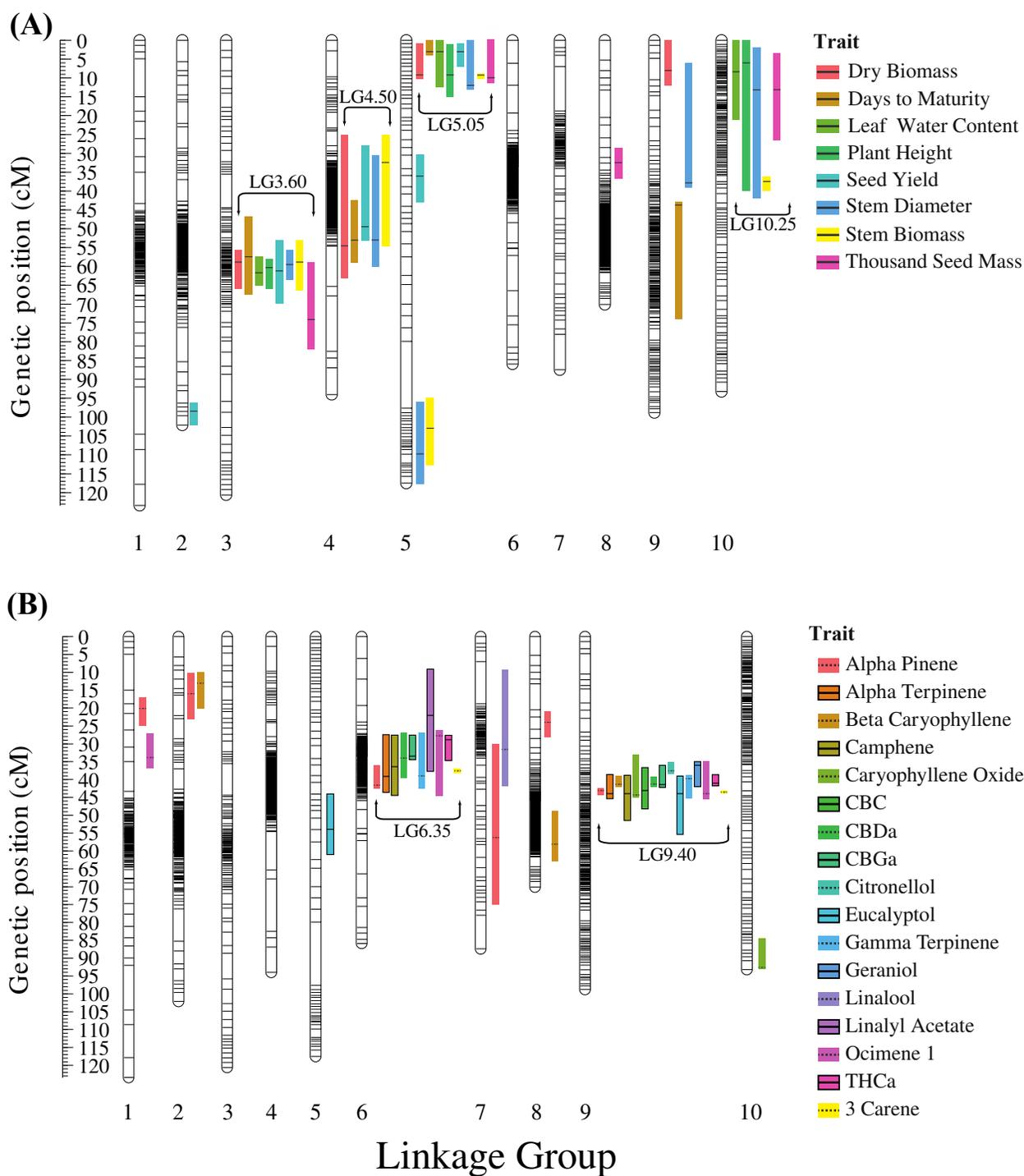


Figure 2: Linkage map and QTL intervals. (A) Linkage map showing boxplots depicting the 1.5 LOD confidence intervals of QTL for the 8 measured agronomic traits. (B) Linkage map showing boxplots depicting the 1.5 LOD confidence intervals of QTL for the 17 measured biochemical traits. Black bars within each boxplot indicate the location of the peak LOD values while box colors indicate QTL identified for the respectively colored trait.

Table 1. QTL models, locations and effect sizes for agronomic traits.

Phenotype	QTL	Linkage Group	Marker (bp)	Genetic Position (cM)	LOD	Variance Explained (%)
Leaf Water Content (g)	LWC.1	3	50700151	61.73	14.65	16.25
	LWC.2	5	23530617	3.00	12.37	13.48
	LWC.3	10	58605089	8.27	4.23	4.32
Plant Height (cm)	PLHT.1	3	46248568	60.33	19.41	22.35
	PLHT.2	5	31242637	9.30	14.01	15.38
	PLHT.3	10	74603632	5.88	4.40	4.45
Thousand Seed Mass (g)	TSM.1	3	89444283	74.00	8.44	8.29
	TSM.2	5	31294490	10.00	17.22	18.16
	TSM.3	8	14446033	32.62	6.93	6.73
	TSM.4	10	16666160	13.12	5.78	5.56
Stem Diameter (cm)	SD.1	3	65433658	59.44	18.86	15.71
	SD.2	4	86599640	53.04	6.38	4.80
	SD.3	5	41915966	12.00	23.29	20.13
	SD.4	5	84759484	109.77	4.45	3.30
	SD.5	9	1825282	38.22	8.42	6.44
	SD.6	10	16666160	13.12	5.11	3.81
	SD.1:SD.5				5.84	4.38
Stem Biomass (g)	SB.1	3	22597724	58.82	11.24	9.01
	SB.2	4	15873733	32.36	7.60	5.92
	SB.3	5	31242637	9.30	32.48	31.05
	SB.4	5	84871972	102.83	4.46	3.39
	SB.5	10	24849063	37.50	7.26	5.65
Seed Yield (g)	SY.1	2	94185707	98.47	12.59	11.35
	SY.2	3	41816369	61.18	6.72	5.78
	SY.3	4	84450770	49.52	4.92	4.17
	SY.4	5	23530617	3.00	14.15	12.92
	SY.5	5	57357245	36.00	4.04	3.41
Dry Biomass (g)	DB.1	3	22597724	58.82	8.30	8.04
	DB.2	4	86642457	54.30	4.17	3.91
	DB.3	5	31242637	9.30	22.29	24.19
	DB.4	9	8629041	8.00	4.17	3.91
Days to Maturity	DTM.1	3	26862649	57.70	6.30	5.22
	DTM.2	4	86599640	53.00	6.05	5.01
	DTM.3	5	23530617	3.00	33.10	34.27
	DTM.4	9	3462161	43.60	5.24	4.31

Markers are reported as the physical base pair position of the linkage map marker closest to the LOD peak of the respective QTL.

Genetic positions (centiMorgans), LOD values, and variance estimates of QTL have been rounded to two decimal places.

Colons indicate interactions between the specified QTL.

Table 2. QTL models, location and effect sizes for biochemical traits.

Phenotype	QTL	Linkage Group	Marker (bp)	Genetic Position (cM)	LOD	Variance Explained (%)
Alpha Pinene (ppm)	AP.1	1	730085	20.00	12.11	9.90
	AP.2	2	1811866	15.80	5.94	4.54
	AP.3	6	65509552	41.60	15.76	13.43
	AP.4	7	71316711	56.49	4.84	3.65
	AP.5	8	3907307	24.00	8.64	6.80
	AP.6	9	2778113	43.00	32.09	31.04
	AP.1:AP.5 AP.3:AP.6				6.71 9.61	5.16 7.64
Alpha Terpinene (ppm)	AT.1	6	19863635	39.10	10.17	15.70
	AT.2	9	2941221	44.00	14.31	23.15
	AT.1:AT.2				3.51	5.04
Beta Caryophyllene (ppm)	BC.1	2	2134594	13.00	9.86	9.20
	BC.2	8	47143084	58.32	6.37	5.72
	BC.3	9	2323644	41.28	35.06	43.88
Camphene (ppm)	CAM.1	6	17849490	36.29	10.51	17.36
	CAM.2	9	2941221	44.00	11.40	19.01
	CAM.1:CAM.2				4.65	7.19
Caryophyllene Oxide (ppm)	CO.1	9	2992284	44.38	11.75	20.42
	CO.2	10	34580106	93.00	10.59	18.18
	CO.1:CO.2				6.91	11.38
CBC (%)	CBC.1	9	2778113	43.00	6.56	12.57
CBDa (%)	CBDa.1	6	29867178	33.73	32.56	28.21
	CBDa.2	9	2453237	41.10	44.85	44.79
	CBDa.1:CBDa.2				13.64	9.59
CBGa (%)	CBGa.1	6	45639112	33.37	20.32	26.46
	CBGa.2	9	2453237	41.10	25.53	35.21
	CBGa.1:CBGa.2				7.87	8.97
Citronellol (ppm)	CIT.1	9	2083967	37.46	12.41	23.44
Eucalyptol (ppm)	EUC.1	5	72066134	53.98	3.52	6.39
	EUC.2	9	2941221	44.00	5.52	10.23
Gamma Terpinene (ppm)	GT.1	6	19863635	39.10	7.64	11.03
	GT.2	9	2270995	39.39	17.44	28.12
	GT.1:GT.2				2.84	3.89
Geraniol (ppm)	GE.1	9	2083904	36.00	26.39	43.33
Linalool (ppm)	LI.1	7	35578586	31.65	6.12	12.34
Linalyl Acetate (ppm)	LA.1	6	5006390	22.00	4.56	9.35
Ocimene 1 (ppm)	OC.1	1	2554340	34.00	8.52	13.11
	OC.2	6	39795404	27.98	6.42	9.65
	OC.3	9	2941221	44.00	7.67	11.68
THCa (%)	THCa.1	6	39794190	28.90	20.58	24.54
	THCa.2	9	2453237	41.10	25.56	32.10
	THCa.1:THCa.2				8.04	8.36
3 Carene (ppm)	3C.1	6	55569529	37.18	16.34	26.70
	3C.2	9	3462212	43.48	15.23	24.59
	3C.1:3C.2				9.94	15.11

Markers are reported as the physical base pair position of the linkage map marker closest to the LOD peak of the respective QTL.

Genetic positions (centiMorgans), LOD values, and variance estimates of QTL have been rounded to two decimal places.

Colons indicate interactions between the specified QTL.

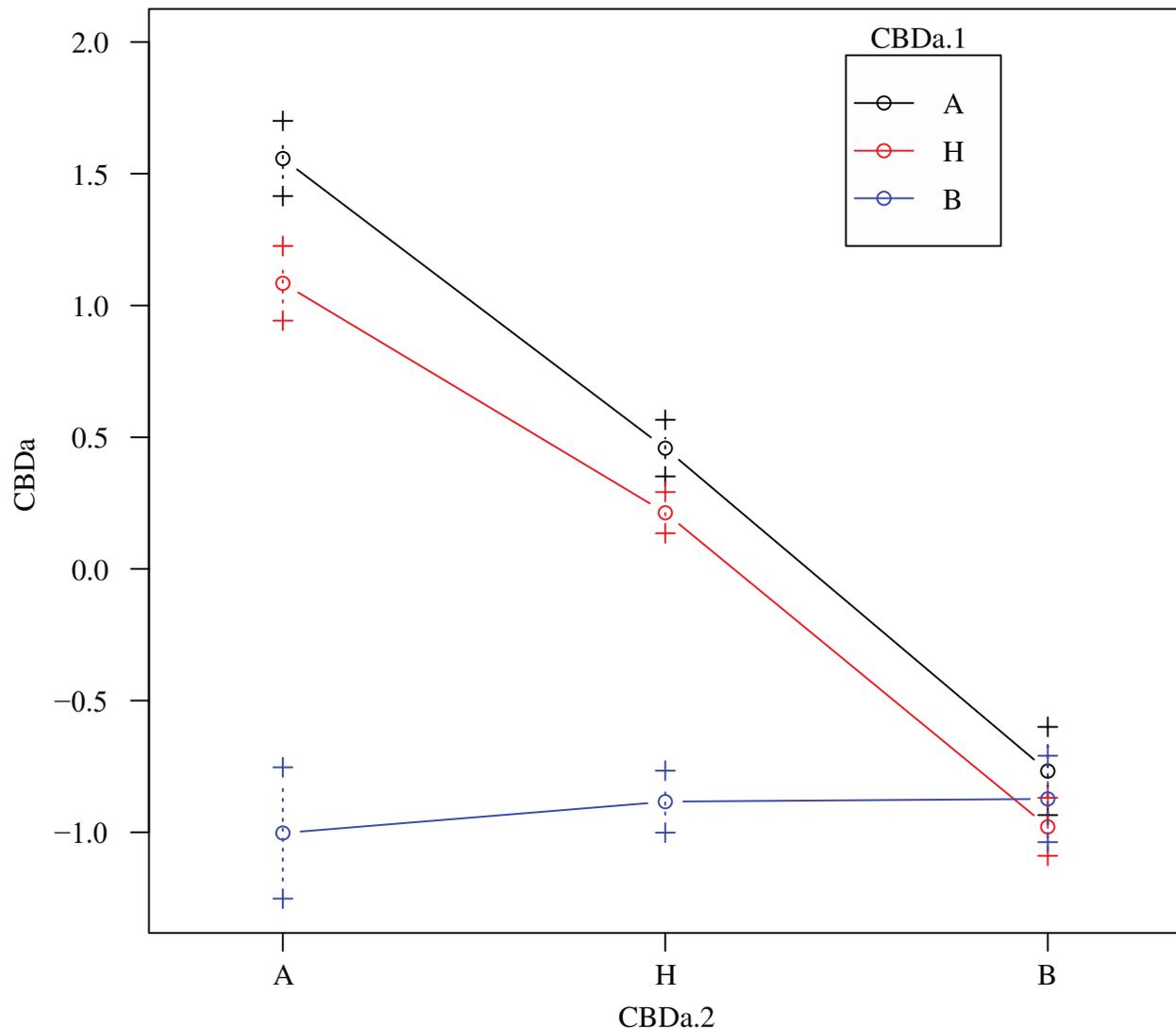


Figure 3: Epistatic interaction of the QTL clusters LG6.35 (CBDa.1) and LG9.40 (CBDa.2). Line colors indicate F₂ plant genotype at CBDa.1 while x-axis positions indicate F₂ plant genotype at CBDa.2 (A, Carmagnola; B, USO31). Y-axis values indicate the mean (\pm standard error) quantile normalized CBDa quantities. Lesser y-axis values indicate lower quantities while higher y-axis values indicate greater quantities of CBDa produced.

QTL candidate gene identification

Of the 34 agronomic QTL identified, SB.3 was the only QTL with a narrow 1.5 LOD interval which spanned \sim 50,000 base pairs (Figure 2 A). Our BUSCO analysis on the region covering this 1.5 LOD interval and its nearby surrounding sequence identified models for five predicted genes. BLAST analyses of the five predicted gene sequences resulted in one of these

predicted genes to have a homolog match to the ethylene responsive transcription factor *TINY* (Wilson et al. 1996 and Xie et al. 2019) with 100% coverage and 99.7% sequence identity. Sequences of the other four predicted gene models did not result in any homolog matches. Further inspection of the Carmagnola and USO31 *TINY* gene sequences identified a single non-synonymous homozygous SNP (Figure 4 A) not represented in the genotype matrix used to create the linkage map because of hard SNP filtering. Extraction and subsequent analysis of variance (ANOVA) of this non-synonymous SNP revealed significant differences ($p < 0.05$) among all F₂ agronomic trait means (Supplementary Table 6 and Supplementary Figure 5).

Using the published cDNA sequence from Taura et al. 2009, we identified olivetol synthase (OLS) as a candidate gene for LG9.40 which aligned inside this region with 100% coverage and 99.00% sequence identity. Comparison of the USO31 and Carmagnola OLS gene sequences revealed 9 non-synonymous homozygous SNPs in the coding regions (Figure 4 B). While no candidate gene linked to LG6.35 could be identified, we note that the partial isopentenyl-diphosphate delta-isomerase (synthesizes the substrate used for geranyl pyrophosphate synthesis) coding sequence from Booth et al. 2017 exhibited the greatest homology to a genomic region within LG6.35.

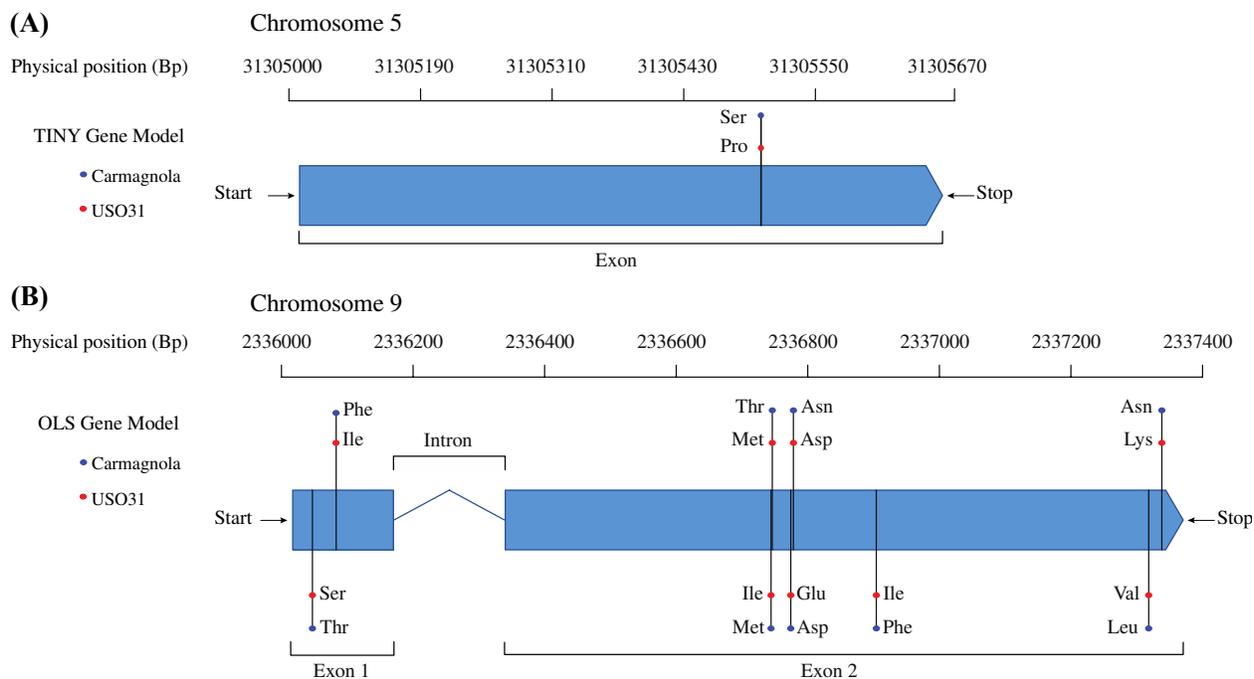


Figure 4: Models for candidate genes underlying LG5.05 and LG9.40. Shown are the physical locations of the *TINY* (A) and *OLS* (B) genes with annotations depicting the three letter codes for amino acid substitution(s) identified between Carmagnola (blue) and USO31 (red).

Functional validation of olivetol synthase alleles

We tested whether the variation in *OLS* alleles between the parents could affect biochemical production by expressing the alleles in *S. cerevisiae* (yeast) supplemented with the substrate hexanoic acid to measure olivetol production. Olivetol standard calibrations using GC-MS displayed a retention time of ~6.40 minutes, a diagnostic ionization fragment of 124 m/z and molecular ion of 180 m/z , which were used to identify olivetol produced by yeast colonies expressing the Carmagnola-derived and USO31-derived *OLS* alleles (Figure 5 C and D). At each sampling time point, yeast colonies expressing the Carmagnola-derived *OLS* allele produced significantly more ($p < 0.05$) olivetol quantities compared to yeast colonies expressing the USO31-

derived OLS allele (Figure 5 B), consistent with the effect of genotype at the QTL. Yeast colonies transformed with empty pYES2 vector did not produce detectable amounts of olivetol.

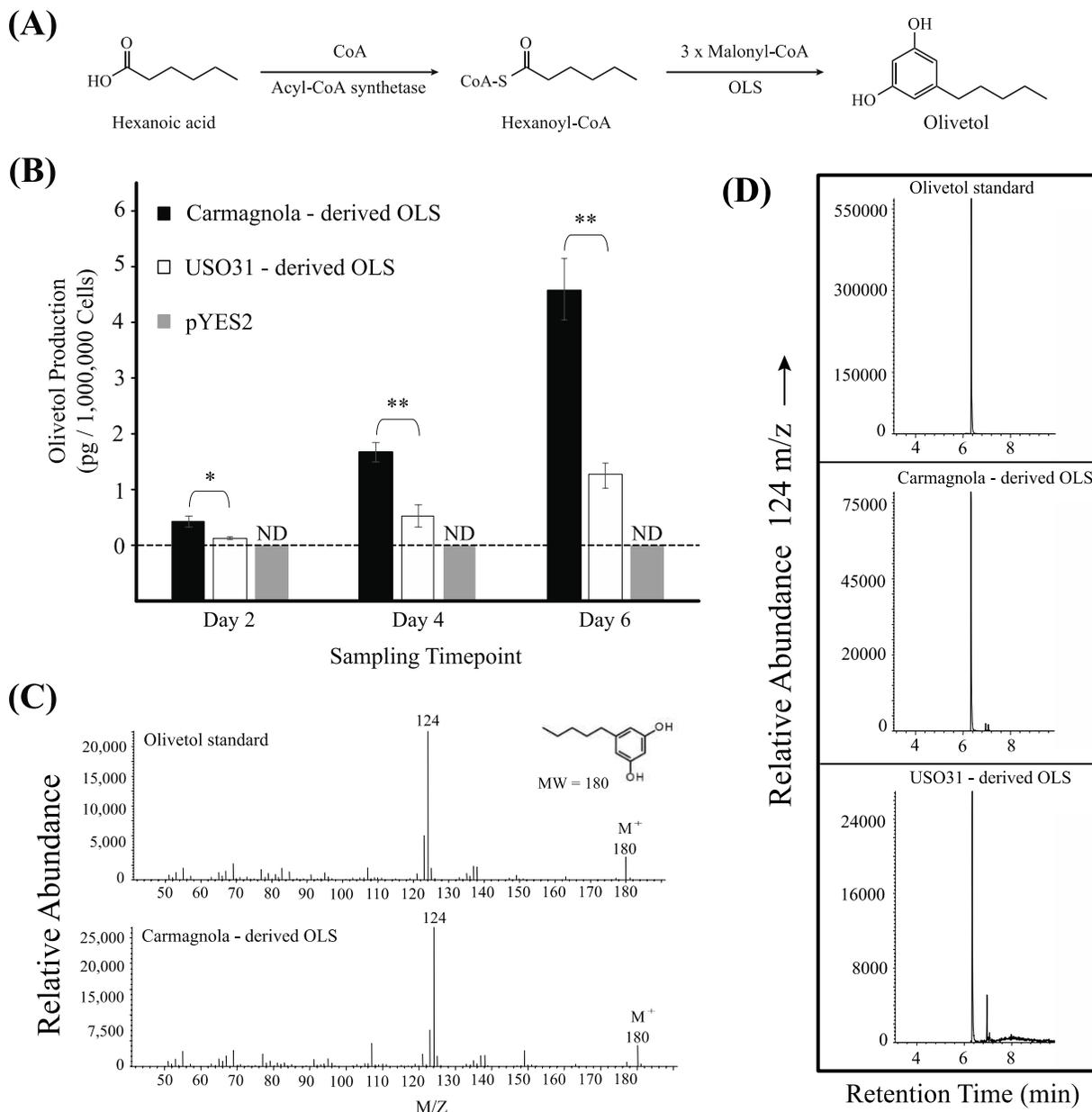


Figure 5: Olivetol production in media supplemented with 1 mM hexanoic acid. (A) Synthesis of olivetol from hexanoic acid and hexanoyl-Coenzyme A (CoA) as described by Taura et al. 2009 and Kovalchuk et al. 2020. (B) Shown are bar plots comparing the mean (\pm standard deviation) olivetol production for biological triplicates of yeast colonies expressing the Carmagnola-derived OLS (black) and USO31-derived OLS (white). Yeast colonies transformed with empty pYES2 vector is shown in grey. Single asterisks indicate a t-test p-value less than 0.05 while double asterisks indicate a t-test p-value less than 0.01. The dashed horizontal line is used to indicate no detectible (ND) quantity of olivetol for colonies transformed with empty pYES2 vectors. (C) Mass

spectra showing the major diagnostic ionization fragment and molecular ion (M^+) for olivetol for the standard (top) and Carmagnola-derived OLS (bottom). (D) Chromatograms showing the retention time (minutes) for the peak containing the major diagnostic ionization fragment for olivetol for the olivetol standard, Carmagnola-derived OLS, and USO31-derived OLS.

Discussion

Using hemp as a system to study fundamental questions regarding the genetics of complex traits, our results are relevant to both theoretical and applied quantitative genetics. Theory predicts that variation of complex traits is often attributed to many loci across the genome that can act in an additive, dominant, or epistatic manner (Lynch and Walsh, 1998). In the context of *C. sativa*, for which genetic studies have largely focused on understanding the inheritance of cannabinoids, the literature has only demonstrated instances of dominance and additivity associated with these traits' variation (de Meijer et al. 2003, Weiblen et al. 2015, Campbell et al. 2020, Petit et al. 2020a, b, and Grassa et al. 2021). Our study builds upon the current literature by suggesting that while epistasis is also a prevalent genetic factor for both cannabinoids and terpenes, epistatic interactions explain considerably less phenotypic variance (range = 3.89% - 15.11%, Table 2) than additive genetic effects. Thus, while our results suggest that the effect sizes of epistatic interactions are low for biochemical traits, they explain a significant amount of trait variation. Identification of specific QTL has been applied for genetic improvement in numerous crop species such as *Maize* and rice (Yousef et al. 2002, Luo et al. 2014 and Kumar et al. 2017). In *C. sativa*, traits of importance such as THCa and CBDa can be improved by identifying QTL that predict trait values (Toth et al. 2020 and Wenger et al. 2020). The numerous instances of epistasis identified here suggest that breeding for enhanced biochemical phenotypes in *C. sativa* may require a more complex selection process in order to take advantage of these epistatic interactions (Holland, 2001). Overall, our results demonstrate how classical quantitative genetics approaches can be used to better understand

complex trait genetic architecture in a non-model species and furthermore they provide evidence for the causal genetics that potentially underlie the numerous trait differences between two industrial hemp cultivars.

Phenotype correlations

The presence of co-located QTL suggests trait correlations (Luo et al. 2017). While many of the QTL identified in this study did co-localize, this was only observed for QTL associated with traits of the same class type (agronomic or biochemical, Figure 2 A and B). Indeed, traits of the same class type exhibited significant correlations ($p < 0.05$) that reflect their co-located QTL (Figure 1). The absence of co-located QTL and significant correlation between trait classes however suggests that the loci associated with variation in agronomic traits are largely independent of loci associated with variation in biochemical traits. Whether or not the lack of association between agronomic and biochemical traits is population-specific is currently not possible to determine since no other whole genome mapping population data exist for *C. sativa*.

Quantitative genetic architecture of C. sativa agronomic traits

Our mapping efforts identified a total of 34 QTL associated with measurable variation among the eight measured agronomic traits. Although most agronomic QTL co-localized into four distinct clusters (Figure 2 A), single QTL were identified for seed yield, thousand seed mass and dry biomass. Overall, the observed QTL clustering patterns suggest that much of the agronomic differences between Carmagnola and USO31 are controlled by a few pleiotropic large-effect loci. We note however that the summation of most agronomic trait's total variance explained in Table 1 are ~ 50% which may indicate the existence of additional small effect QTL that our experiment did not have the power to identify. Interestingly, each agronomic trait model contained a QTL

located within LG3.60 and LG5.05. This suggests that the function of the genes linked with LG3.60 and LG5.05 may control overall plant growth.

Using the narrow 1.5 LOD interval of SB.3 within LG5.05, we identified a predicted gene with high sequence homology to *TINY*, a dehydrin response element which has been shown to affect overall plant growth in *Arabidopsis thaliana* (Wilson et al. 1996 and Xie et al. 2019). After further analyzing variation in this gene's sequence, we identified a single non-synonymous homozygous SNP within the Carmagnola and USO31 *TINY* coding region (Figure 4 A) for which the F₂ genotypic classes were significantly different for all agronomic traits (Supplementary Table 6, and Supplementary Figure 5). While it is possible that stronger genotype-phenotype associations may exist in regulatory regions, our limited ability to annotate the LOD interval of SB.3 only allowed us to identify the coding sequence of *TINY*. Thus, we hypothesize that the single non-synonymous SNP in the *TINY* coding sequence underlies QTL at LG5.05 across all eight agronomic traits. Future studies will need to utilize functional genetics to validate the phenotypic effect of the single non-synonymous SNP identified between Carmagnola and USO31.

Quantitative genetic architecture of C. sativa biochemical traits

Our mapping efforts identified a total of 35 QTL associated with measurable variation among the 17 measured biochemical traits. While our model for alpha-pinene contained 6 QTL that spanned our linkage map, most other trait models contained QTL that co-localized on linkage groups 6 and 9. Similar to agronomic traits, the extensive clustering patterns of biochemical QTL at LG6.35 and LG9.40 suggest that most of the cannabinoid and terpene trait differences between the Carmagnola and USO31 parents are controlled by a small number of regions of the *C. sativa* genome. Again however, we note the power limitation of our experimental design to identify additional small effect QTL which could explain the residual variance of biochemical traits (Table

2). Interestingly, an epistatic interaction between QTL within LG6.35 and LG9.40. was identified across models of three cannabinoids and five terpenes (Table 2, Figure 3). Our ability to detect these numerous instances of epistasis across biochemical traits is reflective of their greater heritability compared to agronomic traits in *C. sativa* (Campbell et al. 2019).

The common epistatic interaction identified for numerous biochemical trait models suggests two possible hypotheses for the genes linked to the QTL within LG6.35 and LG9.40. First, we hypothesize that these instances of epistasis may indicate the locations of genes that synthesize precursor molecules to all biochemical traits measured. Alternatively, we hypothesize that these shared epistatic interactions may suggest the presence of genes involved in the interacting biosynthesis pathways for terpenes and cannabinoids (Booth et al. 2019). In general, terpenes are synthesized through either the plastidial methylerythritol phosphate (MEP) or the cytosolic mevalonate (MEV) pathways while cannabinoids are synthesized through the polyketide pathway (Kovalchuk et al. 2020). The MEP, MEV, and polyketide pathways all utilize geranyl pyrophosphate (GPP) as a substrate to produce their downstream compounds (Szkopińska & Płochocka 2005, Chizzola 2013, Booth et al. 2017, Booth et al. 2019, and Kovalchuk et al. 2020). Evidence also supports that the MEP pathway synthesizes both classes of terpenes in the glandular trichomes and flower tissue (McCaskill 1995, Dudareva et al. 2005, and Wölwer-Rieck et al. 2014). The polyketide pathway, via CBGa synthase (CBGAS), uses GPP derived from the MEP pathway and olivetolic acid as substrates to form CBGa which is the precursor molecule to all downstream cannabinoids (Fellermeier et al. 2001, Gagne et al. 2012 and Kovalchuk et al. 2020).

Using published sequences of genes involved in the MEP, MEV, and polyketide synthesis pathways, we identified olivetol synthase (OLS, Taura et al. 2009) as the candidate gene underlying LG9.40 (Figure 4 B). We hypothesize that OLS is the gene underlying LG9.40 because

of its critical step in the cannabinoid synthesis pathway and the interaction between OLS and the MEP synthesis pathway. OLS acts in conjunction with olivetolic acid cyclase to produce olivetolic acid, the compound that subsequently combines with MEP-derived GPP to form CBG_a via CBGAS (Taura et al. 2009, Gagne et al. 2012 and Kovalchuk et al. 2020). With OLS's critical involvement in the cannabinoid synthesis pathway, variation in the quantity or efficiency of OLS is likely to greatly affect production of cannabinoids (Gagne et al. 2012). We identified nine homozygous amino acid substitutions segregating within the F₂ OLS coding sequence which may be responsible for the highly contrasting quantities of cannabinoids produced between the Carmagnola and USO31 parents through alteration of their OLS enzyme function (Supplementary Table 1). Our functional assays in yeast show that the divergent OLS alleles between Carmagnola and USO31 underlie their differences in cannabinoid quantities which suggests that the USO31 OLS may be less efficient at converting hexanoyl-CoA to olivetol (Figure 5 A and B). If the USO31 OLS is less efficient, this would reduce the quantity of cannabinoids produced by USO31 compared to Carmagnola which was a trend we observed in our field experiment (Supplementary Table 1).

Although the OLS functional assay explains why F₂ plants with the Carmagnola allele at LG9.40 produced significantly more cannabinoids, it is not clear why the genotype at this QTL cluster also causes differences in terpene production. At LG9.40 F₂ plants possessing the Carmagnola allele generally produced more of each terpene compared to individuals with the USO31 allele (Supplementary Table 5). Additional whole genome data from *C. sativa* mapping populations is needed to determine whether or not this phenomenon for cannabinoid and terpenes at LG9.40 is population specific. While general schematics of the MEP, MEV, and polyketide pathways have been described in *C. sativa* (Fellermeier et al. 2001, Taura et al. 2009, Gagne et al.

2012, Booth et al. 2017, Booth et al. 2019, and Kovalchuk et al. 2020), much still remains uncertain about the nuances of their exact mechanisms. These data at LG9.40 suggest two possible hypotheses. First, we hypothesize that these data may indicate complex interactions between cannabinoid and terpene biosynthesis pathways which have not been previously described. Alternatively, we hypothesize that these data may reflect a pleiotropic regulatory mechanism controlling both cannabinoid and terpene biosynthesis pathways in *C. sativa* as suggested by Zager et al. 2019. Thus, the observed patterns of terpene production at LG9.40 necessitate more investigation into the genetic mechanisms of the cannabinoid and terpene biosynthesis pathways.

In conclusion, understanding the genetics of complex traits remains a formidable challenge. This is because complex traits can vary considerably ranging from traits controlled by a few QTL of large effect to other traits controlled by a number of loci of small effect. Other factors such as epistasis and dominance add additional complications which further inhibit our ability to fully understand the genetics of complex traits. However, in the past decade, strides have been made in model species by combining WGS with bi-parental mapping populations to identify numerous QTL associated with variation in complex traits such as drug resistance in *Caenorhabditis elegans* (Burga et al. 2019), ethanol tolerance in yeast (Swinnen et al. 2012), water use physiology in *Arabidopsis thaliana* (Mojica et al. 2016), and nitrogen use efficiency in rice (Yang et al. 2017). In less characterized species, genetic understanding of complex traits lags far behind since these species lack many of the well-established genetic resources available to model species. Therefore, it is necessary that for less characterized species such as industrial hemp, investigations of fundamental questions regarding the genetics of complex traits are conducted because they provide the foundations for understanding these species genetic architectures. Despite the prominence of hemp and its numerous uses in society, the genetics of agriculturally important traits in hemp have

been seldom investigated. To evaluate the genetic architecture of complex traits in hemp, we used a classical quantitative genetics approach paired with WGS for high resolution mapping of QTL and heterologous expression in yeast to functionally validate a candidate gene. Rather than adhering to the additive model whereby traits are controlled by numerous loci of small effect (Fisher, 1919), we show that the phenotypic distinctions between Carmagnola and USO31 are attributed to a small number of loci of relatively large effect. While additional steps remain necessary to: 1) validate the parental alleles for *TINY* and 2) resolve the mechanisms of the cannabinoid and terpene biosynthesis pathways, the results discussed here demonstrate the exploration of fundamental complex trait questions in a non-model species, improving upon the current understanding of the genetics controlling agriculturally important traits in hemp.

Data availability statement

The F₂ linkage map and phenotype data used for all analyses (QTL mapping, phenotype correlations, etc.) have been made available on the GSA figshare portal. File_S1 contains the raw F₂ phenotype data. File_S2 contains the quantile normalized F₂ phenotype data used for mapping. File_S3 contains the F₂ linkage map and genotype information. Yeast strains and plasmids are available upon request. Raw fastq files have been deposited to NCBI's short read archive under BioProject Accession number: PRJNA723060.

REFERENCES

- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- BLAST. (2009). Nucleotide BLAST: Search nucleotide databases using a nucleotide query.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120
- Booth, J. K., & Bohlmann, J. (2019). Terpenes in *Cannabis sativa* – From plant genome to humans. *Plant Science*. <https://doi.org/10.1016/j.plantsci.2019.03.022>
- Booth, J. K., Page, J. E., & Bohlmann, J. (2017). Terpene synthases from *Cannabis sativa*. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0173911>
- Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btg112>
- Buchanan, C. J., Webb, A. L., Mutschall, S. K., Kruczkiewicz, P., Barker, D. O. R., Hetman, B. M., ... Taboada, E. N. (2017). A genome-wide association study to identify diagnostic markers for human pathogenic campylobacter jejuni strains. *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2017.01224>
- Burga, A., Ben-David, E., Lemus Vergara, T., Boocock, J., & Kruglyak, L. (2019). Fast genetic mapping of complex traits in *C. elegans* using millions of individuals in bulk. *Nature Communications*. <https://doi.org/10.1038/s41467-019-10636-9>
- Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype × Environment Interactions of Industrial Hemp Cultivars Highlight Diverse Responses to Environmental Factors. *Age*. <https://doi.org/10.2134/age2018.11.0057>
- Campbell, L. G., Dufresne, J., & Sabatinos, S. A. (2020). Cannabinoid Inheritance Relies on Complex Genetic Architecture. *Cannabis and Cannabinoid Research*. <https://doi.org/10.1089/can.2018.0015>
- Chizzola, R. (2013). Regular monoterpenes and sesquiterpenes (essential oils). In *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*. https://doi.org/10.1007/978-3-642-22144-6_130
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr330>

- De Meijer, E. P. M., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. M. C., Ranalli, P., & Mandolino, G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*. <https://doi.org/10.1093/genetics/163.1.335>
- Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., ... Gershenzon, J. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0407360102>
- Fellermeier, M., Eisenreich, W., Bacher, A., & Zenk, M. H. (2001). Biosynthesis of cannabinoids. *European Journal of Biochemistry*. <https://doi.org/10.1046/j.1432-1327.2001.02030.x>
- Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*. <https://doi.org/10.1017/S0080456800012163>
- Gagne, S. J., Stout, J. M., Liu, E., Boubakir, Z., Clark, S. M., & Page, J. E. (2012). Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1200330109>
- Gallily, R., Yekhtin, Z., & Hanuš, L. O. (2015). Overcoming the Bell-Shaped Dose-Response of Cannabidiol by Using Cannabis Extract Enriched in Cannabidiol. *Pharmacology & Pharmacy*. <https://doi.org/10.4236/pp.2015.62010>
- Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., ... Schwartz, C. J. (2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist*. <https://doi.org/10.1111/nph.17243>
- Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*. <https://doi.org/10.1098/rstb.2009.0203>
- Holland, J. B. (2001). Epistasis and Plant Breeding. In *Plant Breeding Reviews* (pp. 27–92). <https://doi.org/10.1002/9780470650196.ch2>
- J.T. Lovell (2019). qtlTools: Data Processing and Plotting in Association with R/qtl. R package version 1.2.0.
- Kassambara A. (2020). Rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.6.0. <https://CRAN.R-project.org/package=rstatix>

- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts199>
- Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., R. Ashnest, J., ... Sharbel, T. F. (2020). The Genomics of Cannabis and Its Close Relatives. *Annual Review of Plant Biology*. <https://doi.org/10.1146/annurev-arplant-081519-040203>
- Kumar, J., Gupta, D. Sen, Gupta, S., Dubey, S., Gupta, P., & Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*. <https://doi.org/10.1007/s00299-017-2127-y>
- Laverty, Kaitlin U., et al. "A Physical and Genetic Map of Cannabis Sativa Identifies Extensive Rearrangements at the THC/CBD Acid Synthase Loci." *Genome Research*, vol. 29, no. 1, 2018, pp. 146–156., doi:10.1101/gr.242594.118.
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 0: 1–3
- 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
- Luo, H., Ren, X., Li, Z., Xu, Z., Li, X., Huang, L., ... Jiang, H. (2017). Co-localization of major quantitative trait loci for pod size and weight to a 3.7 cM interval on chromosome A05 in cultivated peanut (*Arachis hypogaea* L.). *BMC Genomics*. <https://doi.org/10.1186/s12864-016-3456-x>
- Luo, Y., Zakaria, S., Basyah, B., Ma, T., Li, Z., Yang, J., & Yin, Z. (2014). Marker-assisted breeding of Indonesia local rice variety Siputeh for semi-dwarf phenotype, good grain quality and disease resistance to bacterial blight. *Rice*. <https://doi.org/10.1186/s12284-014-0033-2>
- Lynch M. and Walsh B. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, 1998.
- Lynch, R. C., Vergara, D., Tittes, S., White, K., Schwartz, C. J., Gibbs, M. J., ... Kane, N. C. (2016). Genomic and Chemical Diversity in Cannabis. *Critical Reviews in Plant Sciences*. <https://doi.org/10.1080/07352689.2016.1265363>
- MacKay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg2612>
- McCaskill, D., & Croteau, R. (1995). Monoterpene and sesquiterpene biosynthesis in glandular trichomes of peppermint (*Mentha x piperita*) rely exclusively on plastid-derived isopentenyl diphosphate. *Planta*. <https://doi.org/10.1007/BF00239938>

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. <https://doi.org/10.1101/gr.107524.110>
- Mehmedic, Z., Chandra, S., Slade, D., Denham, H., Foster, S., Patel, A. S., ... ElSohly, M. A. (2010). Potency trends of Δ^9 -THC and other cannabinoids in confiscated cannabis preparations from 1993 to 2008. *Journal of Forensic Sciences*. <https://doi.org/10.1111/j.1556-4029.2010.01441.x>
- Mojica, J. P., Mullen, J., Lovell, J. T., Monroe, J. G., Paul, J. R., Oakley, C. G., & McKay, J. K. (2016). Genetics of water use physiology in locally adapted *Arabidopsis thaliana*. *Plant Science*. <https://doi.org/10.1016/j.plantsci.2016.03.015>
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btw044>
- Petit, J., Salentijn, E. M. J., Paulo, M. J., Denneboom, C., & Trindade, L. M. (2020, ab). Genetic Architecture of Flowering Time and Sex Determination in Hemp (*Cannabis sativa* L.): A Genome-Wide Association Study. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2020.569958>
- Petit, J., Salentijn, E. M. J., Paulo, M. J., Denneboom, C., van Loo, E. N., & Trindade, L. M. (2020, ba). Elucidating the Genetic Architecture of Fiber Quality in Hemp (*Cannabis sativa* L.) Using a Genome-Wide Association Study. *Frontiers in Genetics*. <https://doi.org/10.3389/fgene.2020.566314>
- Petit, J., Salentijn, E. M. J., Paulo, M. J., Thouminot, C., van Dinter, B. J., Magagnini, G., ... Trindade, L. M. (2020, c). Genetic Variability of Morphological, Flowering, and Biomass Quality Traits in Hemp (*Cannabis sativa* L.). *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2020.00102>
- Russo, E. B. (2011). Taming THC: Potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *British Journal of Pharmacology*. <https://doi.org/10.1111/j.1476-5381.2011.01238.x>
- Salentijn, E. M. J., Zhang, Q., Amaducci, S., Yang, M., & Trindade, L. M. (2015). New developments in fiber hemp (*Cannabis sativa* L.) breeding. *Industrial Crops and Products*. <https://doi.org/10.1016/j.indcrop.2014.08.011>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btv351>

- Singh, U., Deb, R., Alyethodi, R. R., Alex, R., Kumar, S., Chakraborty, S., ... Sharma, A. (2014). Molecular markers and their applications in cattle genetic research: A review. *Biomarkers and Genomic Medicine*. <https://doi.org/10.1016/j.bgm.2014.03.001>
- Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., & Taura, F. (2004). The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ^1 -tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M403693200>
- Sirikantaramas, S., Taura, F., Tanaka, Y., Ishikawa, Y., Morimoto, S., & Shoyama, Y. (2005). Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology*. <https://doi.org/10.1093/pcp/pci166>
- Speed, D., & Balding, D. J. (2012). Understanding complex traits: from farmers to pharma. *Genome Medicine*. <https://doi.org/10.1186/gm360>
- Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh379>
- Struik, P. C., Amaducci, S., Bullard, M. J., Stutterheim, N. C., Venturi, G., & Cromack, H. T. H. (2000). Agronomy of fibre hemp (*Cannabis sativa* L.) in Europe. *Industrial Crops and Products*. [https://doi.org/10.1016/S0926-6690\(99\)00048-5](https://doi.org/10.1016/S0926-6690(99)00048-5)
- Swinnen, S., Schaerlaekens, K., Pais, T., Claesen, J., Hubmann, G., Yang, Y., ... Thevelein, J. M. (2012). Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome Research*. <https://doi.org/10.1101/gr.131698.111>
- Szkopińska, A., & Płochocka, D. (2005). Farnesyl diphosphate synthase; regulation of product specificity. *Acta Biochimica Polonica*. https://doi.org/10.18388/abp.2005_3485
- Tang, K., Struik, P. C., Yin, X., Calzolari, D., Musio, S., Thouminot, C., ... 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*. <https://doi.org/10.1016/j.indcrop.2017.06.033>
- Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y., & Morimoto, S. (2009). Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Letters*. <https://doi.org/10.1016/j.febslet.2009.05.024>
- Team, R. C. (2019). R: A Language and Environment for Statistical Computing. *Vienna, Austria*.

- Toth, J. A., Stack, G. M., Cala, A. R., Carlson, C. H., Wilk, R. L., Crawford, J. L., ... Smart, L. B. (2020). Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy*. <https://doi.org/10.1111/gcbb.12667>
- 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., & WITHAGEN, J. C. M. (1995, a). The effect of temperature on leaf appearance and canopy establishment in fibre hemp (*Cannabis sativa* L.). *Annals of Applied Biology*. <https://doi.org/10.1111/j.1744-7348.1995.tb05389.x>
- Van der Werf, H. M. G., van Geel, W. C. A., van Gils, L. J. C., & Haverkort, A. J. (1995, b). Nitrogen fertilization and row width affect self-thinning and productivity of fibre hemp (*Cannabis sativa* L.). *Field Crops Research*. [https://doi.org/10.1016/0378-4290\(95\)00017-K](https://doi.org/10.1016/0378-4290(95)00017-K)
- Van Ooijen, J. W. (2006). JoinMap® 4. *JoinMap*.
- Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., & Marks, M. D. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytologist*. <https://doi.org/10.1111/nph.13562>
- Wenger, J. P., Dabney, C. J., ElSohly, M. A., Chandra, S., Radwan, M. M., Majumdar, C. G., & Weiblen, G. D. (2020). Validating a predictive model of cannabinoid inheritance with feral, clinical, and industrial *Cannabis sativa*. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1550>
- Wickham, H. (2016). Ggplot2 Elegant Graphics for Data Analysis. In *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <https://doi.org/10.1007/978-3-319-24277-4>
- Wilson, K., Long, D., Swinburne, J., & Coupland, G. (1996). A Dissociation insertion causes a semidominant mutation that increases expression of TINY, an 45rabidopsis gene related to APETALA2. *Plant Cell*. <https://doi.org/10.2307/3870342>
- Wölwer-Rieck, U., May, B., Lankes, C., & Wüst, M. (2014). Methylerythritol and mevalonate pathway contributions to biosynthesis of mono-, sesqui-, and diterpenes in glandular trichomes and leaves of *Stevia rebaudiana bertonii*. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/jf500270s>
- Xie, Z., Nolan, T., Jiang, H., Tang, B., Zhang, M., Li, Z., & Yin, Y. (2019). The AP2/ERF transcription factor TINY modulates brassinosteroid-regulated plant growth and drought responses in 45rabidopsis. *Plant Cell*. <https://doi.org/10.1105/tpc.18.00918>

- Yang, X., Xia, X., Zhang, Z., Nong, B., Zeng, Y., Xiong, F., ... Li, D. (2017). QTL mapping by whole genome re-sequencing and analysis of candidate genes for Nitrogen use efficiency in rice. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2017.01634>
- Yousef, G. G., & Juvik, J. A. (2002). Enhancement of Seedling Emergence in Sweet Corn by Marker-Assisted Backcrossing of Beneficial QTL. *Crop Science*. <https://doi.org/10.2135/cropsci2002.0096>
- Zager, J. J., Lange, I., Srividya, N., Smith, A., & Markus Lange, B. (2019). Gene networks underlying cannabinoid and terpenoid accumulation in cannabis. *Plant Physiology*. <https://doi.org/10.1104/pp.18.01506>
- Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L. *Journal of Biotechnology*. <https://doi.org/10.1016/j.jbiotec.2018.07.031>

CHAPTER 3

GENOME WIDE POLYMORPHISM AND GENIC SELECTION IN FERAL AND DOMESTICATED LINEAGES OF *CANNABIS SATIVA*

Summary

A comprehensive understanding of the degree to which genomic variation is maintained by selection versus drift and gene flow is lacking in many important species such as *Cannabis sativa* (*C. sativa*), one of the oldest known crops to be cultivated by humans worldwide. We generated whole genome resequencing data across diverse samples of feralized (escaped domesticated lineages) and domesticated lineages of *C. sativa*. We performed analyses to examine population structure, and genome wide scans for F_{ST} , balancing selection, and positive selection. Our analyses identified evidence for sub-population structure and further support the Asian origin hypothesis of this species. Feral plants sourced from the U.S. exhibited broad regions on chromosomes 4 and 10 with high \bar{F}_{ST} which may indicate chromosomal inversions maintained at high frequency in this sub-population. Both our balancing and positive selection analyses identified loci that may reflect differential selection for traits favored by natural selection and artificial selection in feral versus domesticated sub-populations. In the U.S. feral sub-population, we found six loci related to stress response under balancing selection and one gene involved in disease resistance under positive selection, suggesting local adaptation to new climates and biotic interactions. In the marijuana sub-population, we identified the gene *SMALLER TRICHOMES WITH VARIABLE BRANCHES 2* to be under positive selection which suggests artificial selection for increased tetrahydrocannabinol yield. Overall, the data generated, and results obtained from our study help to form a better understanding of the evolutionary history in *C. sativa*.

Introduction

A central goal of evolutionary genetics is to understand the relative roles of recombination, selection, and drift in maintaining genetic variation in populations (Charlesworth & Charlesworth 2012). Since Charles Darwin first described the process of natural selection (Darwin 1859), identifying the differences of how evolution proceeds under natural versus artificial selection (domestication) has long intrigued scientists. Domestication of a species is associated with a “domestication syndrome” whereby adaptive genetic variation that enhances survival traits in nature is often selected against to favor genetics that enhance traits for human use (Brown 2010, Larson et al. 2014). Resequencing technology has been invaluable for understanding how the domestication syndrome affects genomic variation in crops by enabling scientists to compare entire genomes between domesticated lineages and their wild progenitors which has revealed beneficial and deleterious mutations (Mace et al. 2013, Qin et al. 2014). Less understood is how genomic variation is affected when domesticated crop lineages escape farms to become feralized and revert to evolving under natural selection (Scossa and Fernie 2021). Species with both domesticated varieties and feralized crop populations are ideal systems to study these outstanding questions: which genes underlie adaptation under feralization versus domestication and does adaptation under feralization versus domestication occur through fixation of standing variation or through new mutations (Scossa and Fernie 2021)?

Cannabis sativa is one of the oldest known crops to be cultivated by humans for fiber, grain, and secondary metabolites such as terpenes and cannabinoids (Laursen 2015). From a scientific standpoint, *C. sativa* is interesting because it evolved dioecy and an annual life history from ancestors that were monoecious and perennial (Kovalchuk et al. 2020). As a result of extensive human vectoring and hybridization, the evolutionary history of this species remains

convoluted. This domesticated species moved around with humans for thousands of years, but then became prohibited in much of the world in 1937 due to the Marijuana Tax Act. Despite its prohibition, in the U.S., it is estimated that 400 thousand acres of hemp biomass were produced for textile and feedstock between 1942 and 1945 (West 1998). This created the current feralized populations in the U.S., that have been evolving under natural selection for decades (West 1998). As time has progressed, global regulations surrounding the legality of *C. sativa* have increasingly relaxed which enables use of this species to better understand the evolutionary genetics of crop feralization.

Recent progress has been made in understanding the genomic variation of *C. sativa* and has identified several areas remain that require additional investigation (Kovalchuk et al. 2020 and Ren et al. 2021). These areas include: 1) characterization of the genomic variation in broader samples such as those maintained by the Institute of Plant Genetics (IPK), 2) formal testing of the out of Asia hypothesis, and 3) comparison of balancing and positive selection patterns between domesticated and feralized sub-populations to better understand the genetics underlying their domestication/loss of domestication syndrome and local adaptation. Addressing these three areas is necessary for a more comprehensive characterization of this species' evolutionary history.

In this study, the genomic variation of ~80 previously unexamined *C. sativa* accessions, including industrial hemp varieties, feral populations in the U.S., and accessions from the IPK Gatersleben seed bank were characterized using WGS. Several publicly available WGS datasets for additional accessions data were also incorporated into our analyses for a more comprehensive study. Our study identified results consistent with previous reports while also expanding upon the current knowledge of *C. sativa* genomics with new discoveries. Through our analyses, we identify additional evidence for sub-population structure, further support for the Asian origin of *C. sativa*

and find evidence for both balancing and positive selection on loci that potentially reflect the domestication/loss of domestication syndrome and local adaptation of cultivated and feral sub-populations.

Materials and Methods

Sample collection

Seed representing 20 industrial hemp cultivars were imported to Colorado for a set of field trials (Campbell et al. 2019). Additional seed representing 51 *C. sativa* accessions stored at the IPK Gatersleben germplasm bank were imported from Germany to Colorado. Four seeds from each accession were sown in 3” x 3” pots in the Colorado State University greenhouse, and leaf samples were harvested for DNA extraction once the plants were well into their vegetative stage.

We also collected leaf tissue from feral *C. sativa* population located in Merino and Idalia, Colorado. All leaf tissue was freeze dried using a lyophilizer (Labconco, Kansas City, MO) prior to DNA extraction.

Growth conditions and THC quantity measurements

Four seeds of each accession were planted in 3” x 3” pots in wet ProMix soil. The pots were kept damp through the germination window and then plants were watered as needed. General Hydroponics (FloraGro, FloraMicro and FloraBloom) fertilizer was applied 2-3 times in the first 2 months of growth to encourage vegetative growth. No further fertilizing events occurred. Flower tissue samples were harvested from 41 IPK accessions once the plants reached maturity and were oven dried for two weeks. No flower tissue samples were harvested from the 20 industrial hemp cultivars.

To assess the quantity of the cannabinoid tetrahydrocannabinol (THC), within harvested flower tissue, each plant’s dried flower biomass was weighed out into one-gram aliquots and stored

in separate 50mL conical centrifuge tubes. These one-gram flower aliquots were then delivered to Botanacor Laboratories in Denver, Colorado for standard THC quantity analysis (<https://www.botanacor.com>). In short, Botanacor Laboratories assesses flower samples for THC quantity using Agilent HPLC-DAD instruments. We used previously published data on THC percent by mass for 30 accessions from Campbell et al. 2019 and Vergara et al. 2019. In total, we used THC percent by mass for 71 *C. sativa* accessions for which we also had WGS data.

DNA extraction and whole genome shotgun sequencing

DNA was extracted from plant leaf tissue using the Qiagen DNeasy Plant Mini Kit (Valencia, CA). Extracted DNA was then quantified using a Qubit Fluorometer (ThermoFisher Scientific). Leaf tissue from the 20 European industrial hemp accessions was paired-end whole genome sequenced (2x 150 base pairs) at the Duke University Center for Genomic and Computational Biology using an Illumina HiSeq 4000. Leaf tissue from the 51 IPK accessions was paired-end whole genome sequenced (2x 150 base pairs) at the University of Colorado Anschutz Medical Campus using an Illumina NovaSeq. Sequencing efforts aimed for 4x to 7x coverage of the *C. sativa* genome.

Genotyping of samples

In addition to our WGS data, 55 publicly available *C. sativa* WGS datasets representing marijuana, industrial hemp and feral *C. sativa* accessions were obtained from Lynch et al. 2016. Raw sequence data was initially parsed with FastQC (Andrews S. 2010, version 0.11.8) to assess read quality and adapter contamination. Trimmomatic (Bolger et al. 2014, version 0.39) was then used with default parameters to remove low quality reads and any adapter contamination identified in the FastQC report. The trimmed sequence reads were then aligned to version 2 of the CS10 reference genome (Grassa et al. 2021, GenBank assembly accession ID = GCF_900626175.2)

using BWA-MEM with default settings (Li 2010, version 0.7.17). Samtools (Li et al. 2009, version 1.9) was then used to sort sequence alignment files, mark duplicate reads, keep only properly paired reads, and remove reads with a mapping quality less than 10. BCFtools (Narasimhan et al. 2016, version 1.9) was then used with default parameters to identify genetic variants using both the “mpileup” and “call” functions to produce a single variant call file (VCF) for all samples. In total we genotyped 190 samples for our analyses.

VCFtools (Danecek et al. 2011, version 0.1.16) was then used to filter the VCF to contain bi-allelic single nucleotide polymorphisms (SNPs) that possessed a genotyping rate of $\geq 80\%$, quality ≥ 30 , minimum mean read depth ≥ 2 , and minor allele frequency of $\geq 5\%$. Additionally, to reduce the impact of erroneous heterozygous genotype calls due to misalignment of paralog and repetitive element reads, we used VCFtools to remove loci in the VCF that had a heterozygous genotype call rate greater than 50% across the 190 samples. The final VCF contained 8,474,449 SNPs.

Statistical analyses

Population structure, cannabinoid correlations, segregating sites, and Tajima's D

A genome-wide principal component analysis (PCA) was conducted using the R (R Core Team, 2019, version 3.6.0) package “SNPRelate” (Zheng et al. 2012, version 1.18.0). Genetic clustering of samples observed from plotting principal component 1 (PC1) against principal component 2 (PC2) were used to identify sub-populations for downstream analyses. *C. sativa* samples with known source countries were assigned population names based on geographic regions that broadly covered their source and surrounding countries. Thus, we designated “European” (European neighborhood countries), “Asian” (China and North Korea), “U.S. feral” (feral accessions from Colorado, Kansas, and Nebraska) and “South American” (Argentina) as our

geographic populations. Marijuana samples (as described in Lynch et al. 2016) were classified as “Marijuana” because no distinct clustering based on geography was observed for these samples. Since many of the IPK accessions were not labeled with accurate collection site information, these samples were classified as “IPK unknown”.

To further examine population structure, we used ADMIXTURE (Alexander et al. 2009, version 1.3.0) to estimate individual genetic ancestries. The appropriate number of genetic clusters (K) was determined using the cross-validation option for K values 1-10. The K value with the lowest cross validation error was then used for subsequent analyses. After identifying the appropriate K, the output “.Q” file from ADMIXTURE was input into the R package “pophelper” (Francis, R. M. 2017, version 2.2.9) for visualization of individual ancestry.

Genetic divergence between sub-populations was estimated using VCFtools to calculate the mean Weir and Cockerham fixation index (\bar{F}_{ST}) using both a pairwise and sliding window approach with 20,000 base pair window sizes (Weir and Cockerham, 1984). Sliding window \bar{F}_{ST} was visualized using the R package “qqman” (Turner 2017, version 0.1.4).

Means of total potential THC quantities were calculated for each accession with available data using the R package “emmeans” (Lenth et al. 2018, version 1.3.5.1). Correlations between accession mean total potential THC quantity and PC1 were then tested using the base R function `cor.test()` to obtain the Spearman’s rank correlation coefficient (ρ).

Under the out of Asia hypothesis, we would predict deeper coalescence events and thus a greater proportion of segregating sites in the Asian sub-population. To test the hypothesis that Asian accessions represent the ancestral *C. sativa* population, we performed an analysis to compare the proportion of segregating sites across single copy orthologs (SCOs) between the Asian, European, U.S. feral and Marijuana sub-populations. To accomplish this, we first identified 7,540

SCOs between *Ziziphus jujube* (Huang et al. 2016), *Parasponia andersonii* (van Velzen et al. 2018) and *C. sativa* using OrthoFinder (Emms and Kelly, 2019, version 2.2.6) with default parameters. Next, because our Asian population contained 8 samples, we randomly selected 8 samples from the European, U.S. feral and Marijuana sub-populations to use for comparative analyses of segregating sites. Using the SCO gene coordinates contained in the CS10 gff file, we extracted the VCF loci within the SCO gene intervals for each sub-population. The R package “PopGenome” (Pfeifer et al. 2014, version 2.7.5) was then used to count the number of segregating sites within each SCO interval per sub-population. A one-way ANOVA was then performed to compare the mean proportion of segregating sites (per 1kb) between sub-populations. Tukey-adjusted pairwise comparisons were then used to identify which sub-populations exhibited significantly different proportions of segregating sites. To estimate sub-population demography, Tajima’s D (Tajima 1989) was also calculated for each SCO interval using the R package “PopGenome”. A one-way ANOVA was then performed to compare the mean Tajima’s D between sub-populations. Tukey-adjusted pairwise comparisons were then used to identify which sub-populations exhibited significantly different mean Tajima’s D values.

IPK samples without source country information and the two South American samples were not included for statistical analyses beyond our PCA, ADMIXTURE, cannabinoid correlations. Asian samples were not included for statistical analyses beyond PCA, ADMIXTURE, THC correlation, proportion of segregating sites, and mean Tajima’s D analyses due to small sub-population sample size.

Identification of loci under balancing selection

To identify loci under balancing selection we first used the R package “PopGenome” to calculate Tajima’s D for each SCO interval using all samples within the U.S. feral, European and

marijuana sub-populations separately. We then extracted the loci within SCO interval that possessed a Tajima's D value greater than or equal to the 99th percentile Tajima's D of all SCO intervals within each sub-population. Counts of site heterozygosity per locus were then calculated and averaged over each SCO interval. Last, SCO intervals that possessed a mean site heterozygosity count greater than or equal to the 70th percentile mean heterozygosity count of all SCO intervals included in the analysis were considered as genes under balancing selection. Genes under balancing selection were then investigated for their annotation and subsequent interpretation. Additionally, since many genes lacked detailed annotation, we used the TAIR implementation of basic local alignment search tool (BLASTP) to identify the best homolog match in *Arabidopsis thaliana*.

Identification of derived alleles under positive selection

Using the SCOs identified between *Ziziphus jujube*, *Parasponia andersonii* and *C. sativa*, we estimated the ancestral and derived allelic states within the European, U.S. feral and Marijuana sub-populations using the methods described in Price et al. 2018. For this analysis, we filtered our VCF to contain the 40 Marijuana samples, 40 randomly selected European samples and all 19 U.S. feral samples. Additionally, only loci with a 100% genotyping rate across samples were used in our analyses to avoid frequency bias due to missing genotype information. For downstream analyses of positive selection, we focused only on identifying non-synonymous derived alleles.

Three separate unfolded joint site frequency spectra were plotted to estimate derived allelic variation between U.S. feral, European and Marijuana sub-populations. Derived alleles showing evidence for local positive selection (between groups) were identified based on those with non-synonymous derived allele frequencies $\geq 70\%$ in one group and $\leq 30\%$ in the other two groups. The Fisher's exact test base R function was then used to confirm significant allele frequency-

population associations. Genes containing derived alleles under positive selection were then investigated for their annotation and subsequent interpretation. We then used the TAIR implementation of BLASTP to identify the best homolog match in *Arabidopsis thaliana*.

Derived alleles showing evidence for species-wide positive selection were identified based on those with non-synonymous derived allele frequencies greater than 70% across all three groups. To broadly describe the function of all genes under species-wide positive selection, we performed a GO analysis for biological processes, molecular function, and cellular component using the locus names for the top *Arabidopsis thaliana* homolog matches from BLAST in conjunction with the GO Term Enrichment analysis tool implemented in TAIR using the Fisher's Exact test method with a FDR p-value cutoff of 0.05.

Results

Population structure, correlations, segregating sites, and mean Tajima's D

Our genome wide PCA of the ~ 8.5 million SNPs identified evidence for sub-population structure across the 190 samples. PC1 with PC2 explained 7.82% and 4.20% of the genetic variance, respectively (Figure 6 A). Marijuana samples largely formed a single cluster while *C. sativa* from Europe, Asia, and the U.S. generally clustered based on geography (Figure 6 A). Two WGS samples sourced from South America representing one accession (IPK CAN 51) clustered with European samples while three WGS samples representing two accessions sourced from Europe (IPK CAN 37 and IPK CAN 58) clustered near Asian *C. sativa*. IPK accessions with unknown source country information mostly clustered with European and Asian samples except for IPK CAN 36 which clustered near the Marijuana group.

ADMIXTURE tests of cross validation error for K values 1-10 indicated K = 6 as the best choice of cluster value (Supplemental Figure 11). Genetic distinctions between samples that

broadly resembled their clustering patterns in the PCA were observed (Supplemental Figure 12). Marijuana samples largely made-up genetic cluster 4. European *C. sativa* were mostly split between genetic clusters 2 and 6. Genetic cluster 5 mostly consisted of U.S. feral plants. All six genetic clusters were represented within Asian *C. sativa* with the most predominant being 3 and 4. Most IPK samples of unknown origin and the two samples from South America showed evidence for belonging to one of the two clusters associated with European accessions.

Our ANOVA comparing the mean proportion of segregating sites across SCOs identified a significant difference ($p < 0.05$) between sub-populations. Post-hoc pairwise comparison tests found that each sub-population included in the analysis significantly differed ($p < 0.05$) in their proportion of segregating sites. The U.S. feral sub-population exhibited the lowest proportion of segregating sites while the Asian sub-population exhibited the highest proportion of segregating sites (Figure 6 C), consistent with a deeper coalescent time for ancestral sub-populations.

Our ANOVA comparing the mean Tajima's D across SCOs identified a significant difference ($p < 0.05$) between sub-populations as well. The mean Tajima's D value for all sub-populations were positive. Post-hoc pairwise comparison tests found that the U.S. feral and marijuana sub-populations did not significantly differ ($p > 0.05$) from each other and possessed the greatest mean Tajima's D values (Supplemental Figure 13). The U.S. feral and marijuana sub-populations did however significantly differ ($p < 0.05$) from the European and Asian sub-populations. The Asian sub-population possessed the lowest mean Tajima's D value and significantly differed ($p < 0.05$) from the European sub-population.

A significant positive correlation ($p < 2.2e-16$, $R^2 = 0.80$) was observed between % total potential THC (mg/g) and PC1 (Figure 6 B). Samples exhibiting more positive PC1 eigenvalues

(i.e. more closely related to Marijuana samples) tended to produce more THC compared to samples with more negative eigenvalues.

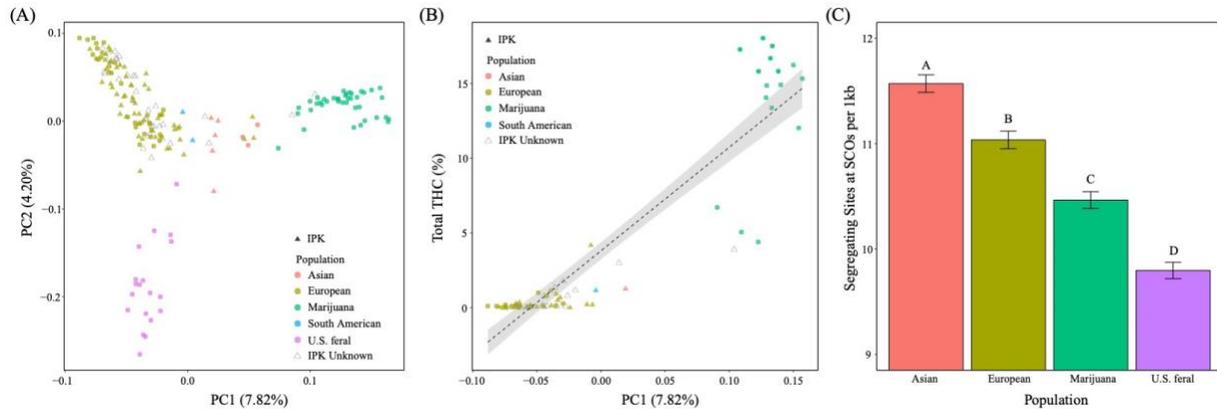


Figure 6: The population structure of *C. sativa*. (A) Genome-wide PCA showing clustering and classification of samples. Color of points indicates sub-population assignment while triangles indicate samples obtained from the Institute of Plant Genetics (IPK). (B) Correlation between PC1 and mean % total THC for 71 accessions. The black line indicates the regression slope while the grey surrounding area indicates the 95% confidence interval. (C) Bar plots showing the comparison of the mean (\pm standard error) proportion of segregating sites at single copy orthologs (SCOs). Letters above bars indicate significance of Tukey adjusted pairwise comparisons such that sub-populations with different letters possess significantly different proportions of segregating sites.

$$\bar{F}_{ST}$$

Our pairwise \bar{F}_{ST} analysis (Supplemental Table 1) indicated that overall, all the *C. sativa* sub-populations are closely related with the most divergent sub-populations being the U.S. feral and Marijuana samples ($\bar{F}_{ST} = 0.116$). Sliding window \bar{F}_{ST} analyses (Figure 7 A) revealed broad genomic regions on chromosomes 4 and 10 that exhibited numerous high \bar{F}_{ST} windows. Plotting a locally weighted scatterplot smoothing (LOWESS) of Tajima's D values along chromosome position for chromosomes 4 and 10 revealed that in these high \bar{F}_{ST} regions, the U.S. feral sub-population exhibited positive values while the European and Marijuana sub-populations exhibited negative values (Figure 7 B). Independent PCA analysis of chromosomes 4 and 10 identified

distant clustering of U.S. feral samples from all other sub-population's samples (Figure 7 C and D).

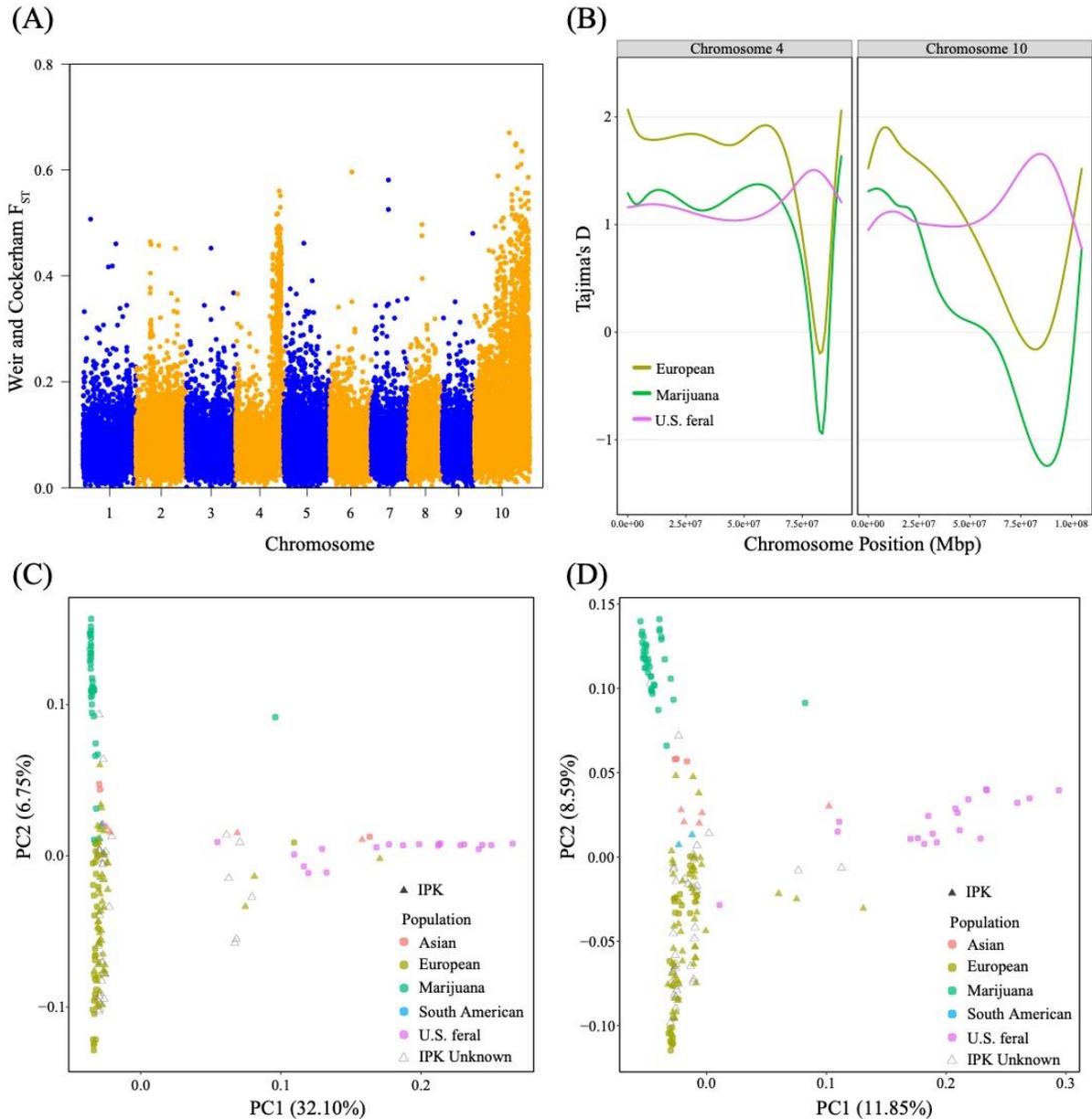


Figure 7: Sub-population divergence on chromosomes 4 and 10. (A) Sliding window \bar{F}_{ST} analysis comparing allelic diversity between Marijuana, European and U.S. feral sub-populations. Y-axis points indicate \bar{F}_{ST} scores across 20,000 base pairs. (B) LOWESS plots showing Tajima's D of SCOs plotting against physical positions of chromosomes 4 and 10 (base pairs). Differently colored lines indicate Tajima's D value for different sub-populations. (C) Plot of the first two principal components of PCA on all SNPs between 77Mb and 88Mb of chromosome 4. (D) Plot of the first two principal components of PCA on all SNPs identified on chromosome 10.

Loci under balancing selection

Of the 75 SCOs possessing mean Tajima's D values greater than or equal to the top 1% in each sub-population, 23 of these SCOs per sub-population possessed a mean site heterozygosity greater than or equal to the top 30% of each sub-population's respective mean site heterozygosity values (Supplemental Table 2). Thus, these 23 genes per sub-population were considered to be under balancing selection.

In the U.S. feral sub-population, the annotation of six genes (GenBank peptide accessions: XP_030484994.1, XP_030484921.1, XP_030504444.1, XP_030511229.1, XP_030509614.1, and XP_030486224.1) under balancing selection indicated their involvement in response to numerous biotic and abiotic stresses such as wounding, pathogen, oxidative, salt, and osmotic stress. In the European sub-population, four genes (XP_030510979.1, XP_030482503.1, XP_030478550.1, and XP_030499670.1) under balancing selection were related by their role in maintaining essential chloroplast processes such as: chloroplast gene expression, chloroplast movement in response to light, protein transport into the chloroplast, and thylakoid architecture. The marijuana sub-population also possessed four genes (XP_030510979.1, XP_030480662.1, XP_030493766.1, and XP_030480127.1) under balancing selection that were related by their roles in maintaining essential chloroplast function such as chloroplast gene expression, iron transport into the chloroplast, and chloroplast development during the embryonic stage. Across all these sub-populations one gene, XP_030510979.1, was constitutively under balancing selection. This gene's annotation is *transcription termination factor MTERF9 (MTERF9)*, a chloroplast transcription termination factor that has also shown evidence for being involved in plant response to abiotic stress in *Arabidopsis thaliana* (Robles et al. 2015).

Derived alleles under positive selection

Our analysis of the non-synonymous derived allele frequency spectrum identified 5,143 derived alleles distributed across 2,205 SCOs (Supplemental Table 3). The plotted joint frequency spectra revealed patterns of rare and common derived alleles across the U.S. feral, European and Marijuana sub-populations while also showing derived alleles at high frequency in one group and low frequency in another (Figure 8 A). Between the three *C. sativa* groups, we identified 43, 19, and 7 derived alleles in the U.S. feral, Marijuana and European samples respectively that fit our criteria for sub-population positive selection (Supplemental Table 4). Fisher's exact test at each of these loci confirmed significant ($p < 0.05$) allele frequency to sub-population associations.

Of the genes containing derived alleles showing evidence for sub-population positive selection, we identified one gene potentially contributing to local adaptation in the U.S. feral group whose gene annotation was *suppressor of RPS4-RLD 1 (SRFR1)*, CS_10 gene ID = XP_030499563.1), a gene which has been shown to be involved in defense response to pathogens in *Arabidopsis thaliana* (Kim et al. 2009). We identified interesting genes potentially contributing to local adaptation in the European and Marijuana sub-populations based on their best *Arabidopsis thaliana* homolog match. In the European sub-population, the best *Arabidopsis thaliana* homolog match for one of these genes was AT1G18740/*BYPASS1-LIKE* (annotated as XP_030493166.1 in CS_10), a gene which has been shown to be involved in freezing tolerance (Chen et al. 2019). In the Marijuana sub-population, the best *Arabidopsis thaliana* homolog match for one gene was AT1G09310/*SMALLER TRICHOMES WITH VARIABLE BRANCHES 2 (SVB2)*, annotated as XP_030506485.1 in CS_10), a gene which has been shown to regulate trichome formation (Hussain et al. 2021).

At the species wide level, we identified 439 derived alleles distributed across 349 genes that fit our criteria for species-wide positive selection (Supplemental Table 4). These 439 derived alleles were evenly distributed throughout the *C. sativa* genome (Supplemental Figure 14). Using the GO Term Enrichment tool implemented in TAIR, we identified 25, 14, and 4 GO terms for biological processes, cellular component and molecular function respectively that were significantly enriched (FDR $p < 0.05$, Supplemental Tables 5, 6, and 7). The GO terms with the greatest fold change for biological processes, cellular component and molecular function were chloroplast RNA processing, Golgi cisterna, and zinc ion binding respectively. Of the genes containing derived alleles showing evidence for species-wide positive selection, one was particularly interesting. This gene was annotated as *mevalonate kinase* (*MEV kinase*), which functions to phosphorylate mevalonate into mevalonate-5-phosphate; a key step in sesquiterpene biosynthesis via the mevalonic acid pathway (Bergman et al. 2019).

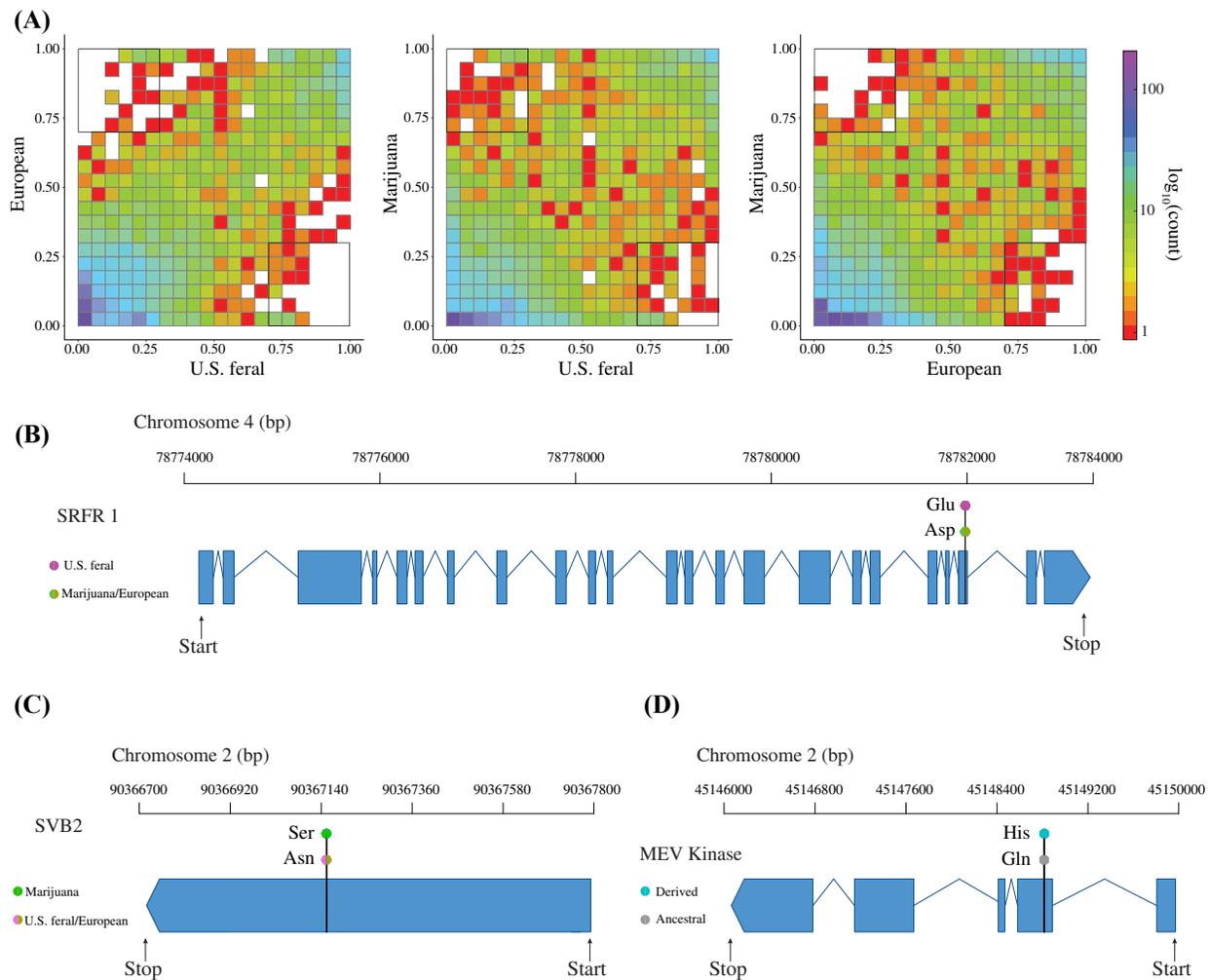


Figure 8: Positive selection in *C. sativa*. (A) Derived joint site frequency spectrums showing the derived non-synonymous allelic diversity between U.S. feral, European and Marijuana sub-populations. X and Y axes indicate the frequency of alleles within each bin (squares). The number of derived alleles with each frequency bin is color coded according to the logarithmic scale. The bolded black squares indicate derived alleles at $\geq 70\%$ frequency. (B) *SRFR1* gene model showing the three letter codes for the most frequent amino acid substitution within each sub-population. (C) *SVB2* gene model showing the three letter codes for the most frequent amino acid substitution within each sub-population. (D) Gene model for *MEV kinase* showing the three letter codes for the most frequent amino acid substitution under species-wide positive selection.

Discussion

Whether or not the out of Asia hypothesis for *C. sativa*'s origin is supported remains a central question when studying the evolution of this species. Recently, Ren et al. 2021 found evidence supporting the long-held hypothesis that Asian accessions represent the ancestral population of *C. sativa*. Interestingly, the central PCA clustering of the Asian sub-population we find suggests that this sub-population shares about equal allelic variation with all other sub-populations (Figure 6 A). If the Asian sub-population is ancestral, we would predict deeper coalescent events and thus a greater proportion of segregating sites in the Asian sub-population. After comparing the proportion of segregating sites at SCOs between sub-populations, our results further support an out of Asia origin of *C. sativa* by showing that the Asian sub-population exhibits a significantly greater ($p < 0.05$) proportion than all other sub-populations (Figure 6 C). Additionally, the significantly greater proportion of segregating sites present in the Asian sub-population suggests a greater effective population size. A greater Asian effective population size was supported by our comparisons of mean Tajima's D which suggested that all other sub-populations have undergone a recent contraction without time for expansion hence their more positive mean Tajima's D (Supplemental Figure 13). Thus, we predict deeper coalescent events in the Asian sub-population which is supported by their greater proportion of segregating sites across 7,540 SCOs.

Population structure

In contrast to Marijuana, the European and Asian sub-populations generally exhibited comparatively greater genomic variation (Figure 6 A and C). The broad PCA clustering and two dominant ancestry likelihoods (2 and 6) of European accessions may reflect subtle genetic changes that occurred within this sub-population through selective breeding of traits desired for

domestication. Interestingly, the population structure of the U.S. feral samples suggests that while this sub-population likely shares a recent common ancestor with European hemp varieties, it has since undergone significant genomic change which we hypothesize reflects natural selection against the domestication traits artificially selected for in their ancestors. Overall, IPK accessions with known source country information clustered with other samples from similar regions such as Europe and Asia (Figure 6 A). Except for the IPK accession CAN 36, IPK accessions without known source country information also clustered with samples sourced from Europe and Asia (Figure 6 A). Future studies should make efforts to collect and sequence additional georeferenced accessions across Asia and Africa.

Recent studies have made use of next generation sequencing methods to describe the population structure of *C. sativa* (Sawler et al. 2015, Lynch et al. 2016, Soorni et al. 2017 and Ren et al. 2021). A common finding of these prior studies is a clear genetic distinction between Marijuana and non-Marijuana type accessions. Additionally, Busta et al. 2022 recently analyzed the population structure of feral Nebraskan *C. sativa* and found that they were more closely related to hemp type cultivars than marijuana type cultivars. While we did find evidence to support these prior findings (Figure 6 A and B), our study also made large scale efforts to generate new WGS data of feral accessions in the U.S. (Colorado), European industrial hemp varieties, and accessions from the IPK seed bank. By including new WGS data of these previously unexamined accessions, in addition to publicly available data, our analyses provide results that expand the current knowledge of *C. sativa* population genomics.

Sub-population divergence

Consistent with our pairwise \bar{F}_{ST} results, numerous studies have previously documented relatively low population divergence between *C. sativa* sub-populations (Sawler et al. 2015, Lynch

et al. 2016 and Ren et al. 2021). The higher pairwise \bar{F}_{ST} scores of the feral U.S. vs. the European and Marijuana sub-population we found (Supplemental Table 1), however, is potentially a result from the broad high \bar{F}_{ST} genomic regions on chromosomes 4 and 10 (Figure 7 A). Additionally, in contrast to the European and marijuana sub-populations, the U.S. feral sub-population exhibited comparatively greater genetic variation in these regions as well based on our Tajima's D analyses (Figure 7 B). Due to their substantial size, we hypothesize these broad high \bar{F}_{ST} regions may be the result of chromosomal inversions maintained at high frequency in the feral U.S. sub-population. Chromosomal inversions have been previously associated with broad high \bar{F}_{ST} regions in other plant species such as sunflower (Todesco et al. 2020). Future studies should make efforts to assemble a chromosome scale reference genome representative of these U.S. feral accessions to validate any potential inversions in this sub-population.

Balancing Selection

Our Tajima's D and mean heterozygosity analyses together identified numerous genes under balancing selection within the U.S. feral, European, and marijuana sub-populations. While collectively all the genes we identified under balancing selection spanned a diversity of functions, subsets of genes within each sub-population were broadly related by their similar annotation. In the U.S. feral sub-population, six genes involved in plant response to various biotic and abiotic stresses were identified to be under balancing selection. This is interesting because a consequence of domestication is the loss of adaptive genetic variation for survival in nature to favor fitness for human purpose (Brown 2010). The U.S. feral sub-populations have been evolving under natural selection, which has likely favored survival traits that were selected against in their domesticated ancestors. Thus, we hypothesize that the maintenance of genetic variation at these stress response loci in the U.S. feral sub-population reflects selective pressure from a loss of domestication

syndrome that selects for traits which allow for adaptation to novel perturbations in this sub-population's wild landscape via heterozygote advantage.

In both the European and marijuana sub-populations, four genes broadly involved in the maintenance of essential chloroplast functions were identified to be under balancing selection. It is interesting that these two sub-populations, which contain many domesticated genotypes, both exhibit evidence for balancing selection at functionally related (yet distinct) loci despite being domesticated for different traits. While it remains unclear as to why, we hypothesize that, in contrast to the U.S. feral sub-population, the maintenance of genetic variation at genes broadly related to chloroplast function in both the European and marijuana sub-populations reflects selection for traits favored by domestication that exhibit a heterozygote advantage.

Interestingly, the gene *MTERF9* met our criteria for balancing selection across all three sub-populations. In *Arabidopsis thaliana*, mutagenesis of *MTERF9* has been shown to cause altered chloroplast gene expression, defective chloroplast development, and altered plant response to abiotic stresses (Robles et al. 2015). Our finding that genetic variation is maintained at *MTERF9* across divergent sub-populations of *C. sativa* is intriguing and suggests a species-wide heterozygous advantage at this gene. Similar findings of genomic regions with species-wide maintenance of heterozygosity have been observed in other domesticated plant species such as maize where it is hypothesized that a homozygous state of these loci produces a deleterious phenotype via inbreeding depression (Brandenburg et al. 2017 & Liu et al. 2018). The maintenance of genetic variation at *MTERF9* across *C. sativa* we find may indicate a similar phenomenon as predicted in maize where a homozygous state of this gene results in a species-wide cost to fitness given its broadly important role in chloroplast gene expression, development, and plant response

to abiotic stress. Thus, we hypothesize that the genetic variation maintained at *MTERF9* in *C. sativa* reflects species-wide balancing selection for heterozygote advantage.

Positive selection

Broadly, our analysis of the derived non-synonymous allele frequency spectra identified results consistent with population genetics theory wherein we observed both rare and fixed alleles in each sub-population (Nielson 2005). Furthermore, the results obtained from the joint derived allele frequency spectra suggest that while overall most high frequency derived alleles are common to the U.S. feral, European, and Marijuana sub-populations, there are some highly diverged loci across the genome which serve as the basis of higher within-group similarity (Figure 8 A).

Interestingly, the U.S. feral *C. sativa* contained the greatest proportion of high frequency derived alleles among the three sub-populations. This finding, in conjunction with the results from our other analyses, suggests that the feral U.S. sub-population has undergone significant genetic change which we hypothesize indicates local adaptation to new climates, biotic interactions, and relaxed selection on traits favored by domestication. Indeed, one gene containing a high frequency derived allele in the U.S. feral group which may contribute to local adaptation to novel pathogens was *SRFRI* (Figure 8 B), a gene for which numerous studies have documented its role in defense against pathogenic bacteria in *Arabidopsis thaliana* (Kim et al. 2009, Kim et al. 2010 and Li et al. 2010). Positive selection on *SRFRI* is also interesting because it suggests that in addition to maintenance of genetic variation at loci associated with stress response, adaptation to natural environments during feralization also occurs through fixation of newly emerged mutations.

By identifying the best homolog matches in *Arabidopsis thaliana* for unannotated genes with high frequency derived alleles, we were also able to discover candidate genes that potentially underlie local adaptation in the European and Marijuana sub-populations. In the European sub-

population, the gene annotated as XP_030493166.1 was found to be under positive selection. The best homolog match for this gene was AT1G18740/*BYPASS1-LIKE*, a gene which has been shown to confer freezing tolerance in *Arabidopsis thaliana* via the C – repeat binding factor pathway (Chen et al. 2019). We hypothesize that the positive selection on the derived allele in XP_030493166.1 reflects selection in domesticated hemp cultivars for increased fitness in lower temperatures experienced across European climates.

In the Marijuana sub-population, our finding that AT1G09310/*SVB2* was the best homolog match for an unannotated gene (XP_030506485.1) containing a derived allele under positive selection was particularly interesting. Variation in *SVB2* has been shown to regulate trichome number in *Arabidopsis thaliana* (Hussain et al. 2021). Positive selection on this gene in Marijuana is interesting because the primary sought after product from cultivation of marijuana is THC – a psychoactive cannabinoid that is largely produced in the glandular trichomes of female flowers via the polyketide pathway (Kovalchuk et al. 2020). Additionally, it has been shown in *C. sativa* that yield of cannabinoids is positively correlated with the number of capitate glandular trichomes present on bracts (Turner et al. 1981). The association between the functionality of *SVB2* and the primary cultivation of marijuana for THC suggests that the positive selection we observe on the derived allele in *SVB2* is driven by humans to develop domesticated cultivars with a greater number of glandular trichomes that subsequently produce greater quantities of THC. Future studies will need to utilize functional genetics to identify potential phenotypes associated with the derived alleles we identified in *SRFR1*, *BYPASS1-LIKE*, and *SVB2*.

Our efforts to identify positively selected derived alleles at the species-wide level yielded genes that spanned a diversity of significantly enriched ontologies (Supplemental Tables 5, 6 and 7). Interestingly, one gene with a derived allele positively selected at the species-wide level was

MEV kinase, which functions to phosphorylate mevalonate into mevalonate-5-phosphate (Bergman et al. 2019); a key step in sesquiterpene biosynthesis via the mevalonic acid pathway (Kovalchuk et al. 2020). While sesquiterpenes, such as beta caryophyllene, are known for their natural role in plant defense response (Rasmann et al. 2005), they are also of great interest to the *C. sativa* science community because of their use as food additives and potential for use in human medicine (Nuutinen 2018). The broad application of sesquiterpenes in both nature and humans suggests two possible hypotheses for why the derived allele in *MEV kinase* has been positively selected at the species wide level in *C. sativa*. First, we hypothesize that the derived allele was positively selected in the ancestral sub-population for enhanced resistance to novel pathogens. Alternatively, we hypothesize that the derived allele was indirectly selected for during the early domestication of *C. sativa* in Asia to produce cultivars with sesquiterpene production more well suited for food or medicinal applications (Li 1973). Future studies will need to utilize functional genetics to identify a phenotype associated with the derived allele we found in *MEV kinase* to validate our hypotheses.

Conclusion

In conclusion, as regulations around the world regarding *C. sativa*'s legality have increasingly relaxed, the volume of genomics studies on this species has steadily grown. This is exciting because it has enabled researchers to use whole genome methods for investigating fundamental population genetic questions about this multifaceted crop's evolutionary history. While progress has been made in understanding this species' evolution, several areas require additional attention such as describing the genetic variation present within the multitude of diverse uncharacterized accessions, formal testing of hypotheses regarding this species' origin, and identification of the loci under selection that genetically distinguish domesticated from feral sub-

populations. Our study addresses each of these areas by 1) large scale WGS of previously unexamined accessions, 2) further supporting the Asian origin of *C. sativa*, and 3) identifying loci under balancing and positive selection that potentially serve as the basis for local adaptation at both the sub-population and species-wide levels. While further steps remain necessary to: 1) sequence additional accessions, 2) assemble a genome representative of feral U.S. plants, and 3) functionally validate the derived alleles under positive selection identified here, we believe overall that the data generated, and results obtained from our study help to form a better understanding of the evolutionary history in *C. sativa*.

Data availability

All raw paired end fastq files produced from our WGS efforts have been deposited to NCBI's short read archive under BioProject Accession number: PRJNA866500. Additional data generated in this study are available in the supplemental information available at the G3 paper website (File S1). Supplemental Figure 15 shows the individual derived allele frequency spectrums for the U.S. feral, European, and Marijuana sub-populations respectively. Supplemental Tables 8-11 describe the sliding window F_{ST} scores, PCA values, means of total potential THC for IPK accessions, and Tajima's D and segregating sites of the SCO intervals respectively. The final filtered VCF has been deposited to Dryad and given the DOI link: <https://doi.org/10.5061/dryad.rv15dv49q>.

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. <https://doi.org/10.1101/gr.094052.109>
- Allaby, R. G., Ware, R. L., & Kistler, L. (2019). A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evolutionary Applications*. <https://doi.org/10.1111/eva.12680>
- Andrews, S. (2010). FastQC. *Babraham Bioinformatics*. <https://doi.org/citeulike-article-id:11583827>
- Bakker, E. G., Toomajian, C., Kreitman, M., & Bergelson, J. (2006). A genome-wide survey of R gene polymorphisms in Arabidopsis. *Plant Cell*. <https://doi.org/10.1105/tpc.106.042614>
- Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. D. W., Poh, Y. P., Hahn, M. W., ... Langley, C. H. (2007). Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*. <https://doi.org/10.1371/journal.pbio.0050310>
- Bergman, M. E., Davis, B., & Phillips, M. A. (2019). Medically useful plant terpenoids: Biosynthesis, occurrence, and mechanism of action. *Molecules*. <https://doi.org/10.3390/molecules24213961>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btu170>
- Brandenburg, J. T., Mary-Huard, T., Rigai, G., Hearne, S. J., Corti, H., Joets, J., ... Tenailon, M. I. (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1006666>
- Brown, A. H. D. (2010). Variation under domestication in plants: 1859 and today. *Philosophical Transactions of the Royal Society B: Biological Sciences*. <https://doi.org/10.1098/rstb.2010.0006>
- Busta, L., Dweikat, I., Sato, S. J., Qu, H., Xue, Y., Zhou, B., ... Zhang, C. (2022). Chemical and genetic variation in feral *Cannabis sativa* populations across the Nebraska climate gradient. *Phytochemistry*, 200, 113206. <https://doi.org/https://doi.org/10.1016/j.phytochem.2022.113206>

- Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype × Environment Interactions of Industrial Hemp Cultivars Highlight Diverse Responses to Environmental Factors. *Age*. <https://doi.org/10.2134/age2018.11.0057>
- Charlesworth B, Charlesworth D. *Elements of Evolutionary Genetics*. Roberts and Company; Greenwood Village, CO: 2010.
- Chen, T., Chen, J. H., Zhang, W., Yang, G., Yu, L. J., Li, D. M., ... An, L. Z. (2019). BYPASS1-LIKE, a DUF793 family protein, participates in freezing tolerance Via the CBF pathway in Arabidopsis. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2019.00807>
- D. Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.00731>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr330>
- Darwin, Charles, and Leonard Keble. On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. London: J. Murray, 1859.
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*. <https://doi.org/10.1186/s13059-019-1832-y>
- Falk, A., Feys, B. J., Frost, L. N., Jones, J. D. G., Daniels, M. J., & Parker, J. E. (1999). EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.96.6.3292>
- Feys, B. J., Moisan, L. J., Newman, M. A., & Parker, J. E. (2001). Direct interaction between the Arabidopsis disease resistance signaling proteins, EDS1 and PAD4. *EMBO Journal*. <https://doi.org/10.1093/emboj/20.19.5400>
- Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.12509>
- Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., ... Schwartz, C. J. (2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana . *New Phytologist*. <https://doi.org/10.1111/nph.17243>
- Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K. K., Zhang, Z., ... Li, X. (2016). The Jujube Genome Provides Insights into Genome Evolution and the Domestication of Sweetness/Acidity Taste in Fruit Trees. *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1006433>

- Hussain, S., Zhang, N., Wang, W., Ahmed, S., Cheng, Y., Chen, S., ... Wang, S. (2021). Involvement of aba responsive svb genes in the regulation of trichome formation in arabidopsis. *International Journal of Molecular Sciences*.
<https://doi.org/10.3390/ijms22136790>
- Kapun, M., & Flatt, T. (2019). The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Molecular Ecology*.
<https://doi.org/10.1111/mec.14871>
- Kim, S. H., Gao, F., Bhattacharjee, S., Adiasor, J. A., Nam, J. C., & Gassmann, W. (2010). The arabidopsis resistance-like gene SNC1 is activated by mutations in SRFR1 and contributes to resistance to the bacterial effector AvrRps4. *PLoS Pathogens*.
<https://doi.org/10.1371/journal.ppat.1001172>
- Kim, S. H., Kwon, S. Il, Saha, D., Anyanwu, N. C., & Gassmann, W. (2009). Resistance to the *Pseudomonas syringae* effector hopA1 is governed by the TIR-NBS-LRR Protein rps6 and is enhanced by mutations in srfr1. *Plant Physiology*. <https://doi.org/10.1104/pp.109.139238>
- Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., R. Ashnest, J., ... Sharbel, T. F. (2020). The Genomics of Cannabis and Its Close Relatives . *Annual Review of Plant Biology*. <https://doi.org/10.1146/annurev-arplant-081519-040203>
- Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., ... Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America*.
<https://doi.org/10.1073/pnas.1323964111>
- Laursen, L. (2015). Botany: The cultivation of weed. *Nature*. <https://doi.org/10.1038/525S4a>
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package ‘emmeans’. Estimated marginal means, aka least-squares means. *R Package Version 1.15-15*.
- Li, H. (2013). [Heng Li - Compares BWA to other long read aligners like CUSHAW2] Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*.
- Li, H. L. (1973). An archaeological and historical account of cannabis in China. *Economic Botany*. <https://doi.org/10.1007/BF02862859>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btp352>
- Li, Y., Li, S., Bi, D., Cheng, Y. T., Li, X., & Zhang, Y. (2010). SRFR1 negatively regulates plant NB-LRR resistance protein accumulation to prevent autoimmunity. *PLoS Pathogens*.
<https://doi.org/10.1371/journal.ppat.1001111>

- Liu, N., Liu, J., Li, W., Pan, Q., Liu, J., Yang, X., ... Xiao, Y. (2018). Intraspecific variation of residual heterozygosity and its utility for quantitative genetic studies in maize. *BMC Plant Biology*. <https://doi.org/10.1186/s12870-018-1287-4>
- Lynch, R. C., Vergara, D., Tittes, S., White, K., Schwartz, C. J., Gibbs, M. J., ... Kane, N. C. (2016). Genomic and Chemical Diversity in Cannabis. *Critical Reviews in Plant Sciences*. <https://doi.org/10.1080/07352689.2016.1265363>
- Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., ... Wang, J. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications*. <https://doi.org/10.1038/ncomms3320>
- Mondragón-Palomino, M., Meyers, B. C., Michelmore, R. W., & Gaut, B. S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Research*. <https://doi.org/10.1101/gr.159402>
- Moreau, M., Degrave, A., Vedel, R., Bitton, F., Patrit, O., Renou, J. P., ... Fagard, M. (2012). EDS1 contributes to nonhost resistance of *Arabidopsis thaliana* against *Erwinia amylovora*. *Molecular Plant-Microbe Interactions*. <https://doi.org/10.1094/MPMI-05-11-0111>
- Moyers, B. T., Morrell, P. L., & McKay, J. K. (2018). Genetic costs of domestication and improvement. *Journal of Heredity*. <https://doi.org/10.1093/jhered/esx069>
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btw044>
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Nuutinen, T. (2018). Medicinal properties of terpenes found in *Cannabis sativa* and *Humulus lupulus*. *European Journal of Medicinal Chemistry*. <https://doi.org/10.1016/j.ejmech.2018.07.076>
- Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., & Daniels, M. J. (1996). Characterization of eds1, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell*. <https://doi.org/10.1105/tpc.8.11.2033>
- Price, N., Moyers, B. T., Lopez, L., Lasky, J. R., Grey Monroe, J., Mullen, J. L., ... McKay, J. K. (2018). Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1719998115>
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., ... Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication

- and specialization. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1400975111>
- Rasmann, S., Köllner, T. G., Degenhardt, J., Hiltbold, I., Toepfer, S., Kuhlmann, U., ... Turlings, T. C. J. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*. <https://doi.org/10.1038/nature03451>
- Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M. L., Yang, Y., ... Fumagalli, L. (2021). Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Science Advances*. <https://doi.org/10.1126/sciadv.abg2286>
- Robles, P., Micol, J. L., & Quesada, V. (2015). Mutations in the plant-conserved MTERF9 alter chloroplast gene expression, development and tolerance to abiotic stress in *Arabidopsis thaliana*. *Physiologia Plantarum*. <https://doi.org/10.1111/ppl.12307>
- Rustagi, N., Zhou, A., Watkins, W. S., Gedvilaite, E., Wang, S., Ramesh, N., ... Xing, J. (2017). Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. *BMC Genomics*. <https://doi.org/10.1186/s12864-017-3767-6>
- Sawler, J., Stout, J. M., Gardner, K. M., Hudson, D., Vidmar, J., Butler, L., ... Myles, S. (2015). The genetic structure of marijuana and hemp. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0133292>
- Scossa, F., & Fernie, A. R. (2021). When a Crop Goes Back to the Wild: Feralization. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2021.02.002>
- Shultz, A. J., & Sackton, T. B. (2019). Immune genes are hotspots of shared positive selection across birds and mammals. *ELife*. <https://doi.org/10.7554/eLife.41815>
- Soorni, A., Fatahi, R., Haak, D. C., Salami, S. A., & Bombarely, A. (2017). Assessment of Genetic Diversity and Population Structure in Iranian Cannabis Germplasm. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-15816-5>
- Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation. *Genetics*.
- Team, R. C. (2019). R: A Language and Environment for Statistical Computing. *Vienna, Austria*.
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J. S., Soudi, S., Burge, D. O., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*. <https://doi.org/10.1038/s41586-020-2467-6>
- Turner, J. C., Hemphill, J. K., & Mahlberg, P. G. (1981). Interrelationships of glandular trichomes and cannabinoid content. I. Developing pistillate bracts of *Cannabis sativa* L. (Cannabaceae). *Bulletin on Narcotics*.

- van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., ... Geurts, R. (2018). Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1721395115>
- Vergara, D., Huscher, E. L., Keepers, K. G., Givens, R. M., Cizek, C. G., Torres, A., ... Kane, N. C. (2019). Gene copy number is associated with phytochemistry in Cannabis sativa. *AoB PLANTS*. <https://doi.org/10.1093/aobpla/plz074>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- West, D. P. (1998). Hemp and Marijuana: Myths & Realities. *North American Industrial Hemp Council*.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts606>

CHAPTER 4

ROOT PULLING FORCE ACROSS DROUGHT IDENTIFIES GENOTYPE BY ENVIRONMENT INTERACTIONS AND CANDIDATE GENES

Summary

High-throughput, field-based characterization of root systems for hundreds of genotypes in thousands of plots is necessary for breeding and identifying loci underlying variation in root traits and their plasticity. We designed a large-scale sampling of root pulling force, the vertical force required to extract the root system from the soil, in a maize diversity panel under differing irrigation levels for two growing seasons. We then characterized the root system architecture of the extracted root crowns. We found consistent patterns of phenotypic plasticity for root pulling force for a subset of genotypes under differential irrigation, suggesting that root plasticity is predictable. Using genome-wide association analysis, we identified 54 SNPs as statistically significant for six independent root pulling force measurements across two irrigation levels and four developmental timepoints. For every significant GWAS SNP for any trait in any treatment and timepoint we conducted post-hoc tests for genotype-by-environment interaction, using a mixed model ANOVA. We found that 8 of the 54 SNPs showed significant GxE. Candidate genes underlying variation in root pulling force included those involved in nutrient transport. Although they are often treated separately, variation in the ability of plant roots to sense and respond to variation in environmental resources including water and nutrients may be linked by the genes and pathways underlying this variation. While functional validation of the identified genes is needed, our results expand the current knowledge of root phenotypic plasticity at the whole plant and gene levels, and further elucidate the complex genetic architecture of maize root systems.

Introduction

Originating approximately 400 million years ago, roots evolved at least twice in early plant lineages and provided anchorage (Raven and Edwards, 2001). Following these origins, roots continued to evolve structural complexity and functions and serve many critical roles in the biology of plants. In addition to water and nutrient uptake, roots are responsible for anchorage in soil. Roots are also the site of rhizosphere biotic interactions, spanning a range of outcomes from pathogenic to beneficial, and involving microbes, insects, and other plants (Johnson and Rasmann, 2015; Lareen et al., 2016; Clarke et al., 2019). Despite these important roles, understanding the genetics and physiology of root systems in soils has been challenging, leading to the description of roots as the “Hidden Half” of the plant (Eshel and Beeckman, 2013). Overall, root systems have complex structures. Root System Architecture (RSA) is a term used to describe the spatial arrangement of multiple individual roots of several distinct root types of an individual plant, each at a different stage along its developmental trajectory (Smith and de Smet, 2012). RSA traits, then, are the result of the cumulative effects of development of many individual roots within an individual of a single genotype in a given environment.

In addition to their fundamental biological importance and intrinsic structural complexity, roots and root systems vary among species, genotypes and environments. Within a given species, RSA is highly plastic, responding to variation in nutrient status and soil composition (Karlova et al., 2021). RSA traits have been shown to respond to levels of nutrients such as phosphorus and nitrogen by regulating root growth and branching (Gaudin et al., 2011; de Jong et al., 2019). Differences in particle texture within a growth substrate have also been shown to affect RSA phenotypes (Rogers et al., 2016). Water availability regulates both timing of root development and directional patterns of root branching (Bao et al., 2014; Sebastian et al., 2016). These examples

highlight our growing understanding of the robust and diverse responses of root growth to environmental stimuli.

Phenotypic plasticity in the form of altered rates of growth and timing of development of organs and traits are common adaptive responses for both RSA and above ground traits in plants (White and Castillo, 1992; Strock et al., 2018). While phenotypic plasticity is a property of an individual, the degree to which individuals sense and respond to the environment can be represented as a genetic component (Falconer, 1990). This variation in how genotypes sense and respond to the environment is described statistically in an ANOVA as a genotype by environment interaction (GxE). GxE is a common property of quantitative traits (Lynch and Walsh, 1998), affecting both the range of phenotypic values and the rank of genotypes in different environments. Multi-environment genetic mapping studies have shown that this GxE is polygenic and can be resolved to individual genome regions and loci, where the effect size of quantitative trait loci (QTL) changes across different environments (Paterson et al., 2003; El-Soda et al., 2014; Lowry et al., 2019).

Phenotypic plasticity for agriculturally important traits, has been proposed as a breeding target for optimizing response to environmental stress (Nicotra and Davidson, 2010; Kusmec et al., 2017; Kusmec et al., 2018). Implementing this strategy will require a deeper understanding of the extent and the genetic underpinning of GxE for the key traits. The core function of roots in obtaining water and nutrients highlights the potential for utilizing RSA in crops to buffer against environmental volatility due to climate change (Voss-Fels et al., 2018). Root phenotypic plasticity has an unclear, and potentially complicating, role in RSA under variable water and nutrient conditions (Schneider and Lynch, 2020). Future breeding efforts for RSA will be aided by a fuller understanding of GxE interactions affecting roots.

Like many traits, our understanding of the fundamental genetic control of RSA has largely been driven by studies in the model dicot *Arabidopsis thaliana* (Petricka et al., 2012). In addition to elucidating the core conserved pathway of root development and signaling, genetic loci associated with GxE responses of *Arabidopsis* roots to nutrient stress have been identified (Rosas et al., 2013). In contrast with the relatively simple roots of *Arabidopsis*, root systems of important monocot cereal crops such as maize, wheat, and rice have fundamentally different and more complex structures (Smith and de Smet, 2012). Accordingly, our understanding of the genetic architecture of RSA in these crop species is poor in comparison to *Arabidopsis* and related dicots. Most studies examining the genetics of RSA in cereal crops have been done on young plants grown in controlled conditions (Tracy et al., 2020).

Relatively few studies have attempted to characterize the genetic architecture and phenotypic plasticity of RSA in mature field grown cereal crops due to the challenges of measuring these traits. Destructive phenotyping of roots using excavation followed by image-based techniques have been effective in mapping RSA traits across different crop species (Trachsel et al., 2011; Schneider et al., 2020; Zheng et al., 2020). Analysis tools to extract phenotypic data from images of excavated root systems have been developed (Das et al., 2015; Zheng et al., 2020). Loci associated with variation in maize RSA traits have been identified in two recent genome wide association studies (GWAS), both of which used shovelomics (Trachsel et al., 2011) based techniques for root system excavation (Schneider et al., 2020; Zheng et al., 2020). Schneider et al. (2020) used two different diversity panels across three years in two locations including two levels of irrigation at one of these locations and measured image-based RSA traits such as lateral root length and root angle. Their multi-year and multi environment phenotyping efforts were followed by GWAS analysis in FARMCPU which identified candidate genes associated with variation in

both the plasticity and mean trait values within each environment. Zheng et al. (2020) used the shoot apical meristem (SAM) diversity panel in a single environment and growing season. While their efforts did identify candidate genes associated with numerous image-based RSA traits, the lack of multiple environments and growing seasons in Zheng et al. (2020) motivates further use of the SAM population to investigate the plasticity and genetic control of RSA. Overall, the consensus from these shovelomics based GWAS studies is that maize RSA is highly polygenic, controlled by multiple low to moderate effect loci, and that GxE is a strong component of RSA genetics because of the inconsistent set of identified candidate genes.

We focus on root pulling force (RPF) which is an alternative and a higher throughput technique compared to shovelomics based excavation for removing root systems from a field. RPF is quantified through the process of attaching the base of the shoot with a rope to a digital force gauge and applying manual, vertical force until the root system is extracted from the soil. The force required to pull the root system out of the ground is measured by the force gauge and recorded as RPF. This method provides a simple, instantaneous and quantitative measurement of the roots system during extraction. We have successfully used the RPF technique to identify QTL for root system size in *Brassica napus* (Fletcher et al., 2015), and QTL for it have been found in maize (Lebreton et al., 1995). While RPF can serve as a simple measure for root architecture traits, the extracted root systems remain amenable to image-based measurements as in excavation-based root extraction. Imaging of pulled root systems has shown RPF to be highly indicative of root system volume and surface area (Shao et al., 2021). Due to the simplicity and individual plant sampling of the RPF method, it is also amenable to automation to increase the throughput of RSA measurements in the field for improved mapping studies.

Here we used RPF, along with subsequent imaging of pulled root systems, to examine means and plasticity of RSA traits in the SAM diversity maize panel. We phenotyped this population at multiple developmental stages across two field seasons and differential irrigation treatments to investigate GxE in RSA traits. We then performed GWAS on these measurements, which identified a number of candidate genes potentially underlying the variation in root traits. Important questions addressed in this work are: (1) What proportion of trait variance in RPF and RSA in maize is explained by genetics, the environment and GxE? (2) What is the genetic architecture of RPF and RSA in maize across development and irrigation treatments? (3) Is there evidence for GxE at the individual gene level?

Materials and methods

Field experiments

367 lines from the SAM diversity panel (Leiboff et al., 2015) and 4 inbred check lines were grown at the Colorado State University Agricultural Research Development & Education Center in Fort Collins, CO, USA (40.649 N, -105.000 W) in 2018 and 2019. In 2018 seeds were planted on 22-25 May and in 2019 on 14 May using a split-plot design with full irrigation (FI) or limited irrigation (LI) treatments, with three field replicates per treatment. Prior to planting, the fields were fertilized to recommendations for 200 bu/ac yield, amounting to 190 lb/ac N and 25 lb/ac P₂O₅ in 2018 and 190 lb/ac N, 60 lb/ac P₂O₅ and 4 lb/ac Zn in 2019. Each plot consisted of two 12-foot rows with 30-inch spacing between rows and 9-inch spacing between plants within rows. The irrigated treatments received approximately 1 inch of water per week, while the drought treatments were irrigated until well established (approx. 5 weeks after planting) and then received only natural precipitation, except at the root harvesting when it also received irrigation to homogenize the root harvesting process (Supplemental Table 10). This irrigation differential began 500-600

growing degree days after planting, a time in which the average developmental stage was approximately V4. The timing of the root harvests was categorized developmentally relative to the timing of anthesis in the population (Figure 9), pre-flowering for the pulling at 71-73 d after planting, prior to any lines reaching anthesis, early flowering at 61-66 d after planting when roughly 5% of lines had reached anthesis, mid-flowering at 91-93 d from planting when roughly 30% had reached anthesis, and post-flowering at 109-119 d from planting when 98% of lines had reached anthesis. The variation in RPF timepoints allowed us to assess how the various developmental stages might affect root system excavation in this diverse population. This information will improve the efficiency of future root-pulling events. Days to anthesis was measured for each plot as the date at which 50% of the plants in the plot had begun to shed pollen and was determined approximately 3 times per week.

For root phenotyping, a rope was attached to the stalk above the root crown with a slip knot, and a digital force gauge (DS2, Imada Inc., Northbrook, IL, USA) was used to pull the root crown vertically from the soil. Root crowns were cleaned with water and imaged to obtain

additional RSA measurements using the DIRT image analysis software (Das et al., 2015) before being air-dried to constant weight for measurement of mass.

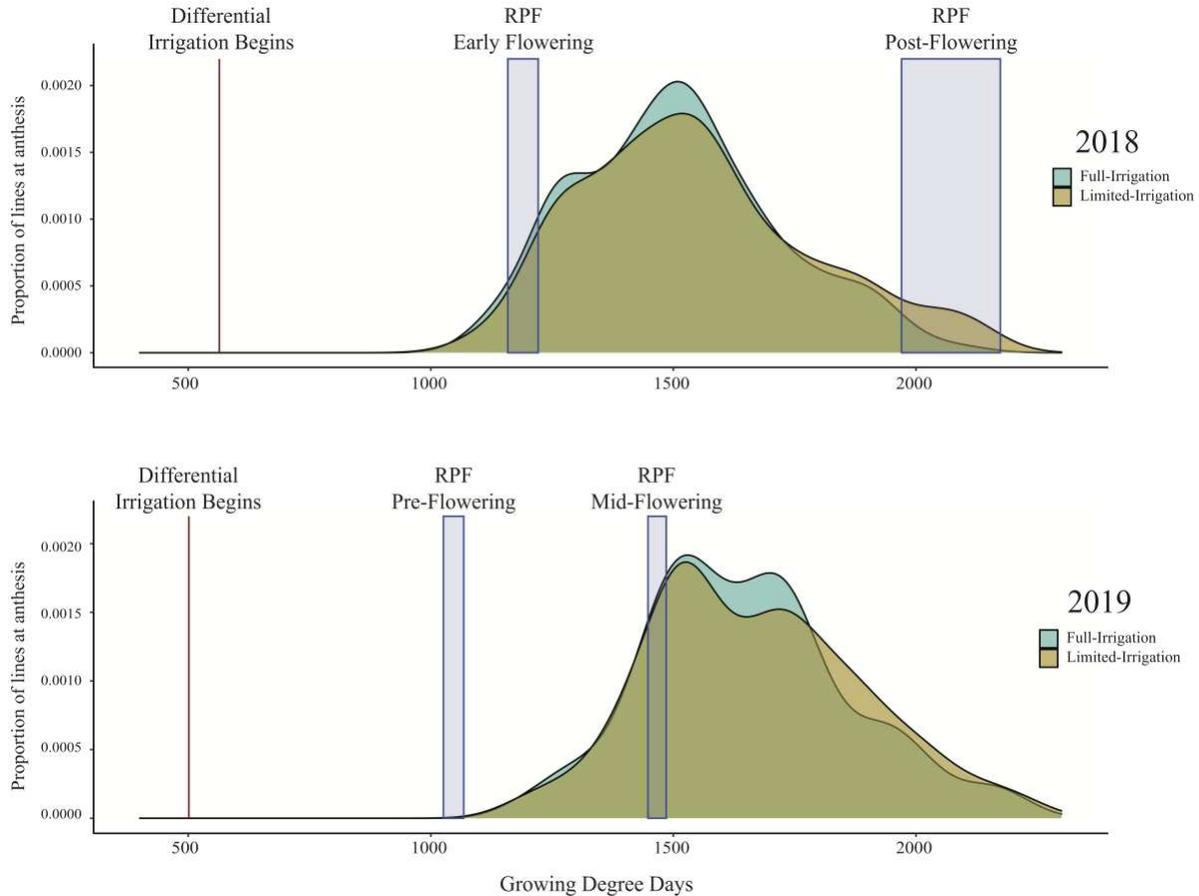


Figure 9 Histograms of days to anthesis for the field experiments. The timing of the root pulling events are highlighted and assigned developmental stages relative to flowering: pre-flowering for the pulling at 71-73 d after planting, prior to any lines reaching anthesis, early flowering at 61-66 d after planting when roughly 5% of lines had reached anthesis, mid-flowering at 91-93 d from planting when roughly 30% had reached anthesis, and post-flowering at 109-119 d from planting when 98% of lines had reached anthesis.

Genetic Correlations

Spearman’s rank correlation coefficients were calculated using the “chart.Correlation” function in the “Performance Analytics” package in R

(<https://github.com/braverock/PerformanceAnalytics>).

Quantitative Genetic Analyses

We estimated broad-sense heritability of traits using a random effect one-way ANOVA model. We split phenotypic data by trait, timepoint, and treatment and used the lmer function in the R package “lme4” v1.1-21 (Bates et al., 2015), treating genotype as a random effect. Additionally, we estimated phenotypic variance due to genotype, the environment and GxE by using a random effect two-way ANOVA model with interaction using the lmer function from the R package “lme4” v1.1-21. Genotype, the environment and GxE were all treated as random effects in the model.

GWAS

To analyze phenotype data via GWAS, the best linear unbiased predictors (BLUPs) were calculated from genotype simple means by treating genotype as a random effect using the R package “lme4” v1.1-21 (Bates et al., 2015) (See Supplemental Dataset S1 with BLUP values used for GWAS). A genotype matrix in HapMap format containing 1.2 million single nucleotide polymorphism (SNP) calls for the SAM diversity panel was then obtained from Leiboff et al. (2015). To reduce the impact of rare erroneous SNP calls, we imposed a minor allele frequency filter of 5% on the genotype matrix. After filtering for minor allele frequency, the genotype matrix contained approximately 860,000 SNPs. GWAS was conducted using an R implementation of FarmCPU (Liu et al., 2016) via GAPIT (Lipka et al., 2012) by following the guidelines at <https://www.zzlab.net/GAPIT/>. The first three principal components as well as a kinship matrix calculated using GAPIT were used as covariates to control for population structure. Normality of p-value distributions were assessed by inspecting each trait’s Q-Q plot output by GAPIT (Supplemental Figure 16). Because each trait’s Q-Q plot displayed p-value distributions that followed approximate normality, we proceeded with the FarmCPU results. SNPs whose

significance passed the Benjamini – Hochberg false discovery rate threshold (0.05) were further investigated to identify the gene or genes within a 10-kb window with which they may be associated.

Statistical tests for significant GWAS hits

To identify QTLxTreatment effects among significant GWAS hits, we performed ANOVA on each of these SNPs. We constructed models using the `lm` function in R. We estimated the effects of SNP, treatment, and SNP-by-treatment interactions. We performed PCA analysis using our SNP genotype data. We included values from the first three principal components in each ANOVA to account for population structure among genotypes. Using the `Anova()` function from the R package “`Car`” (Fox and Weisberg, 2018) we performed a type 3 ANOVA.

Gene expression data are from (Sekhon et al., 2011). We compared mean normalized expression for root tissue and pooled leaves from the V1-stage plants.

Results

RPF Measures Root System Size and is Correlated with RSA Traits in the Field across Multiple Developmental Timepoints

We performed a field experiment during Summer 2018 and 2019 in Fort Collins, Colorado using 367 diverse lines from the SAM maize panel (Leiboff et al., 2015) with full irrigation (FI) and limited irrigation (LI) treatments. We measured RPF at four different timepoints across 2018 and 2019, roughly corresponding to pre-flowering, early flowering, mid-flowering, and post-flowering (see methods). RPF measurements for the SAM population predictably increased with time with a mean of 47 kg pre-flowering to a mean of 120 kg at the end of season, slowing after flowering (Figure 10). Plants in the LI treatment had lower RPF measurements, averaging 70-80% of the full-irrigation (FI) RPF values. Above-ground biomass was also reduced in LI, 60-90% of

FI across the four stages. We estimated the broad-sense heritability of RPF as $H^2 = 0.49-0.59$ in the FI treatment, and generally lower estimates for the LI treatment ($H^2 = 0.36-0.50$, Figure 11), but sufficient for mapping experiments.

To quantify covariance between RPF measurements and RSA traits, conducted Spearman's correlation analysis, which showed that RPF measurements were highly positively correlated with RSA traits at all timepoints and treatments (Figure 12 and Supplemental Tables 12, 13). RPF showed the greatest positive correlation with root mass and area. This is consistent with our previous study of 3D X-ray tomography measurements from the root crowns of a subset of SAM lines, where RPF was highly associated with root volume and surface area (Shao et al., 2021). As further evidence for the utility of RPF as a sampling method, estimates of heritability for RPF were greater than or equal to those for RSA traits (Figure 11). In addition to root traits, we measured the shoot mass of pulled plants and recorded flowering time as growing degree days to anthesis (GDDTA) of the SAM panel. Shoot mass was highly positively correlated with RPF and other RSA traits (Figure 12 and Supplemental Tables 12, 13). Flowering time has a more variable and overall weaker correlation with RSA and plant biomass traits across timepoints and treatments (Supplemental Tables 12, 13).

To assess the relative utility of the root systems that we extracted by RPF, we extracted root systems from a subset of 45 plots using both RPF and shovel-based excavation methods (Trachsel et al., 2011). The mass of root systems extracted from the two techniques did not significantly differ ($p > 0.05$, Supplemental Figure 17A), and genotypes varying in RSA showed consistent differences in plant form with both techniques (Supplemental Figure 17B).

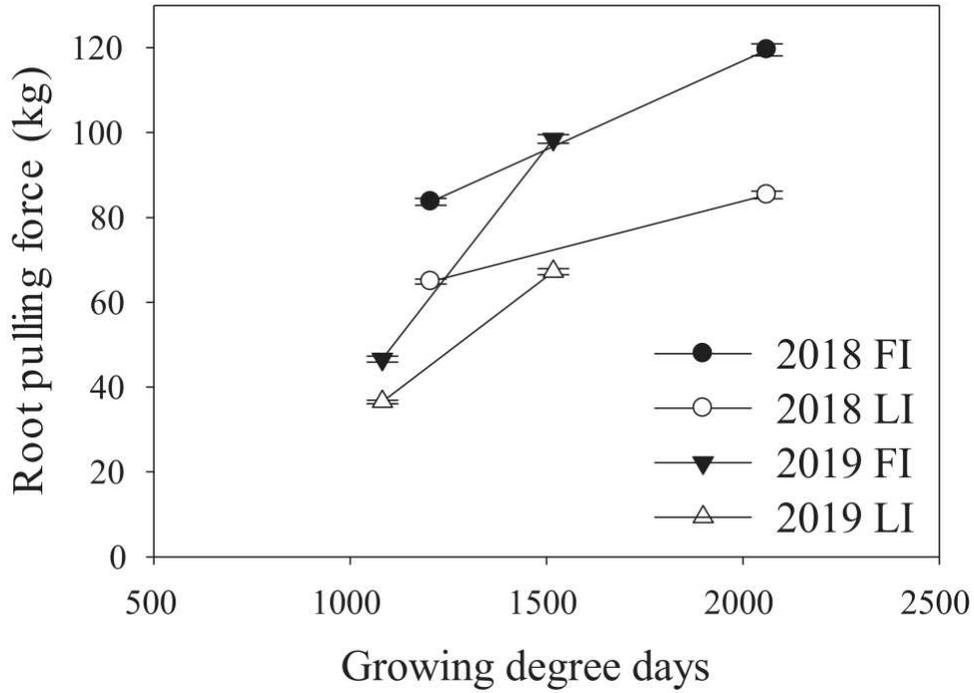


Figure 10 Changes in RPF over the field growing seasons. Measurements for the two field seasons are given (mean \pm SE) for the full (FI) and limited (LI) irrigation treatments.

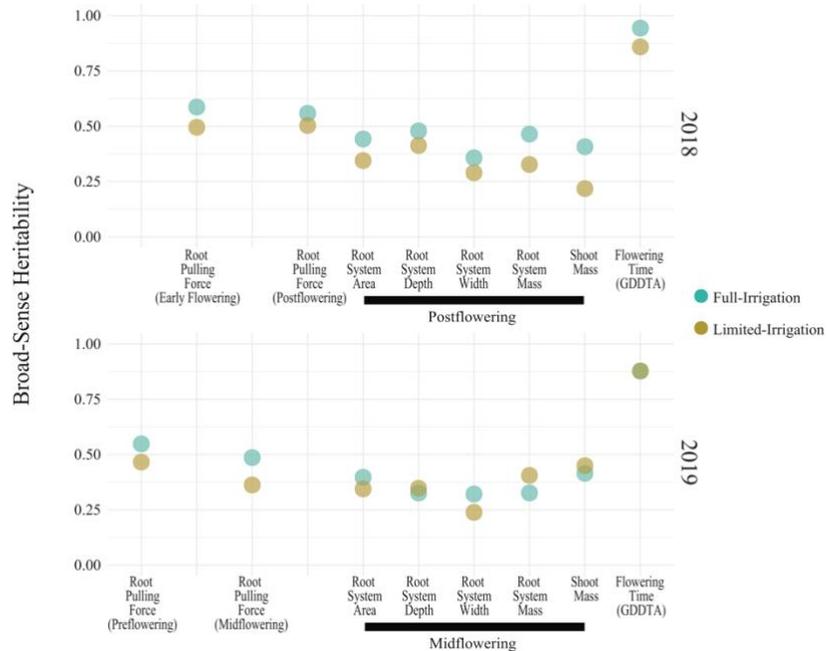


Figure 11 Broad-sense heritability of RPF and other root traits in the field.

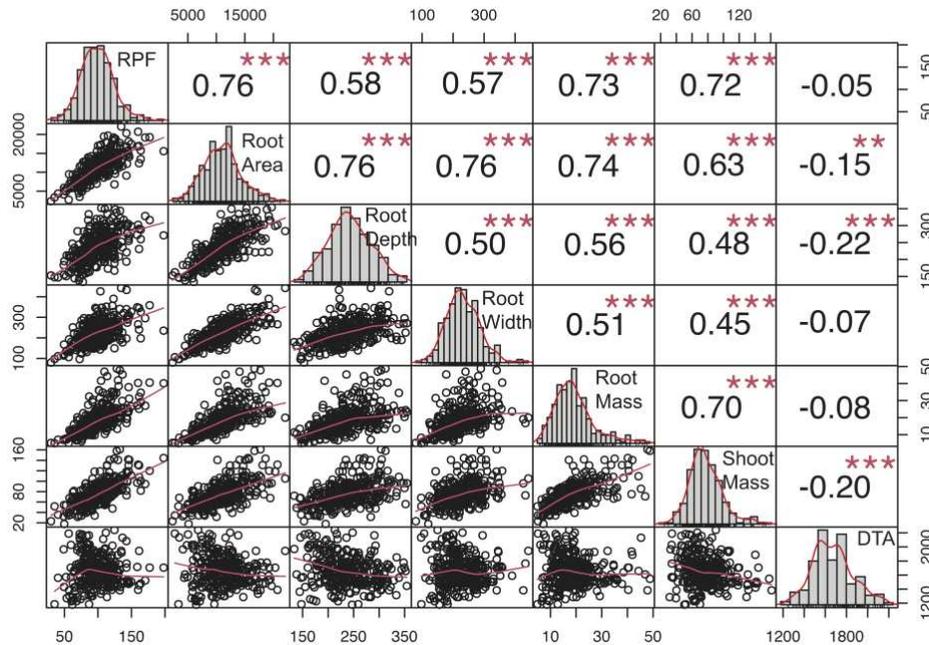


Figure 12 Genotypic correlations among traits in the 2019 full-irrigation treatment, mid-flowering. Values are Spearman's correlation coefficients. ** and *** indicate significant correlations, $P < 0.01$ and 0.001 respectively.

There is a Wide Range of Plasticity in RPF in Response to Water Limitation across the SAM Panel

For RPF, we found a moderate genotypic correlation between irrigation treatments and field seasons (Figure 13), suggesting that there may be differences among genotypes in the SAM population for plasticity to water limitation. For RPF, there was a high average plasticity response across the population, with an overall reduction in RPF under LI treatments (Figure 14). Among individual genotypes, however, there was a large range of plasticity of RPF in response to water limitation. As examples, we highlight six genotypes with consistent responses to irrigation in RPF across field seasons (Figure 14). These groups of genotypes showed large differences in plasticity. One group exhibited a large reduction in root system size under limited irrigation, while the other had increased root size under water limited conditions.

Across treatments and growing seasons, variance component analysis indicated that GxE accounted for a range of 1.2 to 9.8% of the phenotypic variation in RPF (Supplemental Figure 18). Overall, across years we found a significant correlation in plasticity in RPF to water limitation, on a percent of wet treatment basis ($r = 0.2$, $P < 0.01$). This plasticity in RPF is positively associated with plasticity in shoot mass, even on a percent of wet treatment basis ($r = 0.4-0.6$, $P < 0.001$).

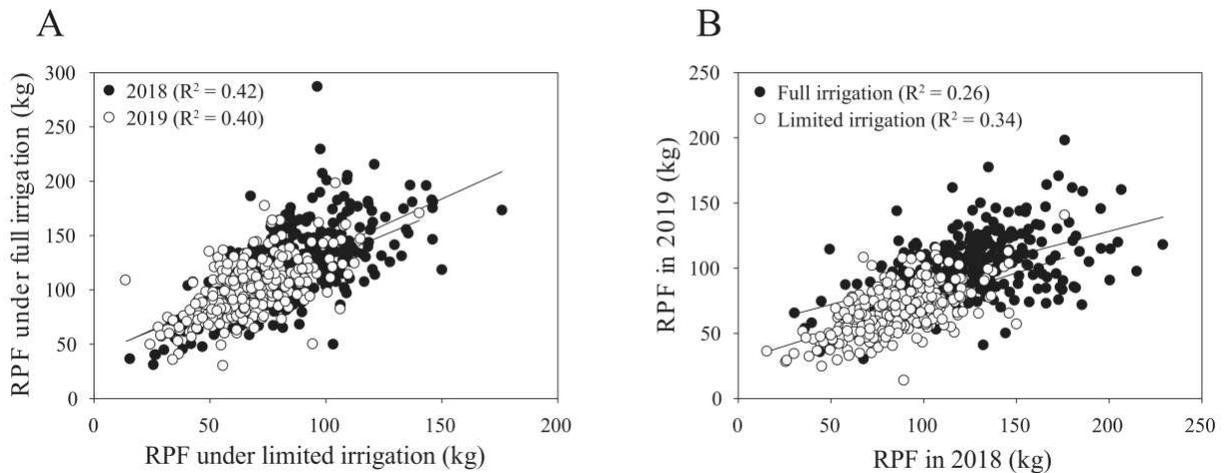


Figure 13 Genotypic correlation for RPF of lines among environments. (A) Correlations between treatments, for 2018 $R^2 = 0.42$ and for 2019 $R^2 = 0.40$. (B) Correlations between years, for FI $R^2 = 0.26$ and for LI $R^2 = 0.34$. Data shown are from the post-flowering stage in 2018 and mid-flowering in 2019.

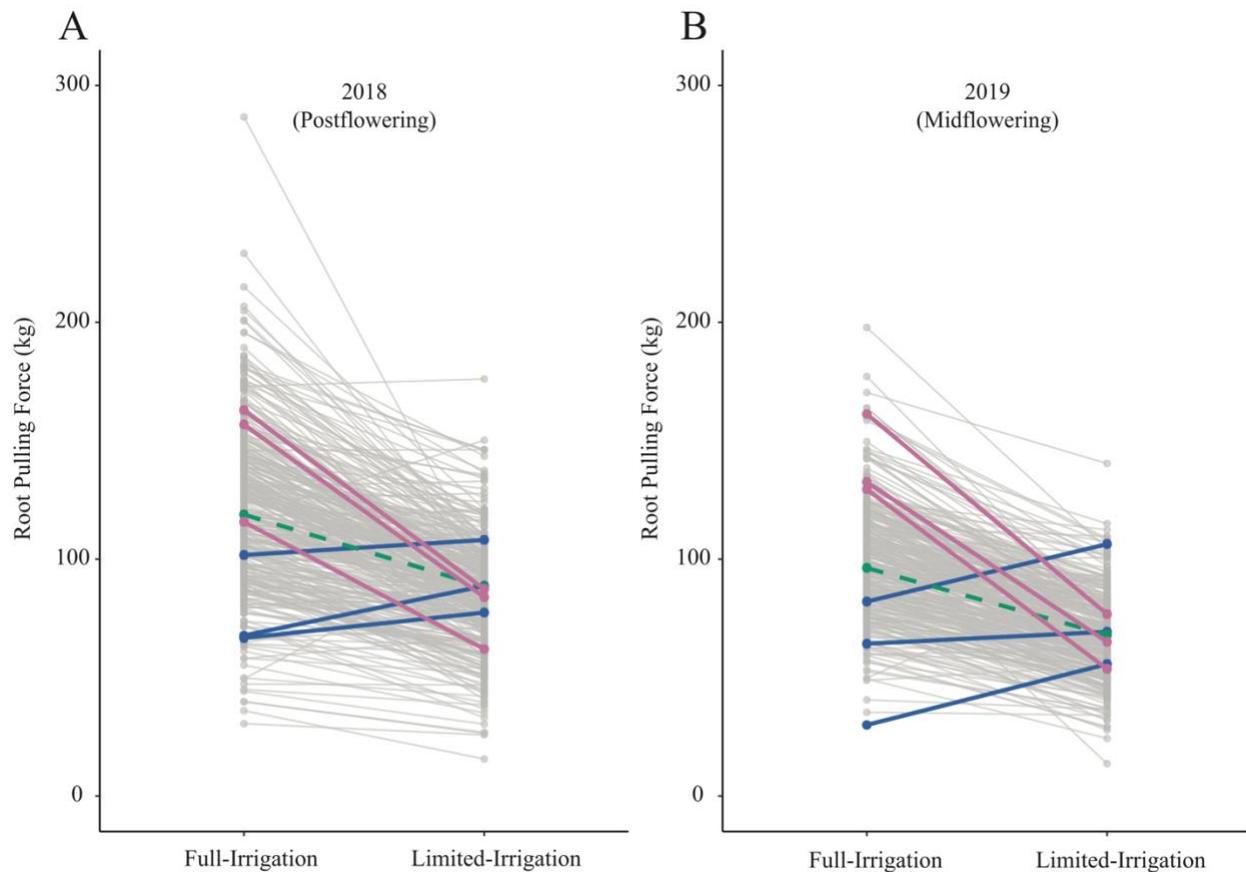


Figure 14 Reaction norms for RPF across irrigation treatments in 2018 (A) and 2019 (B). Blue lines highlight three genotypes (I205, H91, and S8324) that consistently increased RPF under LI; Pink shows three genotypes (A661, 207, and PHN11) where RPF decreased strongly in both years under LI; Green indicates genotype 793, which was approximately average for response to LI.

GWAS Identifies Numerous SNPs Associated with RPF and RSA Traits

The FarmCPU GWAS model controls for false discoveries by accounting for kinship and population structure while still providing high statistical power for candidate gene identification (Liu et al., 2016). Additionally, the relatively rapid LD decay in maize results in good resolution in mapping through GWAS (Yan et al., 2009). We performed GWAS using an R implementation of FarmCPU through GAPIT (Lipka et al., 2012) to identify SNPs associated with RPF, RSA, biomass traits, and flowering time. We identified 54 significant GWAS SNPs for RPF across years,

developmental timepoints, and treatments, along with 6 for root mass, 3 for shoot mass, and 21 for RSA traits (Supplemental Table 14). For RPF, the significant GWAS SNPs varied between treatments with more SNPs identified in FI than LI. None of the RPF significant GWAS SNPs overlapped between the treatments, which may be due to GxE, but is also expected given the way that FarmCPU selects SNPs to include in the model. In addition, we also found 48 significant GWAS SNPs for days to anthesis (Supplemental Table 15); however, we saw no overlap in hits between our root traits and flowering, consistent with the lack of correlation in Figure 12.

The quantitative genetic analysis of RPF revealed GxE for drought (Supplemental Figure 18). In addition, we saw lack of overlap of QTL between FI and LI treatments within a timepoint for RPF and other root traits. This is in contrast to flowering time where GxE effects appear minimal (Supplemental Figure 18) and GWAS analysis identified hits in common between FI and LI treatments (Supplemental Table 15). To test for QTLxTreatment effects for the quantitative traits we measured, we performed a post-hoc ANOVA on each of the 54 significant GWAS SNPs for RPF that we identified. We tested effects of SNP, treatment, and SNP-by-treatment interaction. To account for potentially confounding effects of population structure, we included PCA scores generated from SNP data in the ANOVA model. Through this analysis, we found that eight of the significant GWAS SNPs for RPF showed a significant QTLxTreatment effect (Supplemental Table 14). In contrast, none of the GWAS SNPs for flowering time had a significant QTLxTreatment effect (Supplemental Table 15).

The most significant GWAS hit for root traits was on chromosome 10 and identified from GWAS of the full-irrigation RPF measurement in 2018 (Figure 15A). The candidate gene Zm00001eb427000 is the only gene model within a 10-kb window of the significant GWAS SNP and characterized as the low-affinity ammonium transporter *AMT5* (Dechorgnat *et al.*, 2019)

which shows a prominent root specific expression profile (Figure 17). The alternate allele was associated with lower RPF not only in the FI treatment but also in the LI treatment (Figure 15B). However, the effect size of this polymorphism on RPF was smaller in the dry treatment, with significant GxE at this trait associated SNP. Although this SNP was only reported significant in FarmCPU at a single developmental stage, we found that allele-specific differences in RPF began to emerge at early flowering, but it may be that the effect sizes of this polymorphism varies across development (Figure 15C). The differences in RPF that we observed captured differences in other RSA traits such as root area and depth (Figure 15D,E).

Another significant GWAS SNP showing significant GxE during early flowering was within an exon of the gene Zm00001eb159490, a *SLAH2* nitrate channel (Figure 16A, B). Although the significant GWAS SNP was found at the early-flowering stage, the alternate allele was associated with significantly higher ($p < 0.05$) RPF at all developmental stages sampled (Figure 16C). Image analysis of root systems indicated that there were significant differences in root system depth and area between the alleles at the significant GWAS SNP (Figure 16D, E). *SLAH2* also has a prominent root specific expression profile (Figure 17), in line with the root phenotypes we observe. This is consistent with localization in Arabidopsis where the homologs are found predominantly in the root stele (Zheng et al., 2015).

Another notable candidate gene for RPF was *phosphate transporter 1;2A*, *PHO1;2A* (Supplemental Table 14), is also a nutrient carrier. It shows a similar root specificity with the two root candidates mentioned above (Figure 17). Similar to *SLAH2*, *PHO1;2A* shares a similar localization to the root stele (Hamburger et al., 2002). Additionally, like *AMT5* and *SLAH2*, the significant GWAS SNP in *PHO1;2A* was also associated with differences in basic RSA traits such as root system depth and width (Figure 18). The allele effects in *PHO1;2A* had a significant effect

($p < 0.05$) on root system width. This is consistent with the role of *PHO1;2A* in phosphorus transport, given phosphorus signaling's role in regulating shallow root angles (Liao et al., 2001). These candidates highlight the important role of nutrient signaling in the shaping of RSA.

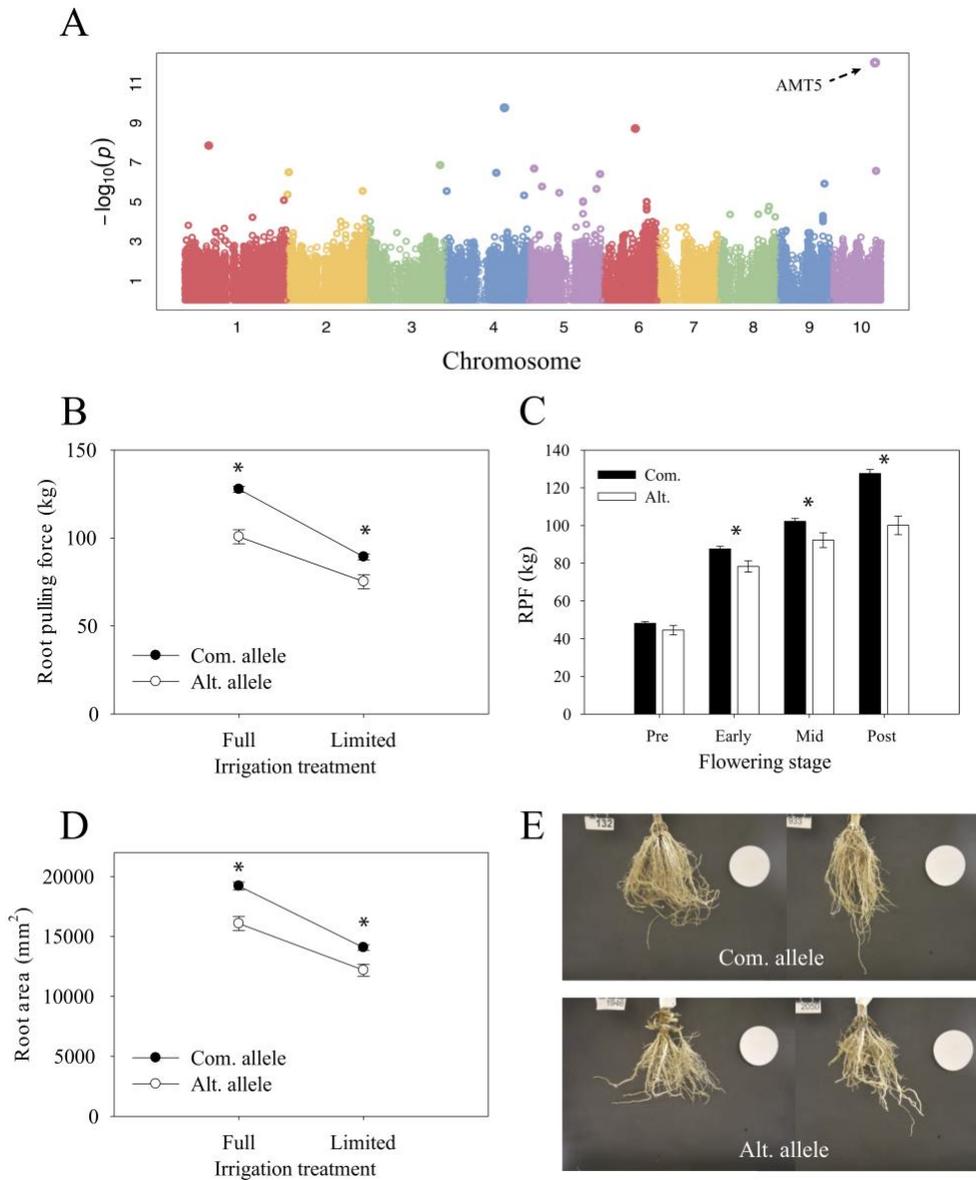


Figure 15 Characterization of candidate gene *AMT5* **A**, Genome-wide association mapping of RPF under well-irrigated conditions at 2018 post-flowering stage. **B**, Reaction norm of RPF from 2018 post-flowering measurements for lines with contrasting alleles (ls mean \pm SE). **C**, Differences in RPF by allele across developmental time points for the full-irrigation treatment (mean \pm SE). * Indicates a significant difference, $P < 0.05$. **D**, Reaction norm of root area (mean \pm SE). * Indicates a significant difference, $P < 0.05$. **E**, Representative root images for lines with the common allele (top) or alternate allele (bottom).

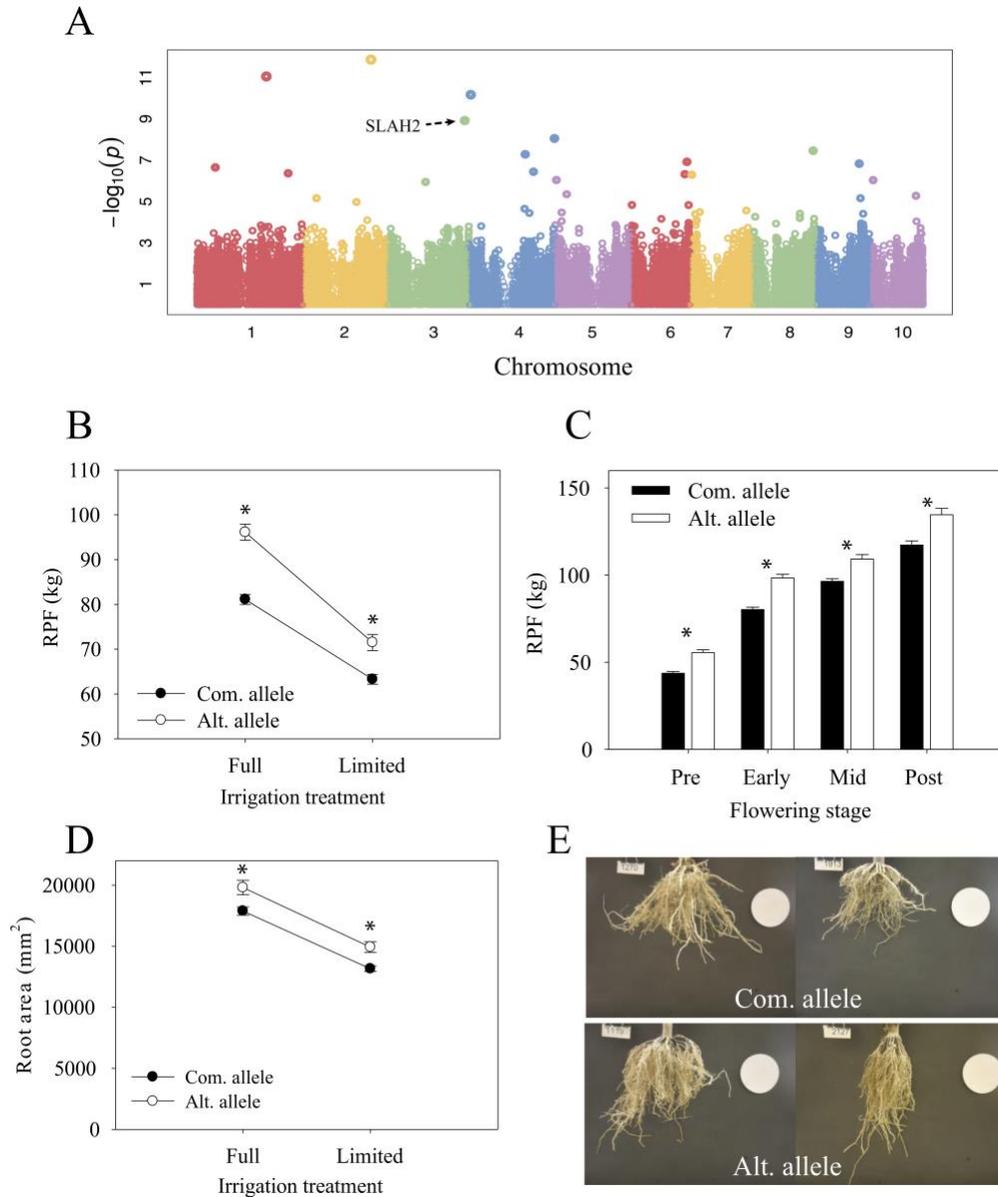


Figure 16 Allele differences at candidate gene *SLAH2*. **A**, Genome-wide association mapping of RPF under well-irrigated conditions at 2018 early-flowering stage. **B**, Reaction norm of RPF from 2018 early-flowering measurements for lines with contrasting alleles (1s mean \pm SE). **C**, Differences in RPF by allele across developmental time points for the full-irrigation treatment (mean \pm SE). * indicates a significant difference, $P < 0.05$. **D**, Reaction norm of root area (mean \pm SE). * indicates a significant difference, $P < 0.05$. **E**, Representative root images for lines with the common allele (top) or alternate allele (bottom).

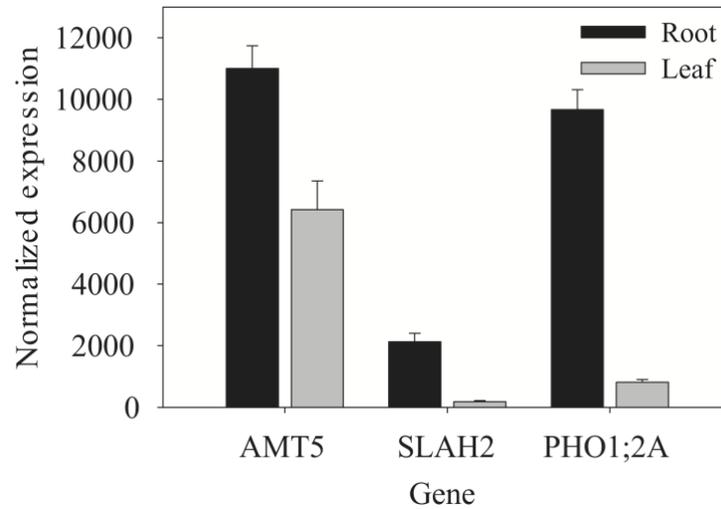


Figure 17 Gene expression in root and shoot organs among nutrient transport candidate genes. Data (mean \pm SD) are for V1-stage plants from Sekhon et al. (2011).

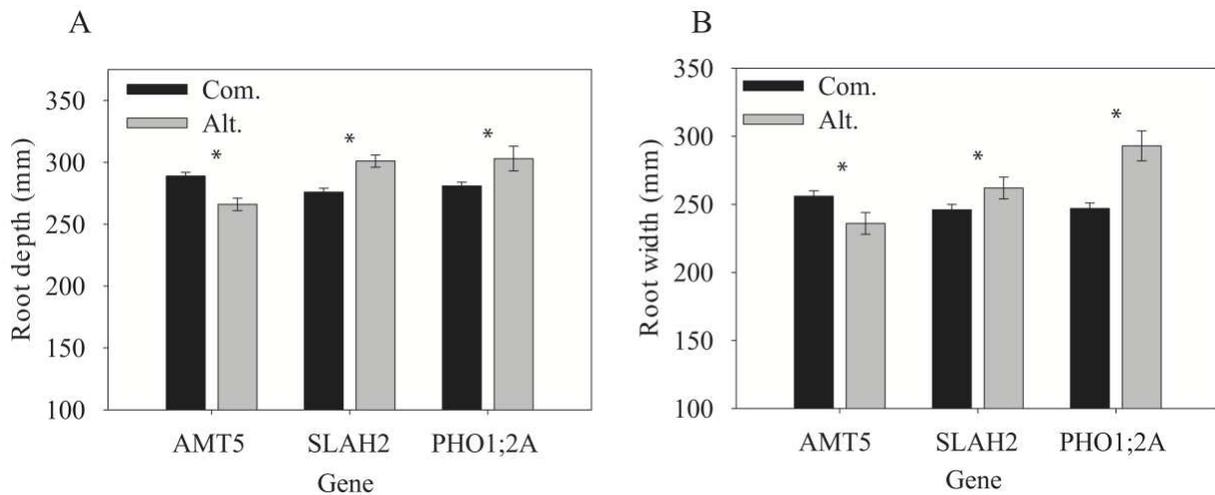


Figure 18 RSA traits of nutrient channel candidate genes, root system depth (A) and width (B). Data are mean \pm SE for the common (Com) and alternate (Alt) alleles under full irrigation post-flowering. * Indicates a significant difference between alleles, $P < 0.05$.

Discussion

Trait correlations and plasticity of maize RSA

Correlated phenotypes often indicate the presence of co-located QTL with pleiotropic effects (Otto, 2004). While we did identify numerous significant correlations between RSA phenotypes, only the irrigated post-flowering RPF and root area traits shared a candidate gene (*flz35*, Zm00001eb428170). The consistent lack of co-located candidate genes across all the RSA traits suggests that despite being correlated, independent phenotypic variation was present which allowed our analyses to identify candidate genes uniquely associated with these trait's measurable variation. We note however that GWAS models such as FarmCPU are not optimal for explicitly testing whether the same loci are affecting traits across developmental stages, environments or growing seasons and thus may be one reason our GWA efforts did not identify overlap in RSA candidate genes. Schneider et al. (2020) and Zheng et al. (2020), who also used FarmCPU for their GWAS analyses, identified a similar phenomenon of few co-localization of root phenotype candidate genes. The lack of shared candidate genes across years could also be due to differences in the developmental stages at which the plants were phenotyped, and we did observe differences in allelic effect sizes at different developmental stages (Figures 7 and 8 C). Despite this, our analyses focused on identifying GxE across irrigation treatments that could be affecting our candidate gene list. Thus, we implemented post-hoc tests for GxE for all significant GWAS SNPs and identified 12 SNPs that showed significant GxE: 8 for RPF, 2 for shoot mass, and 1 each for root mass and area. Although we identified 48 candidate genes for flowering time, none showed significant GxE which is reflective of this trait's overall lower phenotypic plasticity (Supplemental Figure 18).

Understanding the genetic basis of how crops produce alternative phenotypes in response to heterogeneous environments can provide an immense source for crop improvement through breeding (Kusmec et al., 2018). Through our analyses, we found extensive evidence for root phenotypic plasticity among maize lines across growing seasons. While the majority of genotypes (>50%) exhibited lower RPF under drought conditions, a subset of genotypes displayed consistent increases or decreases in their RPF from irrigated to drought conditions across growing seasons (Figure 14). The consistent pattern of RPF plasticity we find suggests that root plasticity is predictable for some maize genotypes across growing seasons and thus heritable, as also found by Schneider et al. (2020). This finding lends credence to the notion that maize breeding programs can breed for genetics that reliably enhance the adaptive value of RSA to a specific target population of environments.

Additionally, our analyses were also able to identify more fine-scale evidence for phenotypic plasticity at the individual gene level (Figures 7-8, Supplemental Table 14). Interestingly, of the four candidate genes for RPF exhibiting significant GxE that also had functional annotation, two were genes involved in nitrogen signaling (*AMT5*, Zm00001eb427000 and *SLAH2*, Zm00001eb159490). Our finding of nitrogen signaling genes exhibiting GxE in response to drought is supportive of previous reports showing differential expression of homologous nitrogen signaling genes in response to water stress in other plant species (Bielsa et al., 2018; Araus et al., 2020; Filiz and Akbudak, 2020). We hypothesize that the alternate SNP alleles in these nitrogen signaling genes may be linked to additional polymorphisms that differentially affect transcriptional responses of these genes to water availability which underlie their gene level GxE. Furthermore, the evidence for GxE we identified could indicate that these

candidate genes are involved in regulating a synergistic response to nitrogen and water availability, which can affect the capture of both as suggested by Araus et al. (2020).

Candidate genes underlying maize RPF variation

Two recent field-based studies quantified variation, covariation and the genetic architecture underlying variation in maize RSA using GWAS (Schneider et al., 2020; Zheng et al., 2020). Schneider et al. (2020) utilized a diversity panel consisting of different genotypes coupled with differential irrigation treatments at different environments. Their results generate a number of mechanistic hypotheses on the adaptive value of specific root traits in particular target environments. For example, they hypothesize that phenotypic plasticity for root angle is potentially more advantageous for plants in environments that experience prolonged stress such as nitrogen deprivation. Schneider et al. (2020) also present a GWAS analyses which highlighted auxin related genes as being responsible for phenotypic variation in resource rich environments. They also found candidate genes related to cytokinin and phosphorus metabolism that were associated with variation in root plasticity that differed from loci controlling variation in a given environment. Zheng et al. (2020) performed their analyses using the same genotypes used in our study but in a single environment and growing season. They also presented GWAS analysis in FarmCPU that identified a candidate gene encoding a MATE transporter with a known function in shaping maize RSA. Key differences between these prior studies and ours include the testing environments, developmental stages, and method of acquiring root systems (shovelomics versus RPF). Our RPF sampling study was performed in two growing seasons in Colorado which is located in the high plains region of the U.S. which frequently experiences drought and thus serves as an ideal location to study genetics related to drought stress in crops. Additionally, using RPF to acquire root systems as opposed to shovelomics techniques (Trachsel et al., 2011) is advantageous because this method

is less prone to missing data and provides an additional quantitative measure of RSA that is generally more heritable than image-based root traits (Figure 11) and has been used previously to identify QTL for drought responses in maize (Lebreton et al., 1995).

Our GWAS efforts identified a total of 54 SNPs associated with measurable variation in RPF across developmental stages, irrigation treatments, and growing seasons (Supplemental Table 14). Most of these 54 SNPs were located within gene models whose annotation spanned a diversity of functions such as monosaccharide transport (Zm00001eb166700), insect resistance (Zm00001eb273440), and auxin biosynthesis (Zm00001eb060250). Interestingly, our GWA efforts found no overlap of candidate genes across developmental stages, irrigation treatments, or growing seasons for RPF despite their similar allelic effects identified from post-hoc analyses (Figures 7C, 8C). The lack of overlap and quantity of functionally diverse RPF candidate genes we find suggest a highly complex genetic architecture for this trait that is determined by numerous factors including developmental stage, environment, and growing season. We hypothesize that the pathway controlling root development may be conserved, but genotypes may vary in pathways that provide environmental signals and at different times of development to the core development pathway. Future studies should design experiments to mechanistically understand when and why these candidate genes affect RPF dynamics at specific developmental stages and environments.

Our GWAS analysis identified candidate genes related by their: 1) roles in nutrient signaling from roots to shoots and 2) greater expression in root tissue such as *AMT5*, *SLAH2*, and *PHO1;2A*, Zm00001eb191650 (Maierhofer et al., 2014; Salazar-Vidal et al., 2016; Filiz and Akbudak, 2020). This is interesting because despite our experiment being conducted in a field that was fully fertilized, SNPs in genes involved in the transfer of the supplemented nutrients were significantly associated with measurable variation in RPF. As previously indicated by Schneider

et al. 2020, this phenomenon suggests the presence of allelic variation in these candidate genes that significantly affects their ability to respond to soil nutrients. We hypothesize that these candidate genes possess alleles that result in alternate signals of nutrient starvation/supplementation which causes root systems to grow differentially in response – underlying the heritable variation we observed in RPF. Previous studies have shown that in *Arabidopsis thaliana* and maize, root systems grew significantly more in response to nitrogen and phosphorus starved conditions respectively using split pot experiments (Ruffel et al., 2011; Wang et al., 2020). Additionally, *PHO1* has been identified in GWAS for RSA traits in *Arabidopsis thaliana* (Rosas et al. 2013), and *AMT5* has been identified in maize for striga resistance, which is likely root-based (Adewale et al., 2020). Future studies should utilize functional genetics to validate the effect of these candidate genes on maize RPF through both characterization of mutants and evaluation of natural alleles of the loci we identified in this study using near-isogenic lines.

Although *AMT5*, *SLAH2*, and *PHO1;2A* were not identified as candidate genes for RSA traits other than RPF via GWA, our post-hoc analyses with the alternate SNP alleles at these genes identified significant differences in other traits such as root area, depth, and width that were consistent with their differences in RPF (Figures 7-8, 10). This suggests that despite the independent variation among the RSA traits we measured, RPF exhibits a strong enough association with other RSA traits to identify candidate genes that significantly affect their variation as well. This connection between RPF and other RSA traits potentially helps to ameliorate the difficulty of studying the genetic basis of RSA in maize. RPF provides a relatively high throughput method for quantifying RSA and discovering candidate genes that affect multiple RSA traits in mature, field grown plants.

Conclusion

Understanding the phenotypic plasticity and genetic control that shapes RSA in maize remains a formidable challenge. This is because of the difficulty of obtaining root phenotype data across development for mature field grown plants. To evaluate the phenotypic plasticity and genetic control of RSA in maize, we used a maize diversity panel coupled with environments of contrasting irrigation levels as well as publicly available genotype matrices from next generation sequencing data. Our results implicate that: 1) root phenotypic plasticity is predictable for some maize genotypes, 2) RSA in maize is a highly complex trait controlled by many functionally diverse genes, and 3) RPF is an efficient phenotype capable of identifying candidate genes associated with variation in additional root architectural traits. Future studies using functional genetic techniques such as mutant screens and QTL mapping in RIL populations are needed to validate the candidate genes identified in this study and accurately quantify effect sizes of alleles across developmental stages and environments. Additionally, there is a need to develop field and analysis designs that improve our ability directly test for QTLx E in statistically robust genome-wide analysis. Overall, the results discussed here extend our knowledge of root phenotypic plasticity in maize at the whole-plant, and gene levels and further elucidate the highly complex genetic architecture controlling variation in maize RSA.

Data Availability

Data are available at: <https://doi.org/10.3389/fpls.2022.883209>.

REFERENCES

- Adewale, S.A., Badu-Apraku, B., Akinwale, R.O., Paterne, A.A., Gedil, M., Garcia-Oliveira, A.L. (2020). Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol* 20, 203. Doi: 10.1186/s12870-020-02360-0
- Araus, V., Swift, J., Alvarez, J.M., Henry, A., and Coruzzi, G.M. (2020). A balancing act: how plants integrate nitrogen and water signals. *Journal of Experimental Botany* 71: 4442-4451.
- Bao, Y., Aggarwal, P., Robbins, N.E., Sturrock, C.J., Thompson, M.C., Tan, H.Q., et al. (2014). Plant roots use a patterning mechanism to position lateral root branches toward available water. *Proceedings of the National Academy of Sciences of the United States of America*, 111: 9319–9324.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw* 67: 1.
- Bielsa, B., Hewitt, S., Reyes-Chin-Wo, S., Dhingra, A., and Rubio-Cabetas, M.J. (2018). Identification of water use efficiency related genes in “Garnem” almond-peach rootstock using time-course transcriptome analysis. *PloS ONE* 13(10): e0205493. Doi: 10.1371/journal.pone.0205493
- Das, A., Schneider, H., Burridge, J., Martinez Ascanio, A.K., Wojchiechowski, T., Topp, C.N., et al. (2015). Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant Methods* 11: 51. Doi: 10.1186/s13007-015-0093-3
- Dechorgnat, J., Francis, K.L., Dhugga, K.S., Rafalski, J.A., Tyerman, S.D., and Kaiser, B.N. (2019). Tissue and nitrogen-linked expression profiles of ammonium and nitrate transporters in maize. *BMC Plant Biology* 19: 206.
- de Jong, M., Tavares, H., Pasam, R.K., Butler, R., Ward, S., George, G. et al. (2019) Natural variation in *Arabidopsis* shoot branching plasticity in response to nitrate supply affects fitness. *PLoS Genetics* (Vol. 15). Doi: 10.1371/journal.pgen.1008366
- El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., and Aarts, M.G.M. (2014). Genotype x environment interaction QTL mapping in plants: lessons from *Arabidopsis*. *Trends in Plant Science*. 19: 390-398. doi: 10.1016/j.tplants.2014.01.001
- Eshel, A., and Beeckman, T. (2013). *Plant roots: the hidden half*. CRC press.
- Falconer, D. S. (1990). Selection in different environments: Effects on environmental sensitivity (reaction norm) and on mean performance. *Genetical Research* 56: 57–70. doi: 10.1017/S0016672300028883.

- Filiz, E., and Akbudak, M.A. (2020). Ammonium transporter 1 (AMT1) gene family in tomato (*Solanum lycopersicum* L.): Bioinformatics, physiological and expression analyses under drought and salt stresses. *Genomics* 112: 3773-3782. doi: 10.1016/j.ygeno.2020.04.009
- Fletcher, R.S., Mullen, J.L., Heiliger, A., and McKay, J.K. (2015). QTL analysis of root morphology, flowering time and yield reveals trade-offs in response to drought in *Brassica napus*. *J Exp Bot* 66: 245-256.
- Fox, J., and Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.
- Gaudin, A.C.M., McClymont, S.A., Holmes, B.M., Lyons, E., and Raizada, M.N. (2011). Novel temporal, fine-scale and growth variation phenotypes in roots of adult-stage maize (*Zea mays* L.) in response to low nitrogen stress. *Plant, Cell and Environment*, 34(12), 2122–2137. doi: 10.1111/j.1365-3040.2011.02409.x
- Hamburger, D., Rezzonico, E., MacDonald-Comber Petétot, J., Somerville, C., and Poirier, Y. (2002). Identification and characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. *Plant Cell* 14: 889–902. doi: 10.1105/tpc.000745
- Johnson, S.N., and Rasmann, S. (2015). Root-feeding insects and their interactions with organisms in the rhizosphere. *Annual Review of Entomology*. 60: 517–535. doi: 10.1146/annurev-ento-010814-020608
- Karlova, R., Boer, D., Hayes, S., and Testerink, C. (2021). Root plasticity under abiotic stress. *Plant Physiology: kiab392*. doi: 10.1093/plphys/kiab392
- Kusmec, A., Srinivasan, S., Nettleton, D., and Schnable, P.S. (2017). Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nature Plants* 3, 715–723. Doi: 10.1038/s41477-017-0007-7
- Kusmec, A., de Leon, N. and Schnable, P. S. (2018). Harnessing phenotypic plasticity to improve maize yields. *Frontiers in Plant Science*. 9: 1–4. doi: 10.3389/fpls.2018.01377
- Lareen, A., Burton, F. and Schäfer, P. (2016). Plant root-microbe communication in shaping root microbiomes. *Plant Molecular Biology*. 90: 575–587. doi: 10.1007/s11103-015-0417-8
- Lebreton, C., Lazic-Jancic, V., Steed, A., Pekic, S., and Quarrie, S.A. (1995). Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. *Journal of Experimental Botany* 46: 853-865.
- Leiboff, S., Li, X., Hu, H.C., Todt, N., Yang, J., Li, X. et al. (2015). Genetic control of morphometric diversity in the maize shoot apical meristem. *Nature Communications* 6: 8974.
- Liao, H., Rubio, G., Yan, X., Cao, A., Brown, K.M., and Lynch, J.P. (2001). Effect of phosphorus availability on basal root shallowness in common bean. *Plant Soil* 23: 269-279.

- Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J. et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. Doi: 10.1093/bioinformatics/bts444
- Liu, X., Huang, M., Fan, B., Buckler, E.S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genetics*. 12: e1005767. Doi: 10.1371/journal.pgen.1005767
- Lowry, D.B., Lovell, J.T., Zhang, L., Bonnette, J., Fay, P.A., Mitchell, R.B. et al. (2019). QTL \times environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proceedings of the National Academy of Sciences of the United States of America*. 116: 12933–12941. doi: 10.1073/pnas.1821543116.
- Lynch, M. and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Assocs, Inc., Sunderland, MA.
- Maierhofer, T., Lind, C., Hüttl, S., Scherzer, S., Papenfuß, M., Simon, J., et al. (2014). A single-pore residue renders the Arabidopsis root anion channel SLAH2 highly nitrate selective. *Plant Cell*. 26: 2554–2567. Doi: 10.1105/tpc.114.125849
- Nicotra, A. B. and Davidson, A. (2010). Adaptive phenotypic plasticity and plant water use. *Functional Plant Biology*. 37: 117–127. doi: 10.1071/FP09139.
- Otto, S. P. (2004). Two steps forward, one step back: The pleiotropic effects of favoured alleles. *Proceedings of the Royal Society B: Biological Sciences*. 271: 705–714. doi: 10.1098/rspb.2003.2635
- Paterson, A.H., Saranga, Y., Menz, M., Jiang, C.-X., and Wright, R.J. (2003). QTL analysis of genotype \times environment interactions affecting cotton fiber quality. *Theoretical and Applied Genetics*. 106: 384–396. doi: 10.1007/s00122-002-1025-y.
- Petricka, J. J., Winter, C. M. and Benfey, P. N. (2012). Control of Arabidopsis root development. *Annual Review of Plant Biology*. 63: 563–590. doi: 10.1146/annurev-arplant-042811-105501.
- Raven, J. A. and Edwards, D. (2001). Roots: Evolutionary origins and biogeochemical significance. *Journal of Experimental Botany*. 52: 381–401. doi: 10.1093/jxb/52.suppl_1.381
- Rogers, E.D., Monaenkova, D., Mijar, M., Nori, A., Goldman, D.I., and Benfey, P.N. (2016). X-ray computed tomography reveals the response of root system architecture to soil texture. *Plant Physiology*. 171: 2028–2040. doi: 10.1104/pp.16.00397.
- Rosas, U., Cibrian-Jaramillo, A., Ristova, D., Banta, J.A., Gifford, M.L., Fan, A.H. et al. (2013). Integration of responses within and across Arabidopsis natural accessions uncovers loci controlling root systems architecture. *Proceedings of the National Academy of Sciences of the United States of America*. 110:15133–15138. doi: 10.1073/pnas.1305883110

- Ruffel, S., Krouk, G., Ristova, D., Shasha, D., Birnbaum, K.D., and Coruzzi, G.M. (2011). Nitrogen economics of root foraging: Transitive closure of the nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand. *Proceedings of the National Academy of Sciences of the United States of America*. 108: 18524-18529. Doi: 10.1073/pnas.1108684108
- Salazar-Vidal, M.N., Acosta-Segovia, E., Sanchez-León, N., Ahern, K.R., Brutnell, T.P., and Sawers, R.J.H. (2016). Characterization and transposon mutagenesis of the maize (*Zea mays*) Pho1 Gene Family. *PloS ONE*. 11: 1–19. doi: 10.1371/journal.pone.0161882.
- Schneider, H.M., Klein, S.P., Hanlon, M.T., Nord, E.A., Kaeppler, S., Brown, K.M., et al. (2020). Genetic control of root architectural plasticity in maize. *Journal of Experimental Botany* 71: 3185-3197. doi: 10.1093/jxb/eraa084.
- Schneider, H.M. and Lynch, J.P. (2020). Should Root Plasticity Be a Crop Breeding Target? *Frontiers in Plant Science*. 11: 1–16. doi: 10.3389/fpls.2020.00546.
- Sebastian, J., Yee, M.C., Viana, W.G., Rellán-Álvarez, R., Feldman, M., Priest, H.D. et al. (2016). Grasses suppress shoot-borne roots to conserve water during drought. *Proceedings of the National Academy of Sciences of the United States of America*. 113: 8861–8866. Doi: 10.1073/pnas.1604021113
- Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N. et al. (2011). Genome-wide atlas of transcription during maize development. *Plant Journal* 66: 553-563. Doi: 10.1111/j.1365-313X.2011.04527.x
- Shao, M.R., Jiang, N., Li, M., Howard, A., Lehner, K., Mullen, J.L., et al. (2021). Complementary phenotyping of maize root system architecture by root pulling force and X-ray imaging. *Plant Phenomics*. 9859254. doi: 10.34133/2021/9859254
- Smith, S., and de Smet, I. (2012). Root system architecture: Insights from Arabidopsis and cereal crops. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367: 1441–1452. Doi: 10.1098/rstb.2011.0234
- Strock, C.F., De La Riva, L.M. and Lynch, J.P. (2018). Reduction in root secondary growth as a strategy for phosphorus acquisition. *Plant Physiology*. 176: 691–703. doi: 10.1104/pp.17.01583
- Trachsel, S., Kaeppler, S.M., Brown, K.M., and Lynch, J.P. (2011). Shovelomics: High throughput phenotyping of maize (*Zea mays* L.) root architecture in the field. *Plant Soil* 341: 75–87. doi: 10.1007/s11104-010-0623-8
- Tracy, S.R., Nagel, K.A., Postma, J.A., Fassbender, H., Wasson, A., and Watt, M. (2020). Crop improvement from phenotyping roots: highlights reveal expanding opportunities. *Trends in Plant Science*. 25: 105-118. Doi: [10.1016/j.tplants.2019.10.015](https://doi.org/10.1016/j.tplants.2019.10.015)
- Voss-Fels, K.P., Snowdon, R.J. and Hickey, L.T. (2018). Designer Roots for Future Crops. *Trends in Plant Science*. 23: 957–960. doi: 10.1016/j.tplants.2018.08.004.

Wang, X., Feng, J., White, P.J., Shen, J., and Cheng, L. (2020). Heterogeneous phosphate supply influences maize lateral root proliferation by regulating auxin redistribution. *Annals of Botany*. 125: 119–130. doi: 10.1093/aob/mcz154.

White, J.W. and Castillo, J.A. (1992). Evaluation of Diverse Shoot Genotypes on selected Root Genotypes of Common Bean under Soil Water Deficits. *Crop Science*. 32: 762–765. doi: 10.2135/cropsci1992.0011183x003200030037x.

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., McMullen, M.D., and Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PloS One*: 4(12), e8451.

Zheng, X., He, K., Kleist, T., Chen, F., and Luan, S. (2015). Anion channel SLAH3 functions in nitrate-dependent alleviation of ammonium toxicity in *Arabidopsis*. *Plant Cell & Environment* 38: 474-486 doi: 10.1111/pce.12389.

Zheng, Z., Hey, S., Jubery, T., Liu, H., Yang, Y., Coffey, L., et al. (2020). Shared genetic control of root system architecture between *Zea mays* and *Sorghum bicolor*. *Plant Physiology* 182: 977-991. doi: 10.1104/pp.19.00752.

APPENDIX I

SUPPLEMENTAL INFORMATION FOR CHAPTER 2: QUANTITATIVE TRAIT LOCI CONTROLLING AGRONOMIC AND BIOCHEMICAL TRAITS IN *CANNABIS SATIVA*

Supplementary Table 1. Summary table of parent, F₁ and F₂ phenotype values.

Supplementary Table 1						
Trait	Mean USO31	Mean Carmagnola	Mean F ₁	Mean F ₂	Minimum F ₂	Maximum F ₂
Days to Maturity	89	118	95	100	82	128
Seed 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
Dry 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
CBDa (%)	0.205	1.917	0.118	0.164	0	1.890
CBGa (%)	0.008	0.037	0.003	0.024	0	0.550
CBC (%)	0.0012	0.0031	0.00	0.0003	0.00	0.0080
Alpha-Pinene (ppm)	28.65	413.34	574.13	84.92	2.6	1081.86
Beta-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
Citronellol (ppm)	0	0.00	0.00	0.679	0	14.85
Alpha-Terpinene (ppm)	0.84	8.25	1.30	2.41	0	51.12
Gamma-Terpinene (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
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
Linalyl Acetate (ppm)	1.96	3.90	1.655	1.29	0	10.24
3-Carene (ppm)	0	3.90	0	0.58	0	22.35
Linalool (ppm)	2.36	12.35	0	0.57	0	23.75

Supplementary Table 2. F₂ phenotype correlations and their significance values.

Trait	CBC	THCa	CBDa	CBGa	3_carene	Alpha-pinene	Alpha-terpinene	Beta-caryophyllene	Camphene	Caryophyllene oxide	Citronellol	Eucalyptol	Geraniol	Gamma-terpinene	Linalool	Linalyl acetate	Ocimene_1	Days to maturity	Leaf water content	Plant height	Thousand seed mass	Stem diameter	Stem biomass	Seed yield	Dry biomass					
CBC	1.000																													
p_value	0.000																													
THCa	0.470	1.000																												
p_value	0.000	0.000																												
CBDa	0.420	0.850	1.000																											
p_value	0.000	0.000	0.000																											
CBGa	0.330	0.290	0.200	1.000																										
p_value	0.000	0.000	0.003	0.000																										
3_carene	0.065	0.410	0.390	0.130	1.000																									
p_value	NS	0.000	0.000	0.049	0.000																									
Alpha-pinene	0.270	0.620	0.720	0.230	0.440	1.000																								
p_value	0.000	0.000	0.000	0.001	0.000	0.000																								
Alpha-terpinene	0.330	0.620	0.690	0.250	0.460	0.680	1.000																							
p_value	0.000	0.000	0.000	0.000	0.000	0.000	0.000																							
Beta-caryophyllene	0.410	0.540	0.630	0.230	0.280	0.730	0.570	1.000																						
p_value	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000																						
Camphene	0.240	0.470	0.520	0.170	0.440	0.670	0.670	0.340	1.000																					
p_value	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000																					
Caryophyllene oxide	-0.180	-0.160	-0.190	-0.220	-0.036	0.088	-0.150	0.051	-0.059	1.000																				
p_value	0.007	0.022	0.004	0.001	NS	NS	0.028	NS	NS	0.000																				
Citronellol	-0.150	-0.230	-0.260	-0.250	-0.110	-0.054	-0.200	-0.330	0.077	0.390	1.000																			
p_value	0.034	0.001	0.000	0.000	NS	NS	0.003	0.000	NS	0.000	0.000																			
Eucalyptol	0.056	0.330	0.300	0.250	0.540	0.400	0.370	0.410	0.310	-0.180	1.000																			
p_value	NS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NS	0.008	0.000																		
Geraniol	-0.230	-0.370	-0.450	-0.370	-0.160	-0.330	-0.310	-0.520	-0.010	0.310	0.670	-0.220	1.000																	
p_value	0.001	0.000	0.000	0.000	0.021	0.000	0.000	0.000	NS	0.000	0.000	0.001	0.000																	
Gamma-terpinene	0.350	0.630	0.700	0.310	0.430	0.740	0.900	0.630	0.660	-0.170	-0.230	0.440	-0.350	1.000																
p_value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.001	0.000	0.000	0.000																
Linalool	0.110	0.120	0.130	0.014	0.250	0.260	0.190	0.047	0.290	0.170	0.290	0.160	0.200	0.170	1.000															
p_value	NS	NS	NS	NS	0.000	0.000	0.005	NS	0.000	0.014	0.000	0.018	0.003	0.013	0.000															
Linalyl acetate	0.190	0.260	0.200	0.094	0.310	0.049	0.270	-0.019	0.380	-0.072	0.120	0.170	0.100	0.210	0.220	1.000														
p_value	0.006	0.000	0.003	NS	0.000	NS	0.000	NS	NS	0.010	NS	NS	0.002	0.001	0.000															
Ocimene_1	0.290	0.430	0.470	0.110	0.280	0.450	0.610	0.490	0.470	-0.043	-0.100	0.110	-0.190	0.610	0.190	0.260	1.000													
p_value	0.000	0.000	0.000	NS	0.000	0.000	0.000	0.000	0.000	NS	NS	NS	0.006	0.000	0.005	0.000	0.000													
Days to maturity	-0.230	-0.140	-0.140	-0.058	0.034	-0.057	-0.091	-0.150	0.026	-0.072	-0.170	-0.097	-0.017	-0.110	-0.140	-0.110	-0.073	1.000												
p_value	0.001	0.043	0.040	NS	NS	NS	NS	0.025	NS	0.013	NS	NS	NS	NS	NS	NS	NS	0.000												
Leaf water content	-0.130	0.079	0.130	0.100	0.075	0.160	0.140	0.120	0.130	0.046	-0.066	0.095	-0.085	0.150	0.068	0.032	0.120	0.400	1.000											
p_value	NS	NS	NS	NS	NS	0.016	0.042	NS	NS	NS	NS	NS	NS	0.030	NS	NS	NS	0.000	0.000											
Plant height	-0.160	0.079	0.110	0.065	0.088	0.110	0.091	0.013	0.120	0.079	0.032	-0.015	0.000	0.091	0.094	0.073	0.150	0.310	0.690	1.000										
p_value	0.025	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.035	0.000	0.000	0.000										
Thousand seed mass	-0.074	0.031	-0.020	0.028	0.097	0.093	0.160	0.057	0.140	0.073	0.010	0.150	-0.050	0.140	0.042	0.110	0.140	0.200	0.340	0.320	1.000									
p_value	NS	NS	NS	NS	NS	NS	0.016	NS	0.038	NS	NS	0.029	NS	0.046	NS	NS	0.036	0.001	0.000	0.000	0.000									
Stem diameter	-0.170	0.015	0.039	0.130	0.088	0.140	0.083	0.012	0.130	-0.007	-0.058	0.034	-0.091	0.092	0.015	-0.035	0.083	0.480	0.630	0.800	0.340	1.000								
p_value	0.009	NS	NS	0.050	NS	0.047	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.000	0.000	0.000	0.000	0.000								
Stem biomass	-0.170	-0.009	0.006	0.078	0.021	0.045	0.015	-0.044	0.065	-0.011	-0.069	-0.056	-0.087	0.018	-0.019	0.004	0.072	0.540	0.600	0.840	0.230	0.570	1.000							
p_value	0.009	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.000	0.000	0.000	0.000	0.000	0.000	1.000						
Seed yield	-0.150	0.048	0.050	0.110	0.069	0.089	0.120	0.003	0.130	0.075	0.023	0.011	-0.032	0.096	0.071	0.053	0.190	0.290	0.420	0.640	0.460	0.650	0.740	1.000						
p_value	0.028	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
Dry biomass	-0.180	-0.012	-0.030	0.070	0.033	0.017	0.028	-0.061	0.052	0.015	-0.016	-0.060	-0.040	0.009	-0.008	-0.015	0.085	0.470	0.540	0.800	0.380	0.840	0.950	0.830	1.000					
p_value	0.007	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				

Spearman rank correlation coefficients and p-values reported have been rounded to three decimal places. Non-significant p-values above 0.05 are reported as "NS".

Supplementary Table 3. Parent allele effects on phenotype values at each identified QTL for agronomic traits.

Supplementary Table 3

Phenotype	QTL	Linkage Group	Carmagnola Effect	USO31 Effect
Leaf Water Content (g)	LWC.1	3	+	-
	LWC.2	5	+	-
	LWC.3	10	+	-
Plant Height (cm)	PLHT.1	3	+	-
	PLHT.2	5	+	-
	PLHT.3	10	+	-
Thousand Seed Mass (g)	TSM.1	3	+	-
	TSM.2	5	-	-
	TSM.3	8	+	-
	TSM.4	10	+	-
Stem Diameter (cm)	SD.1	3	+	-
	SD.2	4	+	-
	SD.3	5	+	-
	SD.4	5	+	-
	SD.5	9	+	-
	SD.6	10	+	-
Stem Biomass (g)	SB.1	3	+	-
	SB.2	4	+	-
	SB.3	5	+	-
	SB.4	5	+	-
	SB.5	10	+	-
Seed Yield (g)	SY.1	2	-	+
	SY.2	3	+	-
	SY.3	4	+	-
	SY.4	5	+	-
	SY.5	5	+	-
Dry Biomass (g)	DB.1	3	+	-
	DB.2	4	+	-
	DB.3	5	+	-
	DB.4	9	+	-
Days to Maturity	DTM.1	3	+	-
	DTM.2	4	+	-
	DTM.3	5	+	-
	DTM.4	9	-	+

Effects listed as "+" indicate F₂ plants homozygous for this allele exhibited greater trait values.

Effects listed as "-" indicate F₂ plants homozygous for this allele exhibited lesser trait values.

Supplementary Table 4. Summary information regarding the percent variance explained by each identified QTL cluster.

Supplementary Table 4

QTL Cluster	Variance Minimum (%)	Variance Maximum (%)
LG3.60	5.22	22.35
LG4.50	3.91	5.92
LG5.05	12.92	34.27
LG10.25	3.81	5.65
LG6.35	9.35	28.21
LG9.40	10.23	44.79

Supplementary Table 5. Parent allele effects on phenotype values at each identified QTL for biochemical traits.

Supplementary Table 5				
Phenotype	QTL	Linkage Group	Carmagnola Effect	USO31 Effect
Alpha Pinene (ppm)	AP.1	1	+	-
	AP.2	2	+	-
	AP.3	6	+	-
	AP.4	7	+	-
	AP.5	8	+	-
	AP.6	9	+	-
Alpha Terpinene (ppm)	AT.1	6	+	-
	AT.2	9	+	-
Beta Caryophyllene (ppm)	BC.1	2	+	-
	BC.2	8	+	-
	BC.3	9	+	-
Camphene (ppm)	CAM.1	6	+	-
	CAM.2	9	+	-
Caryophyllene Oxide (ppm)	CO.1	9	-	+
	CO.2	10	-	+
CBC (%)	CBC.1	9	+	-
CBDa (%)	CBDa.1	6	+	-
	CBDa.2	9	+	-
CBGa (%)	CBGa.1	6	-	+
	CBGa.2	9	+	-
Citronellol (ppm)	CIT.1	9	-	+
Eucalyptol (ppm)	EUC.1	5	+	-
	EUC.2	9	+	-
Gamma Terpinene (ppm)	GT.1	6	+	-
	GT.2	9	+	-
Geraniol (ppm)	GE.1	9	-	+
Linalool (ppm)	LI.1	7	+	-
Linalyl Acetate (ppm)	LA.1	6	+	-
Ocimene 1 (ppm)	OC.1	1	-	+
	OC.2	6	+	-
	OC.3	9	+	-
THCa (%)	THCa.1	6	+	-
	THCa.2	9	+	-
3 Carene (ppm)	3C.1	6	+	-
	3C.2	9	+	-

Effects listed as "+" indicate F₂ plants homozygous for this allele exhibited greater trait values.

Effects listed as "-" indicate F₂ plants homozygous for this allele exhibited lesser trait values.

Supplementary Table 6. ANOVA table showing results for tests of mean differences among F₂ agronomic traits at the non-synonymous SNP within the predicted TINY coding sequence.

Supplementary Table 6

Phenotype	Sum of Squares	Mean Square	F	P-value
DTM	62.48	31.24	42.72	< 2.2e-16
Stem Biomass	51.97	25.99	32.18	2.62E-13
Stem Diameter	39.96	19.98	23.63	3.28E-10
Dry Biomass	40.17	20.09	23.56	3.48E-10
TSM	36.59	18.30	20.95	3.25E-09
Seed Yield	31.00	15.50	17.56	6.52E-08
LWC	23.30	11.65	12.84	4.61E-06
PLHT	18.64	9.32	10.10	5.97E-05

Supplementary Table 7: Table displaying the corresponding chromosomes from the Finola, CS_10 and Purple Kush *C. sativa* assemblies.

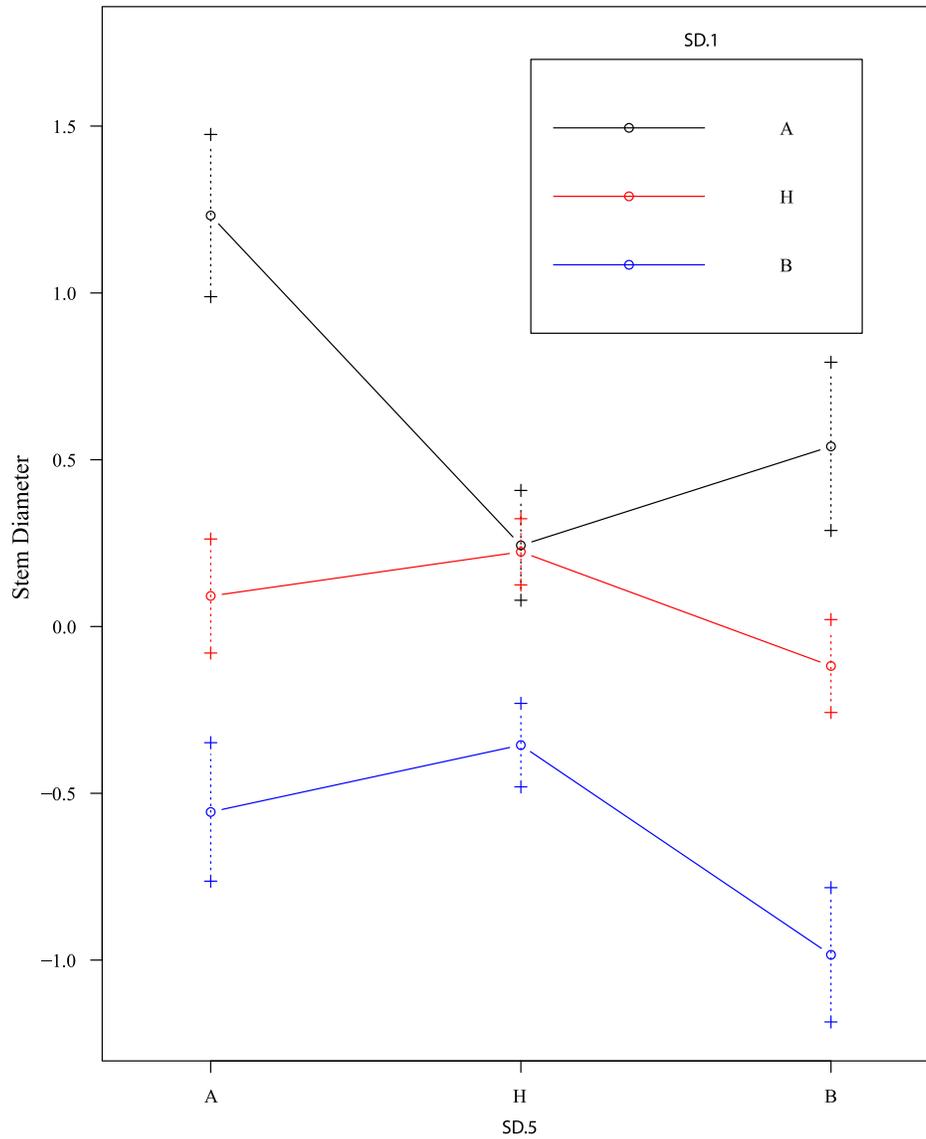
	Finola	CS_10	Purple Kush
Chromosome 1	CM011605.1	NC_044374.1	CM010790.1
Chromosome 2	CM011606.1	NC_044377.1	CM010792.1
Chromosome 3	CM011607.1	NC_044375.1	CM010793.1
Chromosome 4	CM011608.1	NC_044372.1	CM010794.1
Chromosome 5	CM011609.1	NC_044371.1	CM010796.1
Chromosome 6	CM011610.1	NC_044378.1	CM010797.1
Chromosome 7	CM011611.1	NC_044373.1	CM010799.1
Chromosome 8	CM011612.1	NC_044376.1	CM010795.1
Chromosome 9	CM011613.1	NC_044379.1	CM010798.1
Chromosome 10	CM011614.1	NC_044370.1	CM010791.1

Supplementary Table 8: Additive and dominance effects for individual agronomic trait QTL.

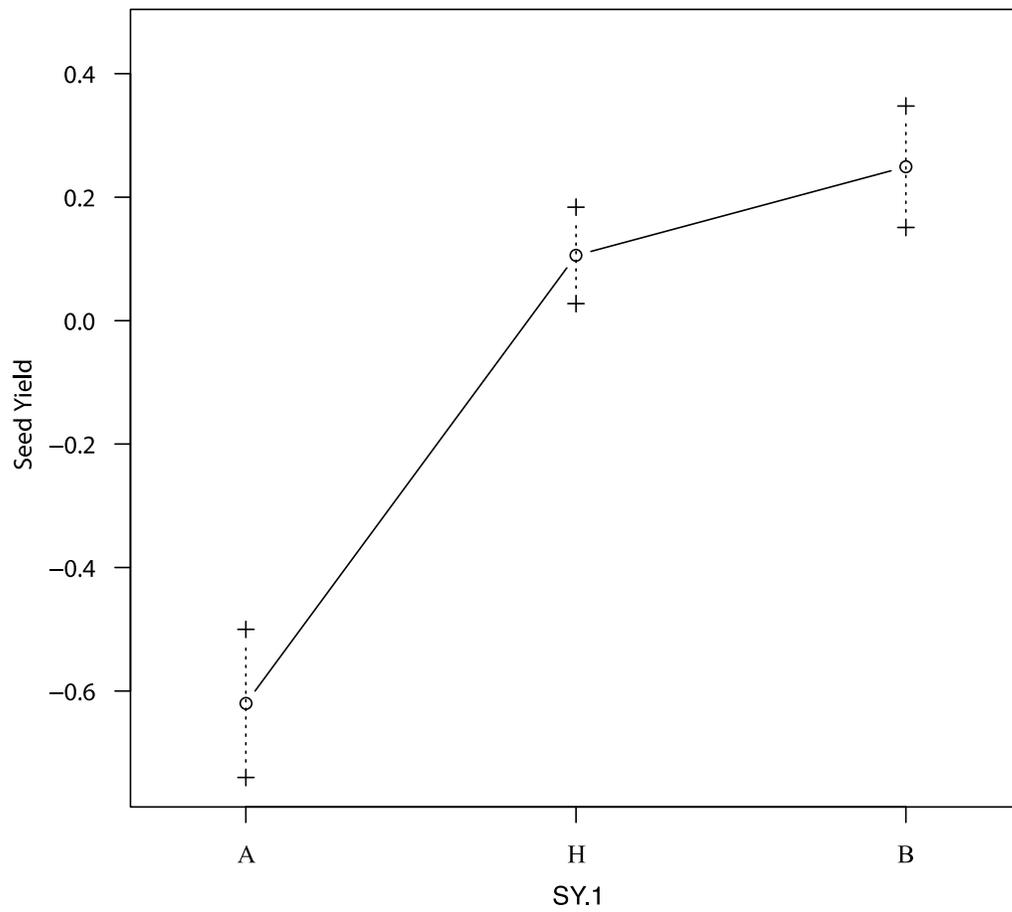
Supplementary Table 8. Additive and Dominance Effects for agronomic QTL			
Phenotype	QTL	Additive Effects	Dominance Effects
Leaf Water Content (g)	LWC.1	-0.567	-0.095
	LWC.2	-0.527	0.432
	LWC.3	-0.304	0.076
Plant Height (cm)	PLHT.1	-0.646	0.124
	PLHT.2	-0.702	0.464
	PLHT.3	-0.298	0.111
Thousand Seed Mass (g)	TSM.1	-0.353	0.271
	TSM.2	-0.202	0.886
	TSM.3	-0.381	-0.058
	TSM.4	-0.307	0.249
Stem Diameter (cm)	SD.1	-0.458	0.161
	SD.2	-0.318	0.070
	SD.3	-0.747	0.336
	SD.4	-0.268	-0.064
	SD.5	-0.170	0.028
	SD.6	-0.286	0.085
Stem Biomass (g)	SB.1	-0.395	0.162
	SB.2	-0.331	0.089
	SB.3	-0.875	0.427
	SB.4	-0.269	-0.136
	SB.5	-0.238	0.362
Seed Yield (g)	SY.1	0.470	0.234
	SY.2	-0.186	0.371
	SY.3	-0.300	0.025
	SY.4	-0.589	0.500
	SY.5	-0.244	0.326
Dry Biomass (g)	DB.1	-0.344	0.235
	DB.2	-0.285	-0.017
	DB.3	-0.736	0.518
	DB.4	-0.264	0.202
Days to Maturity	DTM.1	-0.325	-0.003
	DTM.2	-0.316	0.046
	DTM.3	-0.937	0.271
	DTM.4	0.202	-0.287

Supplementary Table 9: Additive and dominance effects for individual biochemical trait QTL.

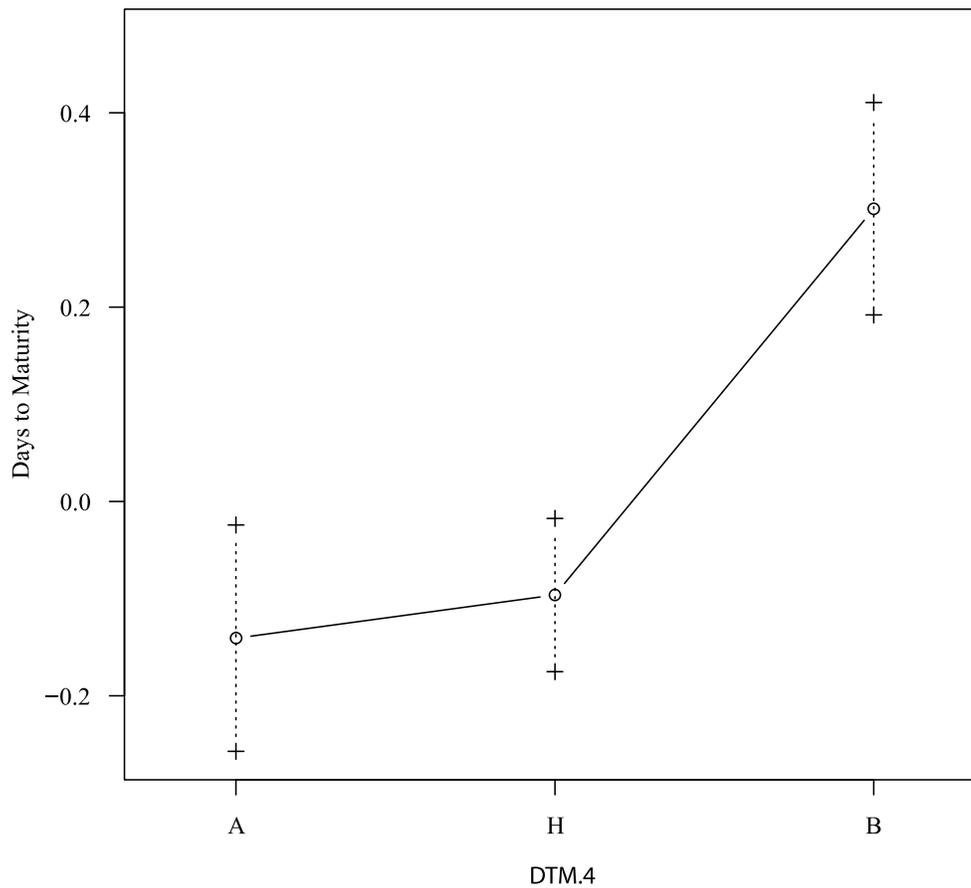
Supplementary Table 9. Additive and Dominance effects for biochemical QTL.			
Phenotype	QTL	Additive Effects	Dominance Effects
Alpha Pinene (ppm)	AP.1	-0.2384	0.3029
	AP.2	-0.2978	0.0202
	AP.3	-0.4272	0.0316
	AP.4	-0.2659	0.0953
	AP.5	-0.1364	-0.0013
	AP.6	-0.5620	0.5423
Alpha Terpinene (ppm)	AT.1	-0.4146	-0.0036
	AT.2	-0.4695	0.2159
Beta Caryophyllene (ppm)	BC.1	-0.3936	-0.2409
	BC.2	-0.3371	0.0240
	BC.3	-0.6755	0.8586
Camphene (ppm)	CAM.1	-0.4168	-0.0721
	CAM.2	-0.3976	-0.0673
Caryophyllene Oxide (ppm)	CO.1	0.3827	0.4519
	CO.2	0.3054	0.0340
CBC (%)	CBC.1	-0.3577	-0.0364
CBDa (%)	CBDa.1	-0.6737	0.3639
	CBDa.2	-0.7774	0.1308
CBGa (%)	CBGa.1	0.5558	-0.4430
	CBGa.2	-0.7809	-0.1202
Citronellol (ppm)	CIT.1	0.3664	-0.2890
Eucalyptol (ppm)	EUC.1	-0.2647	-0.1846
	EUC.2	-0.3578	0.0841
Gamma Terpinene (ppm)	GT.1	-0.3574	0.0271
	GT.2	-0.5591	0.2149
Geraniol (ppm)	GE.1	0.5668	-0.5478
Linalool (ppm)	LI.1	-0.3182	-0.1643
Linalyl Acetate (ppm)	LA.1	-0.3507	-0.0708
Ocimene 1 (ppm)	OC.1	0.4145	0.2130
	OC.2	-0.3399	-0.0034
	OC.3	-0.2886	0.3552
THCa (%)	THCa.1	-0.5228	0.3961
	THCa.2	-0.6045	0.0822
3 Carene (ppm)	3C.1	-0.2877	-0.2009
	3C.2	-0.2651	-0.0886



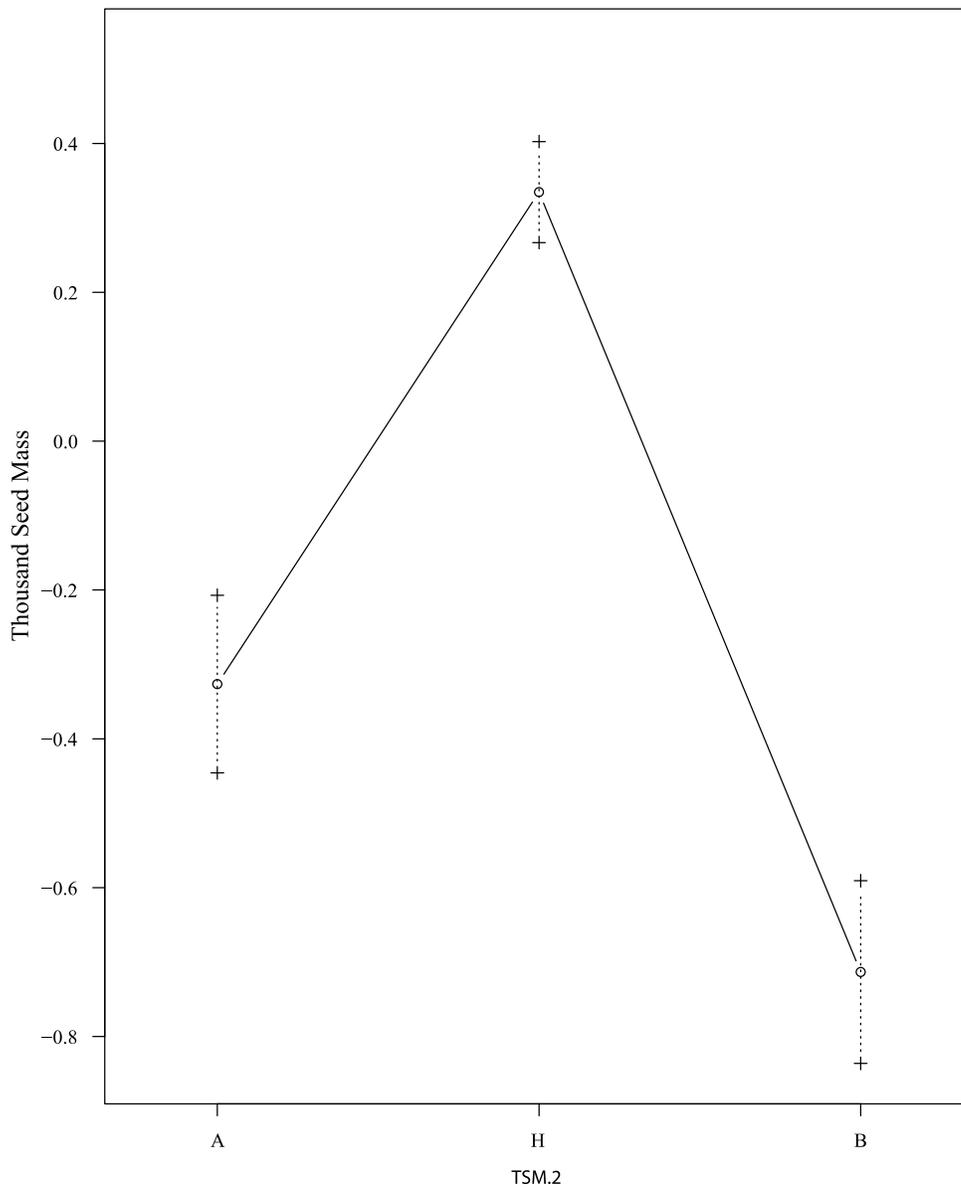
Supplementary Figure 1. Epistatic interaction of the stem diameter QTL SD.1 and SD.5. Line colors indicate F₂ plant genotype at SD.1 while x-axis positions indicate F₂ plant genotype at SD.5. Y-axis indicates the mean (\pm standard error) quantile normalized stem diameter quantities. Lesser y-axis values indicate smaller stem diameters while higher y-axis values indicate larger stem diameters.



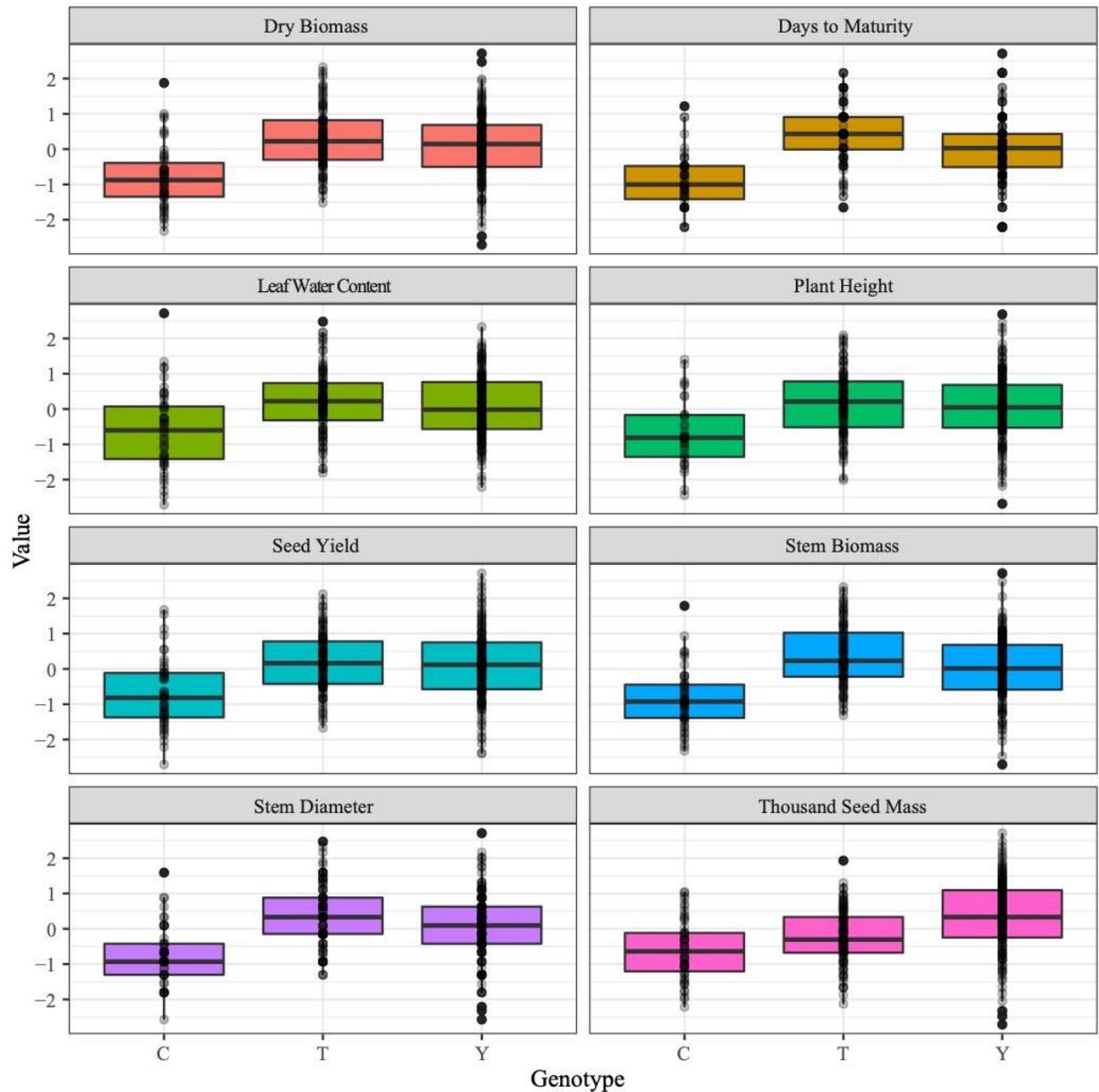
Supplementary Figure 2: Shown are the F₂ quantile normalized phenotype values (\pm standard error) across each F₂ plant genotype. Lesser y-axis values indicate lower seed mass while higher y-axis values indicate greater seed yield.



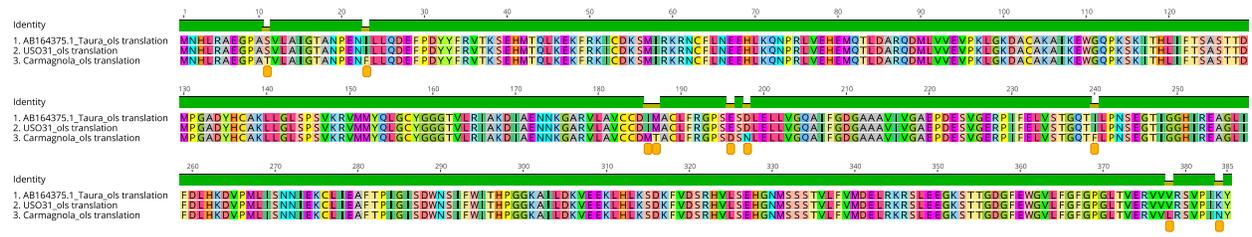
Supplementary Figure 3: Shown are the F₂ quantile normalized phenotype values (\pm standard error) across each F₂ plant genotype. Lesser y-axis values indicate lower seed mass while higher y-axis values indicate greater days to maturity.



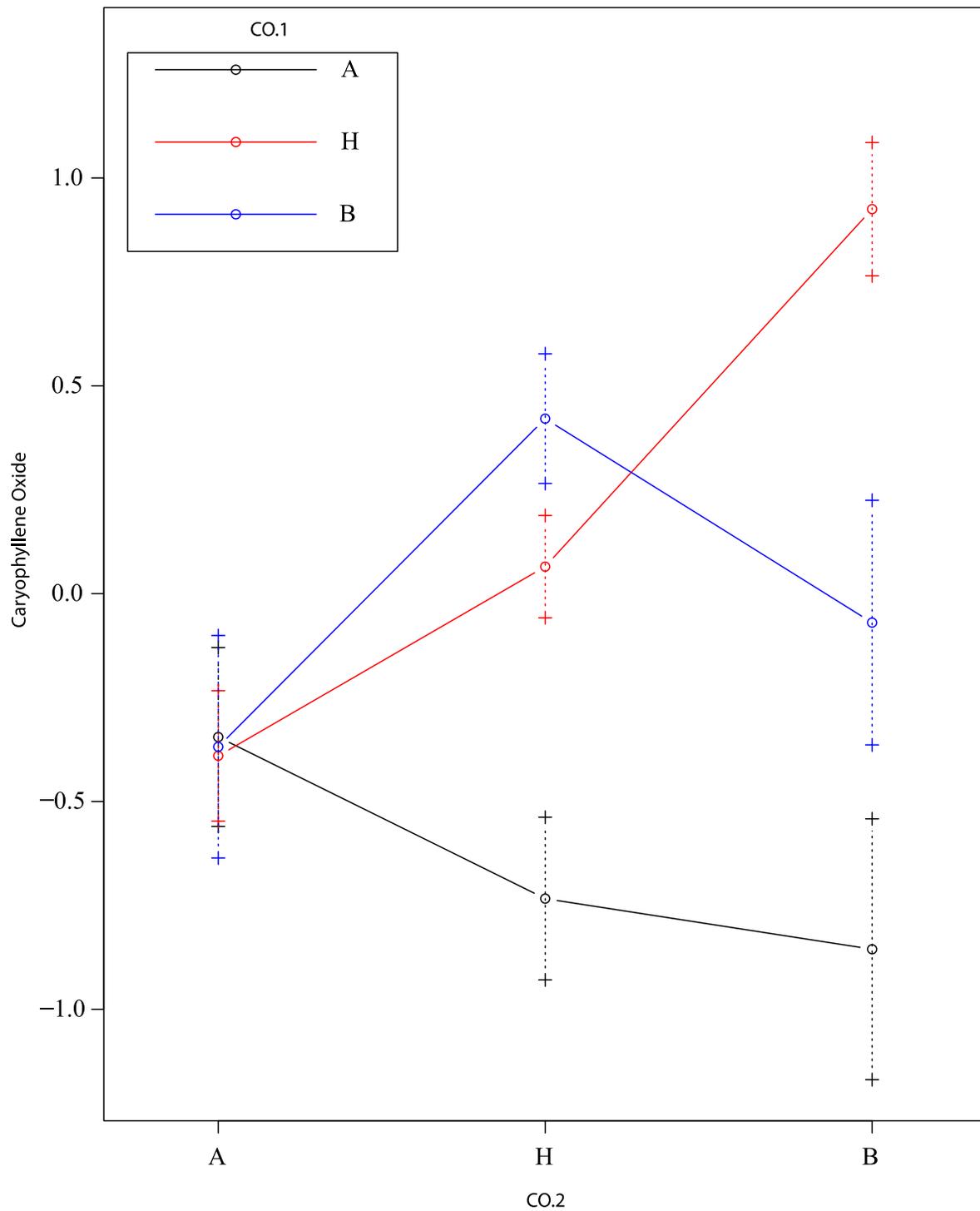
Supplementary Figure 4. Shown are the F₂ quantile normalized phenotype values (\pm standard error) across each F₂ plant genotype. Lesser y-axis values indicate lower seed mass while higher y-axis values indicate greater thousand seed mass.



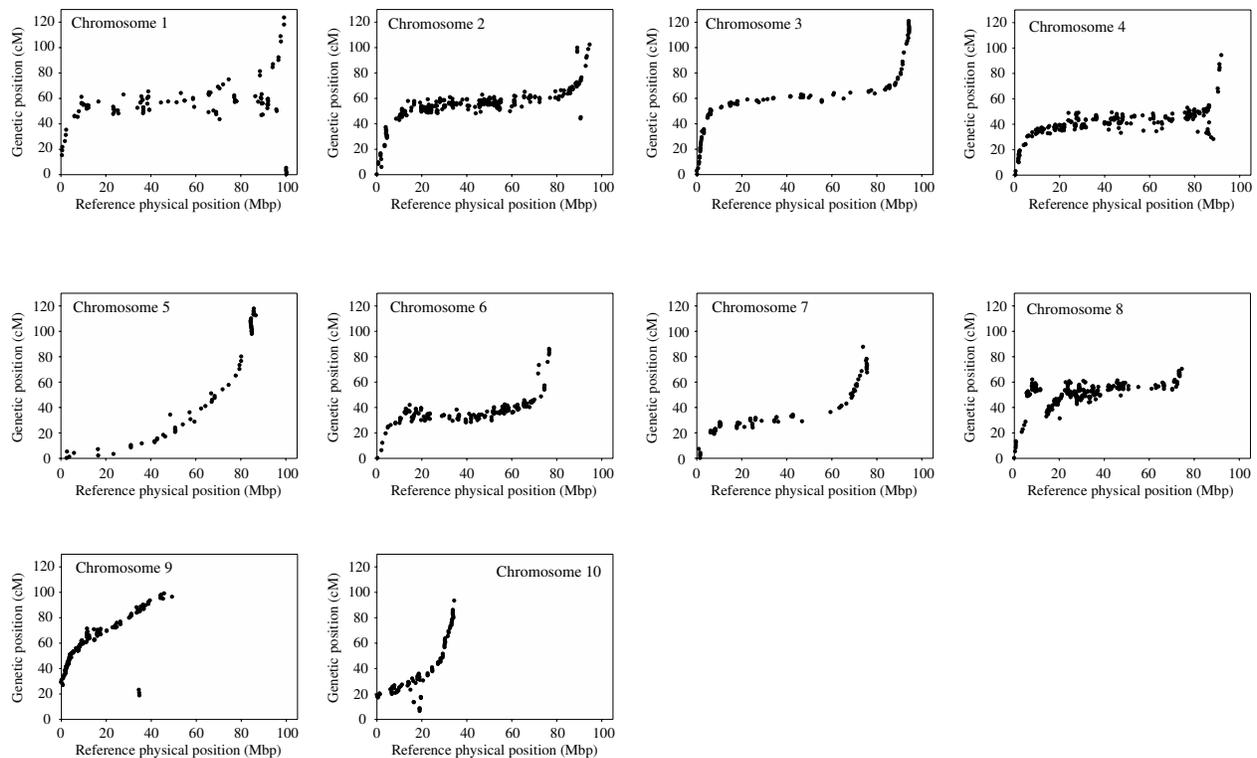
Supplementary Figure 5. Boxplots with individual points showing differences for agronomic trait performance among F₂ line genotypes at the single non-synonymous SNP identified in TINY. X-axis positions indicate the F₂ individual allelic state at the TINY non-synonymous SNP. “C” indicates homozygous for cytosine, “T” indicates homozygous for thymine and “Y” indicates a heterozygote. “C” alleles are derived from the USO31 parent while “T” alleles are derived from the Carmagnola parent. Y-axis positions indicate the quantile normalized values for the respective trait.



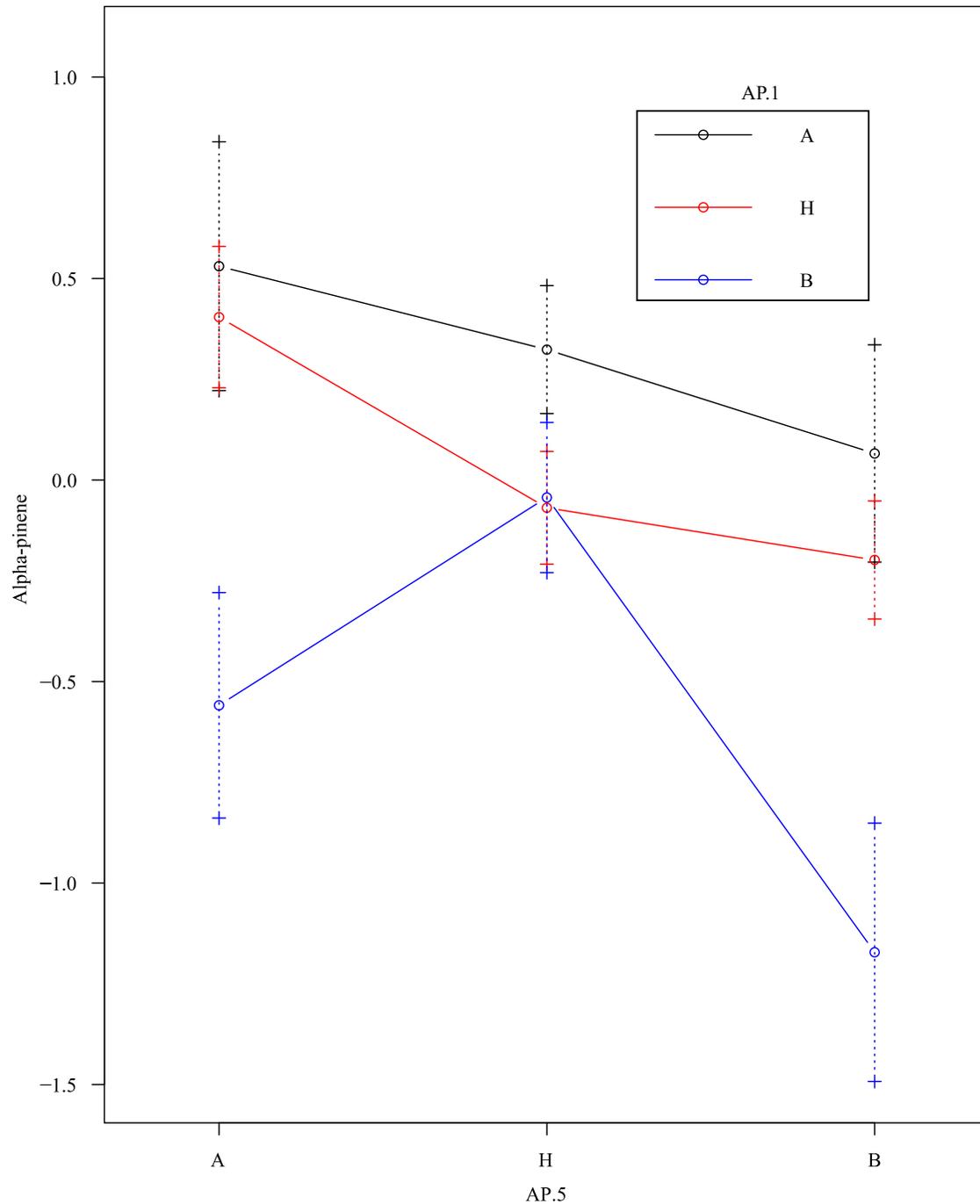
Supplementary Figure 6. Amino acid alignments between olivetol synthase from Taura et al. 2009 and the two transcripts we designed for Carmagnola and USO31. Amino acid substitutions are indicated by the orange points.



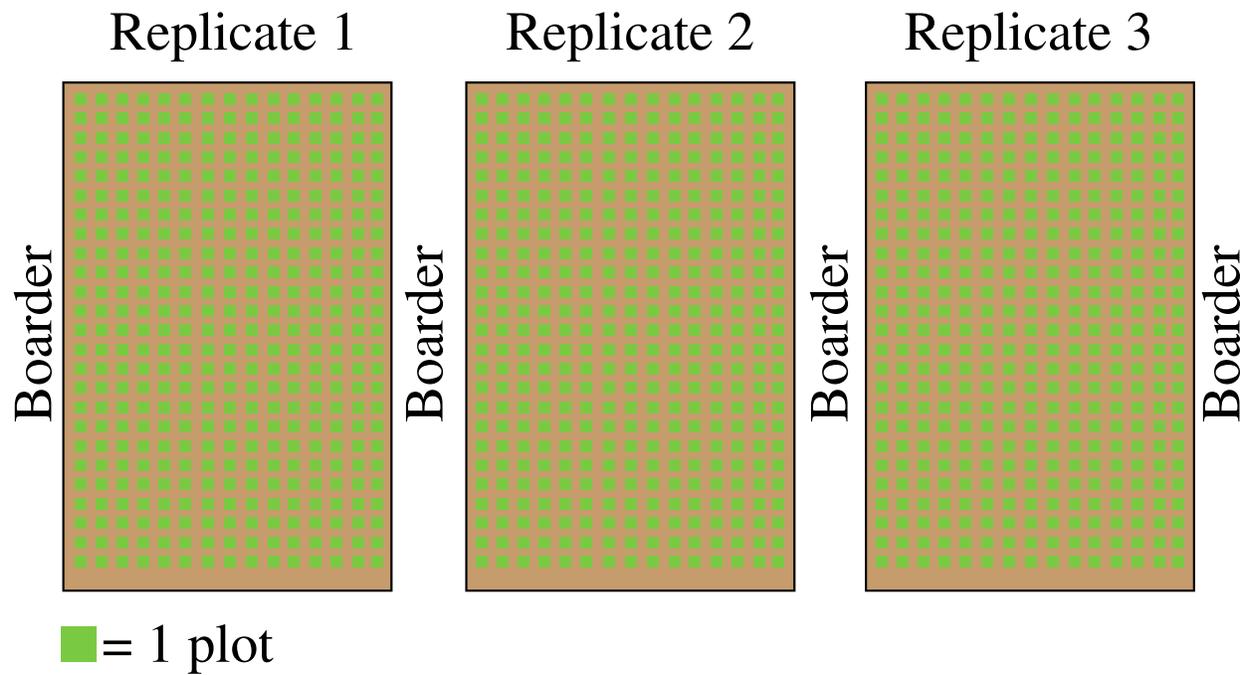
Supplementary Figure 7. Epistatic interaction of the caryophyllene oxide QTL CO.1 and CO.2. Line colors indicate F₂ plant genotype at CO.1 while x-axis positions indicate F₂ plant genotype at CO.2. Y-axis indicates the mean (\pm standard error) quantile normalized caryophyllene oxide quantities. Lesser y-axis values indicate lower quantities while higher y-axis values indicate greater quantities of caryophyllene oxide produced.



Supplementary Figure 8: Shown are the physical versus genetic positions for the SNPs used in linkage map construction that correspond to each chromosome in version 2 of the Finola assembly. X-axis indicates the physical positions (mega base pairs) of the SNPs in the Finola assembly while y-axis positions indicate the genetic positions (centiMorgans) of the SNPs in the linkage map.



Supplementary Figure 9. Epistatic interaction of the alpha-pinene QTL AP.1 and AP.5. Line colors indicate F₂ plant genotype at AP.1 while x-axis positions indicate F₂ plant genotype at AP.5. Y-axis indicates the mean (\pm standard error) quantile normalized alpha-pinene quantities. Lesser y-axis values indicate lower quantities while higher y-axis values indicate greater quantities of alpha-pinene produced.



Supplementary Figure 10: Design of field experiment. Shown is an illustration of the field design used for our field experiment. The experiment had three replicate blocks of 375 plants (Carmagnola, USO31, F₁ and 372 F₂ plants). A single plant was contained in each experimental plot (green squares).

>Carmagnola_TINY_predicted_cds

ATGAGAACATGGGGAAAATGGGTATCCGAAATTAGGGAGCCCCGGAAGAAGAATC
GGATCTGGCTAGGAACTTTCTCAACACCAGAAATGGCGGCGCGTGCACACGACGTG
GCGGCCTTAAGCATCAAGGGTAACTCTGCAATACTCAACTTCCCCGAGCTAGCCGGA
TCATTACCCCGGCCGGAATCTAACTCACCCCGAGACGTTCAAGCCGCGGCTACAAA
AGCGGCTTTAATGGAGTTTCCGAACCCATCTCCTTCACCTTCCACGTCATCATCGAC
AACAACGAGTACTTCATCGACAATGTCGCAGACTTCGTCTTCGTCTTCTTCGTCGTTG
GTGGCAACTTCGTCATCGAGCTACGACGTGTCGTCTTCGCCGGAAGAGCTGAGCGAG
ATAGTTGAGCTGCCGAGTTTGGATACGAGTTTTCGAGTCGGGGAACGAGTTCGTTTTC
GAAGACTCGGTAGAGGGGTGGCTGTATCCTCCTCCGCCGTGGTACCATCATCAGAGC
TTGTTTGAAGAAAATTATGGGTACTTTAGGGAAAGATCAGATAGCAATGTCAGAGCC
AGAGTCTGCTATTCTATCATCTTTTGAGCCTTTATTATGGCAACATTAA

>USO31_TINY_predicted_cds

ATGAGAACATGGGGAAAATGGGTATCCGAAATTAGGGAGCCCCGGAAGAAGAATC
GGATCTGGCTAGGAACTTTCTCAACACCAGAAATGGCGGCGCGTGCACACGACGTG
GCGGCCTTAAGCATCAAGGGTAACTCTGCAATACTCAACTTCCCCGAGCTAGCCGGA
TCATTACCCCGGCCGGAATCTAACTCACCCCGAGACGTTCAAGCCGCGGCTACAAA
AGCGGCTTTAATGGAGTTTCCGAACCCATCTCCTTCACCTTCCACGTCATCATCGAC
AACAACGAGTACTTCATCGACAATGTCGCAGACTTCGTCTTCGTCTTCTTCGTCGTTG
GTGGCAACTTCGTCATCGAGCTACGACGTGTCGTCTTCGCCGGAAGAGCTGAGCGAG
ATAGTTGAGCTGCCGAGTTTGGATACGAGTTTTCGAGCCGGGGAACGAGTTCGTTTTC
GAAGACTCGGTAGAGGGGTGGCTGTATCCTCCTCCGCCGTGGTACCATCATCAGAGC
TTGTTTGAAGAAAATTATGGGTACTTTAGGGAAAGATCAGATAGCAATGTCAGAGCC
AGAGTCTGCTATTCTATCATCTTTTGAGCCTTTATTATGGCAACATTAA

>Carmagnola_derived_olivetol_synthase_cds

GGATCCAAAATGAATCACTTGCGAGCAGAAGGCCCCGCAACCGTCTTGGCTATTGG
AACTGCTAATCCGGAGAATTTTCTGTTGCAAGATGAGTTCCTGGACTATTATTTCCGT
GTTACTAAATCAGAACACATGACTCAGTTGAAGGAGAAGTTTCGTAAGATTTGCGAT
AAATCTATGATCCGAAAGCGGAATTGTTTCCTGAATGAGGAACACCTCAAACAAAA
TCCACGGCTCGTGGAACATGAAATGCAAACCTCTCGACGCGGACAAGATATGTTGG
TTGTGGAAGTCCCCAAATTGGGAAAAGATGCTTGCCTAAAGCTATCAAAGAATGG
GGACAACCCAAATCTAAAATTACACATCTGATCTTTACTAGTGCCAGCACTACTGAT
ATGCCAGGAGCCGACTATCACTGTGCTAAATTACTCGGATTGTCGCCCTCTGTCAA
CGTGTGATGATGTACCAGTTAGGATGCTATGGTGGTGGCACTGTCTTGC GGATCGCT
AAAGACATCGCCGAAAACAATAAAGGTGCCCGTGTTTTAGCAGTTTGCTGTGATATG
ACAGCGTGTGTTTCGCGGACCCTCTGACTCCAATTTGGAATTGTTAGTGGGTCAA
GCTATCTTTGGAGACGGAGCTGCTGCTGTTATAGTTGGTGCAGAACCAGATGAATCG
GTGGGTGAACGCCCATTTTCGAATTGGTGTGACAGGACAAACCTTCTTGCCTAAT
AGCGAAGGAACCATTTGGCGGACATATCCGGGAAGCTGGTCTGATTTTCGATTTGCAC
AAGGATGTTCCAATGCTGATCTCTAATAATATCGAGAAATGCCTTATCGAGGCGTTC
ACTCCGATTGGTATCTCTGACTGGAACCTCATCTTTTGGATTACACATCCCGGTGGTA
AAGCCATCTTAGATAAAGTGAAGAAAAGCTGCATTTAAAATCAGACAAGTTTGTG
GATAGCCGCCATGTCTTGTCTGAACATGGAAATATGTCTTCGTCTACTGTGCTGTTCC
TCATGGACGAGTTGCGCAAACGGTCATTGGAAGAAGGAAAATCTACCACAGGCGAT
GGTTTCGAATGGGGAGTCTTGTGTTGGCTTCGGTCTGGCTTGACCGTTCGAACGCGTG
GTTCTCCGTTCCGGTGCCTATCAATTATTAAGAATTCGCGGCCGCT

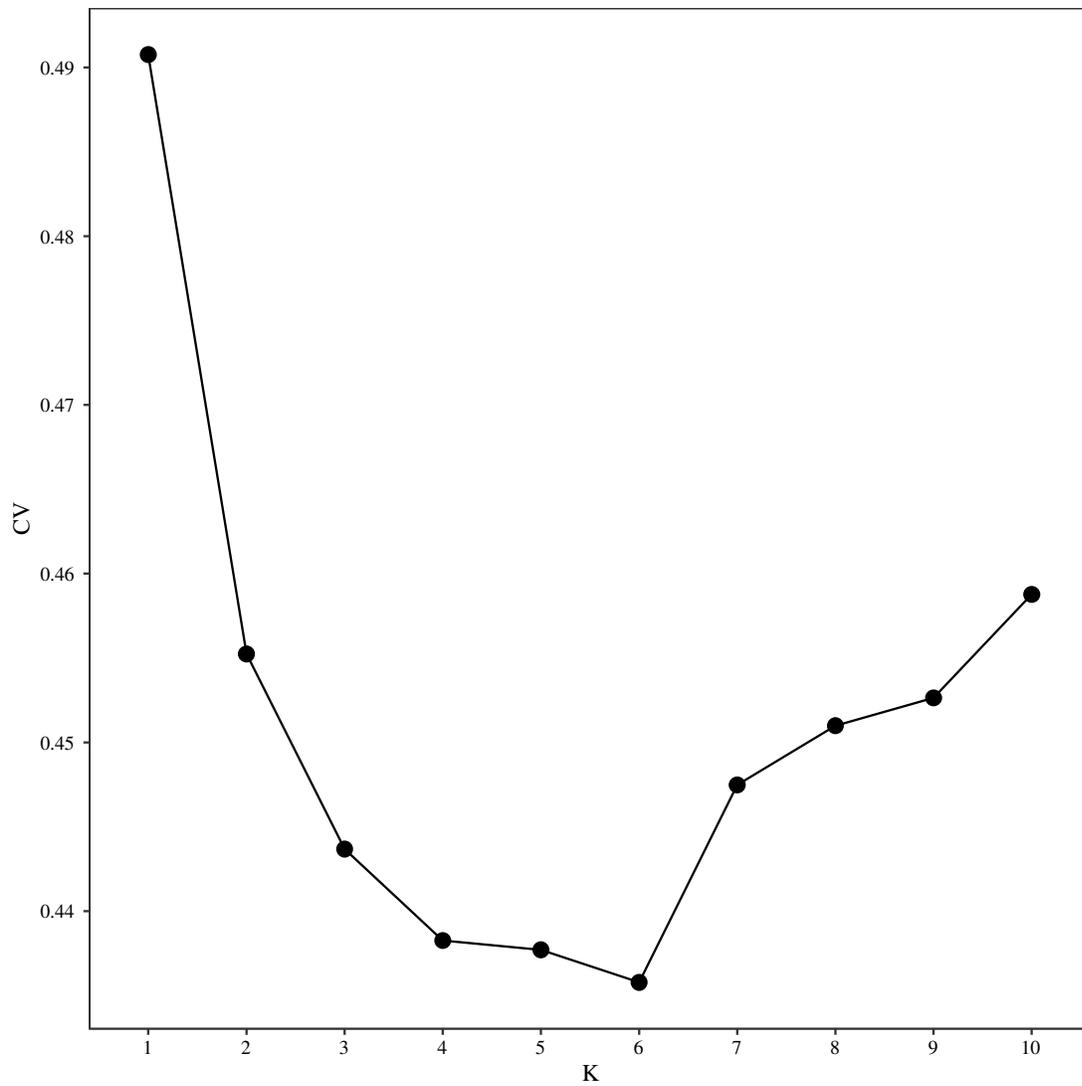
>USO31_derived_olivetol_synthase_cds

GGATCCAAAATGAATCACTTGCGAGCAGAAGGCCCCGCATCCGTCTTGGCTATTGGA
ACTGCTAATCCGGAGAATATCCTGTTGCAAGATGAGTTCCTGGACTATTATTTCCGT
GTTACTAAATCAGAACACATGACTCAGTTGAAGGAGAAGTTTCGTAAGATTTGCGAT
AAATCTATGATCCGAAAGCGGAATTGTTTCCTGAATGAGGAACACCTCAAACAAAA
TCCACGGCTCGTGGAACATGAAATGCAAACCTCTCGACGCGGACAAGATATGTTGG
TTGTGGAAGTCCCCAAATTGGGAAAAGATGCTTGCCTAAAGCTATCAAAGAATGG
GGACAACCCAAATCTAAAATTACACATCTGATCTTTACTAGTGCCAGCACTACTGAT
ATGCCAGGAGCCGACTATCACTGTGCTAAATTACTCGGATTGTCGCCCTCTGTCAA
CGTGTGATGATGTACCAGTTAGGATGCTATGGTGGTGGCACTGTCTTGC GGATCGCT
AAAGACATCGCCGAAAACAATAAAGGTGCCCGTGTTTTAGCAGTTTGCTGTGATATC
ATGGCGTGTGTTTCGCGGACCCTCTGAGTCCGACTTGGAAATTGTTAGTGGGTCAA
GCTATCTTTGGAGACGGAGCTGCTGCTGTTATAGTTGGTGCAGAACCAGATGAATCG
GTGGGTGAACGCCCATTTTCGAATTGGTGTGACAGGACAAACCATCTTGCCTAAT
AGCGAAGGAACCATTTGGCGGACATATCCGGGAAGCTGGTCTGATTTTCGATTTGCAC
AAGGATGTTCCAATGCTGATCTCTAATAATATCGAGAAATGCCTTATCGAGGCGTTC
ACTCCGATTGGTATCTCTGACTGGAACCTCATCTTTTGGATTACACATCCCGGTGGTA
AAGCCATCTTAGATAAAGTGAAGAAAAGCTGCATTTAAAATCAGACAAGTTTGTG
GATAGCCGCCATGTCTTGTCTGAACATGGAAATATGTCTTCGTCTACTGTGCTGTTCC
TCATGGACGAGTTGCGCAAACGGTCATTGGAAGAAGGAAAATCTACCACAGGCGAT
GGTTTCGAATGGGGAGTCTTGTGTTGGCTTCGGTCTGGCTTGACCGTTCGAACGCGTG
GTTGTCCGTTCCGGTGCCTATCAAGTATTAAGAATTCGCGGCCGCT

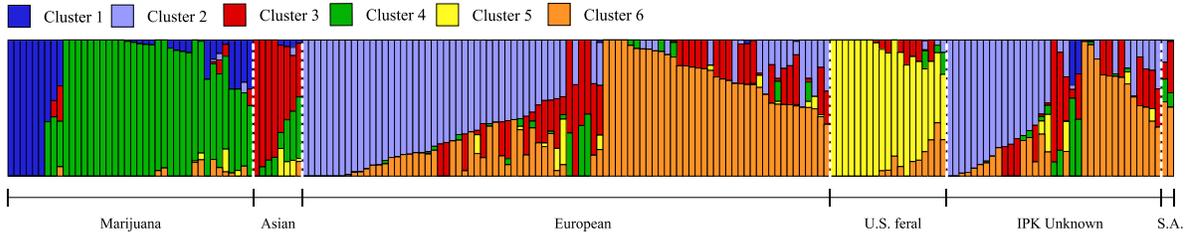
APPENDIX II

SUPPLEMENTAL INFORMATION FOR CHAPTER 3: GENOME WIDE POLYMORPHISM
AND GENIC SELECTION IN FERAL AND DOMESTICATED LINEAGES OF *CANNABIS*

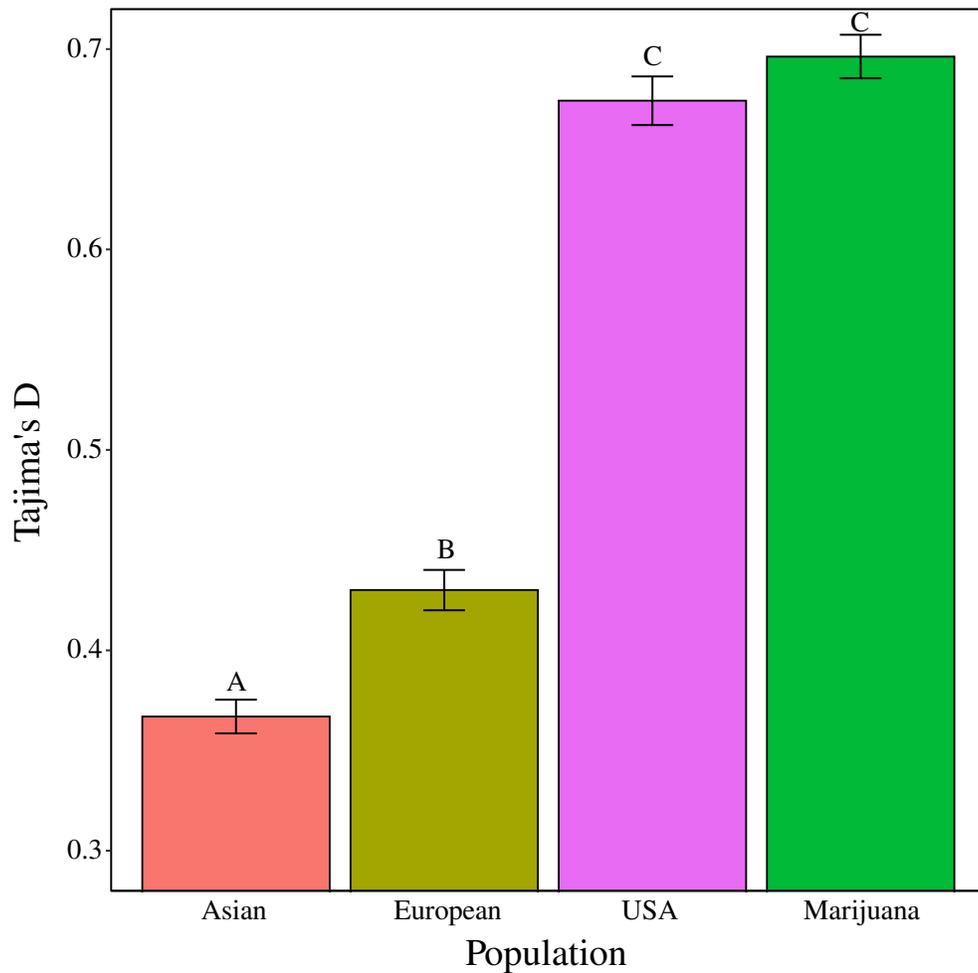
SATIVA



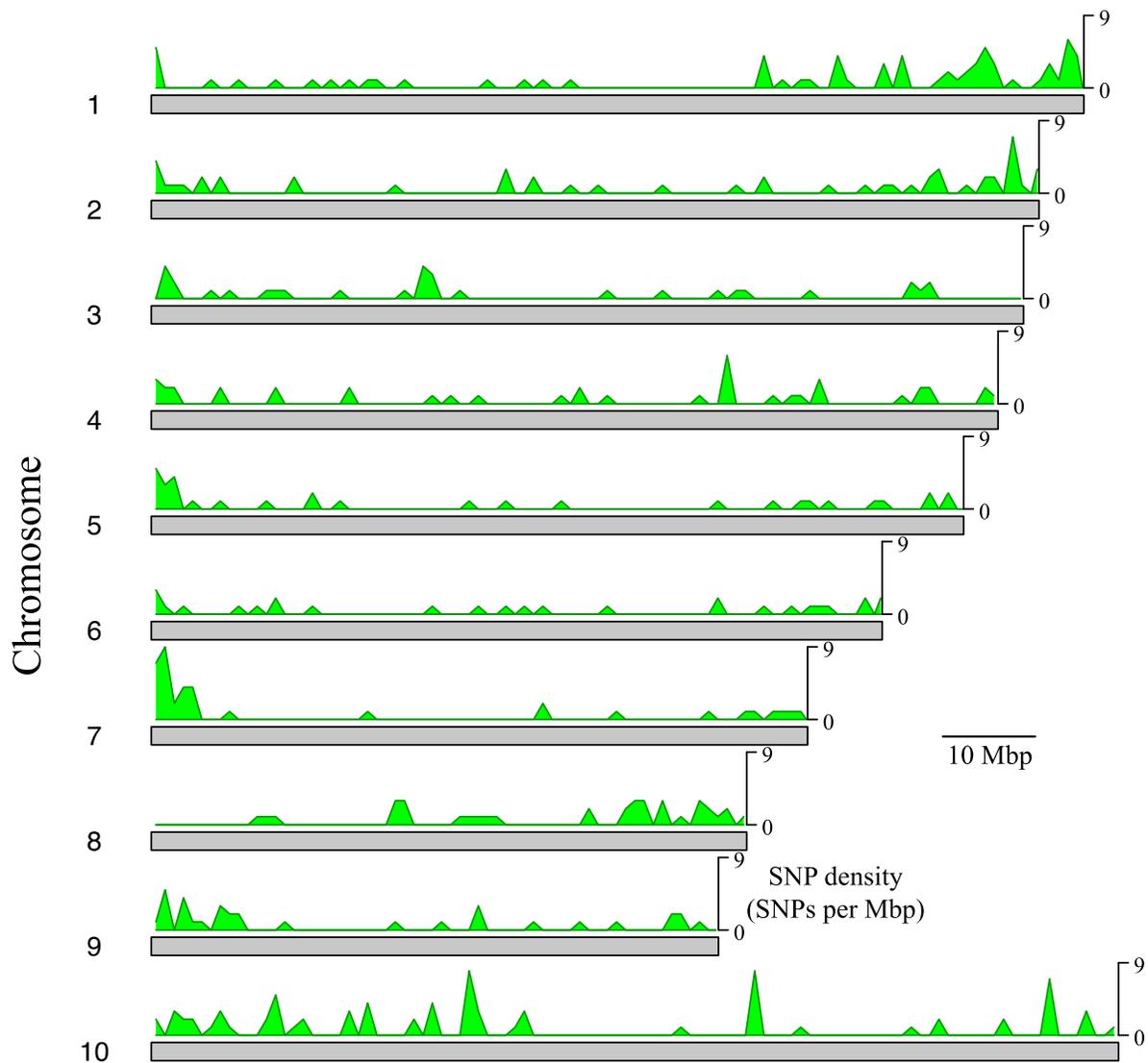
Supplemental Figure 11. Cross validation (CV) errors calculated by ADMIXTURE for K values 1-10. K = 6 exhibited the lowest cross validation error.



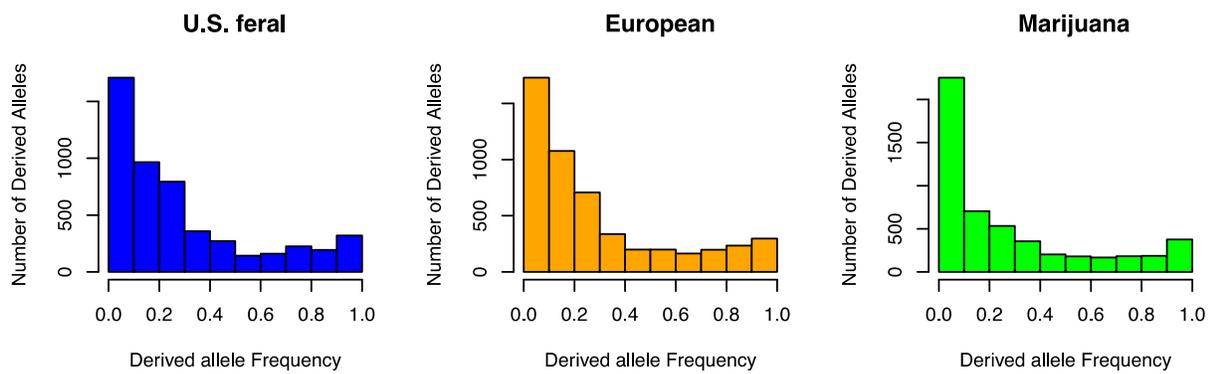
Supplemental Figure 12. Stacked bar plots representing ADMIXTURE results showing individual ancestry likelihoods assuming $K = 6$.



Supplemental Figure 13. Bar plots showing the comparisons of mean Tajima's D (\pm standard error) across the Asian, European, U.S. feral and Marijuana sub-populations at the 7540 single copy orthologs. Letters above bars indicate significance of Tukey adjusted post-hoc pairwise comparisons such that different letters indicate significant differences ($p < 0.05$).



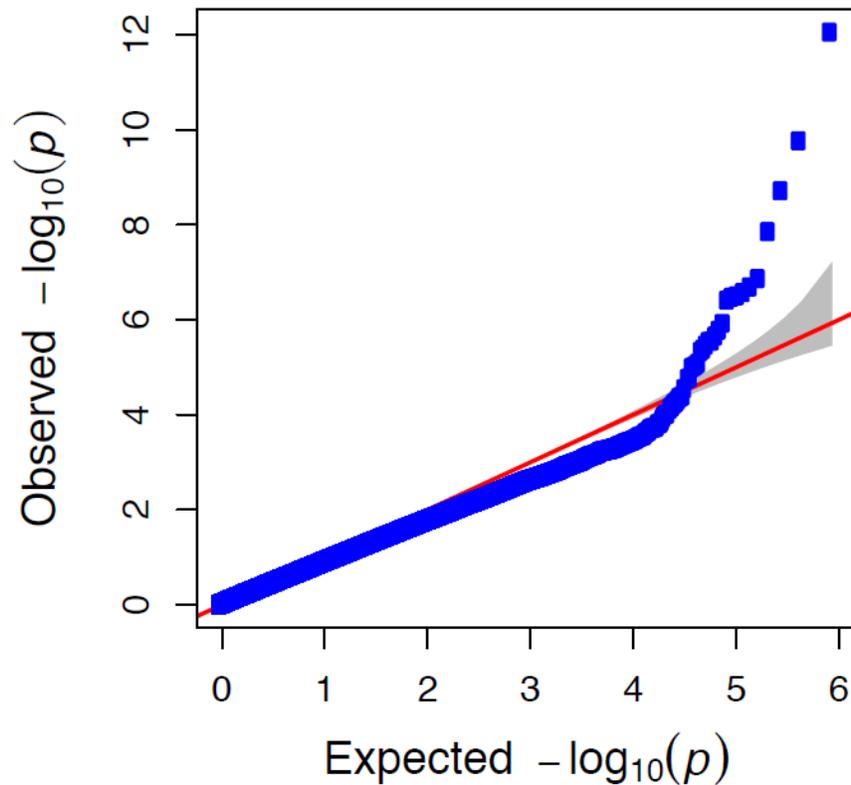
Supplemental Figure 14. SNP density plot depicting the locations of derived alleles showing evidence for species wide selection. Shown are horizontal grey bars representing the ten *C. sativa* chromosomes with local green density plots that depict the number of species wide high frequency derived alleles in 1 megabase intervals.



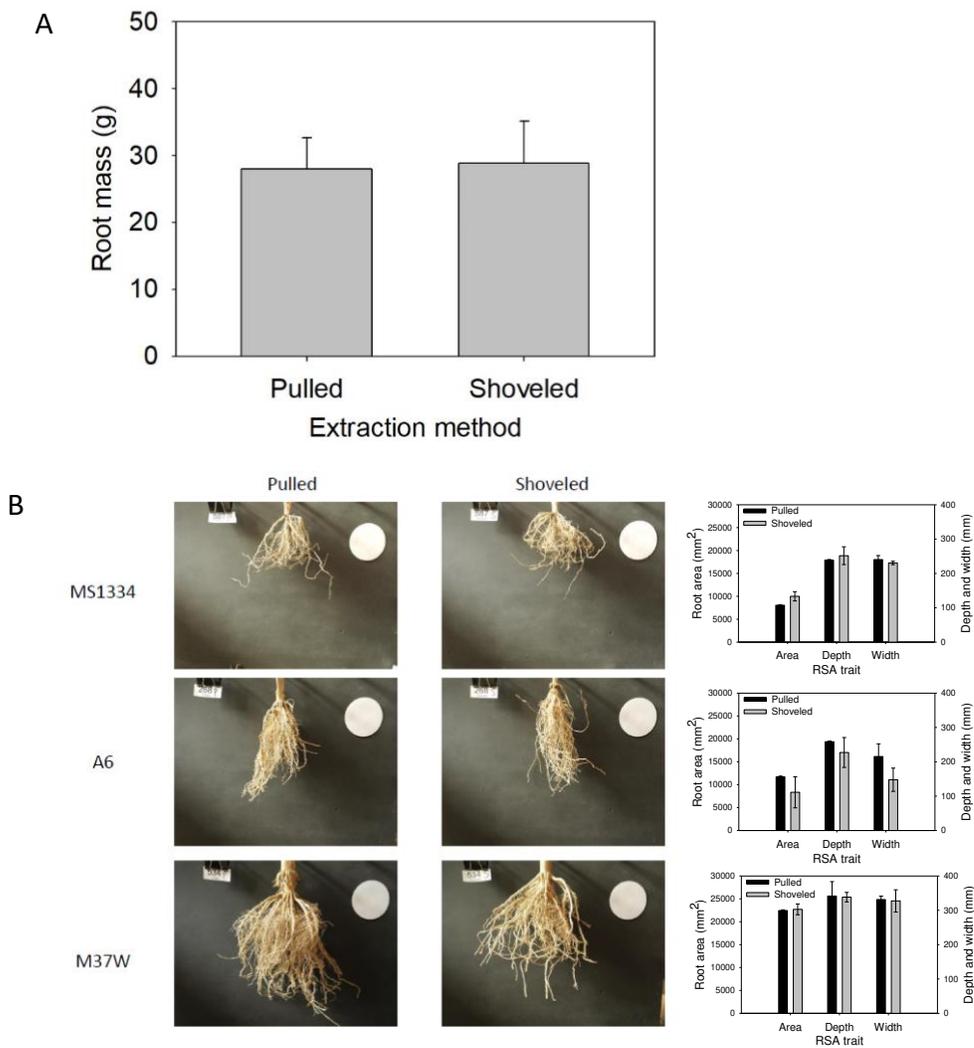
Supplemental Figure 15. Individual derived allele frequency spectrums for the U.S. feral, European and Marijuana sub-populations respectively.

APPENDIX III

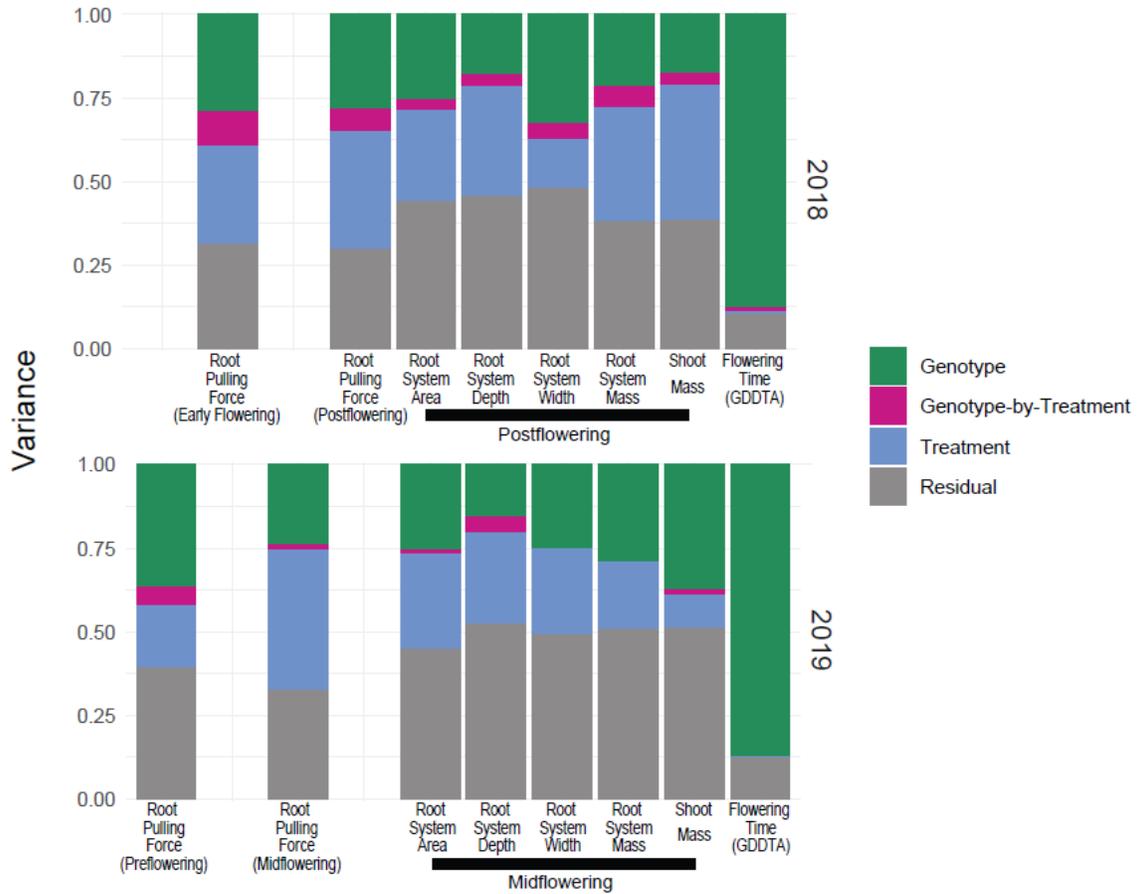
SUPPLEMENTAL INFORMATION FOR CHAPTER 4: ROOT PULLING FORCE ACROSS DROUGHT IN MAIZE IDENTIFIES GENOTYPE BY ENVIRONMENT INTERACTIONS AND CANDIDATE GENES



Supplemental Figure 16. Q-Q plot assessing p-value distributions for the GWA conducted on the fully irrigated post-flowering RPF measurement. Blue squares indicate the comparison of the expected $-\log_{10}(p)$ (X axis) values versus the observed $-\log_{10}(p)$ (y axis) values. The red line indicates the null hypothesis of complete concordance between expected and observed $-\log_{10}(p)$ values.



Supplemental Figure 17. Comparison of RSA for pulled and shovel excavated root systems. A, Crown root mass for the two extraction methods (mean \pm SE, $n = 42-44$; $P = 0.84$, paired t test). B, Image-based root system architecture traits extracted by the different methods for three genotypes with varying root form (left, sample images; right, average values for root area, depth and width (mean \pm SE, $n = 2$)). For all three of the traits, there was a significant effect of genotype (ANOVA, $P < 0.01$) but not of extraction method ($P > 0.10$).



Supplemental Figure 18. Variance component analysis of traits from the 2018 and 2019 field seasons.

Supplemental Table 10. Summary of weather and irrigation for the four environments. FI, full irrigation; LI, limited irrigation.

Environment	June			July			August		
	Ave Temp (°C)	Precip (mm)	Irrigation (mm)	Ave Temp (°C)	Precip (mm)	Irrigation (mm)	Ave Temp (°C)	Precip (mm)	Irrigation (mm)
2018 FI	20.9	21	104	22.0	38	133	20.4	3	140
2018 LI	20.9	21	76	22.0	38	44	20.4	3	0
2019 FI	17.7	75	51	22.4	33	114	21.8	14	102
2019 LI	17.7	75	25	22.4	33	25	21.8	14	51

Supplemental Table 12. Genotypic correlations for the 2018 field season. Values are Spearman's correlation coefficients.

		Full-Irrigation								Limited-Irrigation								
		RPF (Early-flowering)	RPF (Postflowering)	Root System Area (Postflowering)	Root System Depth (Postflowering)	Root System Width (Postflowering)	Root System Mass (Postflowering)	Shoot Mass (Postflowering)	Flowering Time (GDDTA)	RPF (Early-flowering)	RPF (Postflowering)	Root System Area (Postflowering)	Root System Depth (Postflowering)	Root System Width (Postflowering)	Root System Mass (Postflowering)	Shoot Mass (Postflowering)	Flowering Time (GDDTA)	
Full Irrigation	RPF (Early-flowering)	1																
	(Postflowering)	0.6	1															
	Root System Area (Postflowering)	0.44	0.67	1														
	Depth (Postflowering)	0.46	0.51	0.68	1													
	Width (Postflowering)	0.27	0.4	0.7	0.4	1												
	Root System Mass (Postflowering)	0.39	0.59	0.63	0.56	0.33	1											
	Shoot Mass (Postflowering)	0.27	0.53	0.53	0.35	0.34	0.61	1										
	Flowering Time (GDDTA)	-0.23	0.21	0.1 (NS)	0 (NS)	-0.03 (NS)	0.26	0.17 (**)	1									
	Limited Irrigation	RPF (Early-flowering)	0.6	0.5	0.44	0.44	0.27	0.41	0.26	-0.09 (NS)	1							
		(Postflowering)	0.52	0.63	0.49	0.43	0.33	0.48	0.33	0.23	0.59	1						
Root System Area (Postflowering)		0.47	0.47	0.54	0.4	0.48	0.4	0.34	0.07 (NS)	0.47	0.69	1						
Depth (Postflowering)		0.37	0.33	0.35	0.41	0.23	0.34	0.27	0.02 (NS)	0.37	0.53	0.7	1					
Width (Postflowering)		0.32	0.34	0.42	0.26	0.53	0.27	0.27	0.05 (NS)	0.31	0.48	0.78	0.48	1				
Root System Mass (Postflowering)		0.43	0.38	0.4	0.36	0.26	0.52	0.38	0.07 (NS)	0.5	0.65	0.72	0.52	0.56	1			
Shoot Mass (Postflowering)		0.26	0.32	0.34	0.23	0.29	0.41	0.54	0.12 (*)	0.31	0.51	0.57	0.4	0.47	0.64	1		
Flowering Time (GDDTA)		-0.2	0.23	0.11 (NS)	-0.03 (NS)	-0.01 (NS)	0.28	0.21	0.96	-0.1 (NS)	0.23	0.09 (NS)	0.03 (NS)	0.08 (NS)	0.08 (NS)	0.13 (*)	1	
All others are (***)																		

Supplemental Table 13. Genotypic correlations for the 2019 field season. Values are Spearman's correlation coefficients.

		Full-Irrigation								Limited-Irrigation								
		RPF (Preflowering)	RPF (Midflowering)	Root System Area (Midflowering)	Root System Depth (Midflowering)	Root System Width (Midflowering)	Root System Mass (Midflowering)	Shoot Mass (Midflowering)	Flowering Time (GDDTA)	RPF (Preflowering)	RPF (Midflowering)	Root System Area (Midflowering)	Root System Depth (Midflowering)	Root System Width (Midflowering)	Root System Mass (Midflowering)	Shoot Mass (Midflowering)	Flowering Time (GDDTA)	
Full Irrigation	RPF (Preflowering)	*																
	RPF (Midflowering)	0.54	*															
	Area (Midflowering)	0.49	0.76	*														
	Root System Depth (Midflowering)	0.39	0.58	0.76	*													
	Root System Width (Midflowering)	0.32	0.57	0.76	0.5	*												
	Root System Mass (Midflowering)	0.46	0.73	0.74	0.56	0.51	*											
	Shoot Mass (Midflowering)	0.49	0.72	0.63	0.48	0.45	0.7	*										
	Flowering Time (GDDTA)	-0.39	-0.05 (NS)	-0.15 (**)	-0.22	-0.07 (NS)	-0.08 (NS)	-0.2	*									
	Limited Irrigation	RPF (Preflowering)	0.64	0.52	0.52	0.47	0.36	0.48	0.51	-0.36	*							
		RPF (Midflowering)	0.49	0.63	0.5	0.39	0.41	0.5	0.43	0.05 (NS)	0.5	*						
Area (Midflowering)		0.44	0.58	0.56	0.36	0.47	0.56	0.47	-0.05 (NS)	0.49	0.75	*						
Depth (Midflowering)		0.35	0.42	0.4	0.38	0.32	0.43	0.3	0.03 (NS)	0.41	0.64	0.75	*					
Width (Midflowering)		0.3	0.47	0.47	0.26	0.55	0.41	0.35	0.04 (NS)	0.37	0.57	0.76	0.52	*				
Root System Mass (Midflowering)		0.45	0.55	0.52	0.44	0.39	0.61	0.47	0.01 (NS)	0.48	0.73	0.77	0.66	0.53	*			
Shoot Mass (Midflowering)		0.44	0.47	0.39	0.31	0.35	0.44	0.58	-0.11 (**)	0.45	0.63	0.6	0.46	0.46	0.69	*		
Flowering Time (GDDTA)		-0.36	-0.05 (NS)	-0.15 (**)	-0.24	-0.06 (NS)	-0.06 (NS)	-0.2	0.94	-0.34	0.04 (NS)	-0.05 (NS)	0.04 (NS)	0.04 (NS)	0 (NS)	-0.12 (**)	*	
All others are (***)																		

Supplemental Table 14. List of significant SNPs for root traits.

Trait	Marker_SNP_V2	Chromosome	B73_V5_Position	P_value	FDR_Adjusted_P_value	MAF	SNP/Treatment_V0	Closest_Gene	NCBI_ID	Annotation	Other_Genes_in_Interval
Area_2019_Midflowering_Limited_Irrigation	1_2886253	1	28000386	1.11E-08	0.00238393	0.244514	0.668875	Zm00001eb009320	LOC103633057	wall-associated_receptor_kinase-like_14	
RFP_2018_Earlyflowering_Full_Irrigation	1_52633486	1	52761549	2.27E-07	0.01955448	0.193215	0.53757	Zm00001eb015420	LOC100283111	Mitochondrial_import_receptor_subunit_TOM40-1	
RFP_2018_Earlyflowering_Full_Irrigation	1_70850891	1	71369452	1.40E-08	0.00301383	0.113772	0.232765	NA	NA	NA	NA
Area_2019_Midflowering_Limited_Irrigation	1_71607329	1	72169664	3.47E-07	0.03321763	0.34326	0.89038	Zm00001eb019800	LOC100274688	pco121831_E3_ubiquitin-protein_ligase_BRE1-like_2	NA
Area_2018_Postflowering_Full_Irrigation	1_8853057	1	89661777	1.68E-12	1.45E-06	0.332308	0.02402	Zm00001eb023020	LOC100274725	NADH_dehydrogenase_ubiquinone_1_beta_subkom2	NA
Area_2019_Midflowering_Limited_Irrigation	1_17464987	1	17876645	1.76E-12	1.52E-06	0.268025	0.441026	Zm00001eb031970	LOC100192777	bsu1-brassinosteroid_insensitive_suppressor_protein	NA
RFP_2018_Earlyflowering_Full_Irrigation	1_18463284	1	188819897	5.03E-08	0.01444064	0.274729	0.6597556	NA	NA	NA	NA
RFP_2018_Earlyflowering_Full_Irrigation	1_19579532	1	200108340	9.08E-12	3.91E-06	0.160767	0.3411	Zm00001eb037030	LOC100279791	uncharacterized	NA
Root_Mass_2018_Postflowering_Full_Irrigation	1_22825493	1	233758700	7.35E-10	0.000316679	0.072131	0.0001781	Zm00001eb048480	NA	ngl447a_similar_to_Arabidopsis_alpha-galactosidase	NA
RFP_2018_Earlyflowering_Limited_Irrigation	1_240090714	1	245781661	1.00E-07	0.01232527	0.141813	0.4664	Zm00001eb047830	LOC103643560	protein_NETWORKED_1D	NA
RFP_2018_Earlyflowering_Full_Irrigation	1_257144174	1	263703479	4.34E-07	0.01157166	0.479351	0.011868	Zm00001eb052050	LOC100277915	uncharacterized	NA
Area_2019_Postflowering_Full_Irrigation	1_284011654	1	290807894	3.62E-07	0.044557601	0.166154	0.05264	Zm00001eb059830	LOC100282761	inositolphosphoryltransferamide-B_C-26_hydroxylase	Zm00001eb059620
RFP_2018_Midflowering_Limited_Irrigation	1_28642560	1	29326580	5.98E-09	0.00257697	0.084639	0.973803	Zm00001eb060250	LOC542238	ac1-aldehyde_oxidase1	NA
RFP_2018_Postflowering_Full_Irrigation	2_2521440	2	5439373	3.52E-07	0.032160857	0.426647	0.2872388	Zm00001eb065870	LOC103645845	UDP_glycosyltransferase_89B1_like	Zm00001eb068380
RFP_2018_Earlyflowering_Full_Irrigation	2_189571970	2	194795692	1.39E-12	1.20E-06	0.10767	1.1304	Zm00001eb101840	NA	znf13_zinc_finger_protein13	Zm00001eb101820
Area_2018_Postflowering_Full_Irrigation	2_19196855	2	197925572	3.81E-08	0.008373172	0.183077	0.25848	Zm00001eb102380	LOC100282453	uncharacterized	NA
RFP_2018_Postflowering_Limited_Irrigation	2_201905510	2	201300073	1.12E-10	4.83E-05	0.44883	0.857392	Zm00001eb103300	LOC100216563	Ypt_Rab_GAP_domain_of_gyp1p_superfamily_prot	NA
RFP_2018_Postflowering_Limited_Irrigation	2_225678072	2	232124026	9.73E-09	0.002795563	0.093567	0.447575	Zm00001eb114070	LOC100193577	bstf1_BSD_transcription_factor_1	Zm00001eb114080
Area_2018_Postflowering_Full_Irrigation	3_2764530	3	2752924	8.30E-10	0.00035746	0.449231	0.2638375	Zm00001eb119560	LOC103649556	probable_inactive_receptor_kinase_AT4g23740	Zm00001eb119570
RFP_2018_Earlyflowering_Limited_Irrigation	3_38837474	3	37036459	6.68E-11	4.83E-05	0.070175	0.6562497	Zm00001eb127880	LOC100281498	ch12_C3H_transcription_factor_312	NA
RFP_2018_Earlyflowering_Limited_Irrigation	3_124189168	3	125626205	6.09E-08	0.01232527	0.488304	0.9587	Zm00001eb136170	LOC100273335	LOC100193821_like_pseudogene	Zm00001eb136160
Area_2019_Midflowering_Limited_Irrigation	3_148281418	3	150367677	2.30E-07	0.02474692	0.26489	0.212652	Zm00001eb140740	LOC100279357	CRRA	NA
Width_2018_Postflowering_Limited_Irrigation	3_19737364	3	202551404	1.65E-09	0.000711957	0.417647	0.97851	Zm00001eb153060	LOC100191195	uncharacterized	NA
RFP_2018_Postflowering_Full_Irrigation	3_213390029	3	218974411	1.36E-07	0.023492807	0.275449	0.184741	Zm00001eb158100	LOC100193891	um18_ustilago_maydis_induced8	NA
RFP_2018_Earlyflowering_Limited_Irrigation	3_216464843	3	221942005	8.35E-08	0.011996347	0.064404	0.6774897	Zm00001eb159080	LOC100216899	S05_ribosomal_protein_L31_chloroplast_like	NA
RFP_2018_Earlyflowering_Full_Irrigation	3_217652755	3	223193889	1.23E-09	0.00064787	0.237611	0.003156	Zm00001eb159380	LOC103654328	probable_calcium_transporting_ATPase_8_plasma_m	Zm00001eb159500
Area_2019_Midflowering_Limited_Irrigation	4_3484831	4	4464213	1.86E-11	8.03E-06	0.257053	0.6195707	Zm00001eb165770	LOC100284245	lipase	NA
RFP_2018_Earlyflowering_Full_Irrigation	4_4737624	4	5770814	6.93E-11	1.99E-05	0.154867	0.34157	Zm00001eb166390	LOC100193315	uncharacterized	NA
Shoot_Mass_2018_Postflowering_Full_Irrigation	4_5243006	4	6257841	1.37E-11	1.18E-05	0.321543	0.04413	Zm00001eb166700	LOC100272411	Monosaccharide_sensing_protein_2	Zm00001eb166690
RFP_2019_Midflowering_Full_Irrigation	4_5243006	4	6257841	2.05E-07	0.019592705	0.322785	0.06776	Zm00001eb166700	LOC100272411	Monosaccharide_sensing_protein_2	Zm00001eb166690
Area_2019_Midflowering_Limited_Irrigation	4_48516169	4	6195853	1.33E-07	0.02051198	0.18652	0.683457	Zm00001eb168910	LOC060409	pd11_protein_disulfide_isomerase1	NA
RFP_2019_Midflowering_Full_Irrigation	4_15603122	4	16930219	3.54E-08	0.005349047	0.310127	0.556292	Zm00001eb169030	LOC103652889	putative_disease_resistance_protein_At1g50180	NA
Width_2018_Postflowering_Limited_Irrigation	4_4023834	4	46160868	6.83E-08	0.014709403	0.247059	0.69795	NA	NA	NA	NA
RFP_2018_Postflowering_Full_Irrigation	4_147900944	4	153130434	3.36E-07	0.032160857	0.067365	0.552648	Zm00001eb185930	LOC100282580	ATP_dependent_RNA_helicase_dhh1	NA
RFP_2018_Earlyflowering_Full_Irrigation	4_167505381	4	162758639	5.26E-08	0.00647682	0.325207	0.6680009	Zm00001eb187730	LOC100276787	plat27_PLAT2_transcription_factor_3	NA
Width_2018_Postflowering_Limited_Irrigation	4_165834025	4	170913660	9.38E-08	0.016164092	0.173529	0.770226	Zm00001eb189900	LOC100383669	hb115_Homeobox_transcription_factor_115	NA
Area_2018_Postflowering_Full_Irrigation	4_167193841	4	17228498	3.48E-07	0.044557601	0.069231	0.0867912	Zm00001eb190390	LOC100501623	uncharacterized	Zm00001eb190380
RFP_2018_Postflowering_Full_Irrigation	4_171921792	4	177077448	1.71E-10	7.37E-05	0.073353	0.0616928	Zm00001eb191650	LOC103654129	phos2_phosphate_transporter2	Zm00001eb191660
RFP_2018_Earlyflowering_Full_Irrigation	4_180673390	4	185608069	3.65E-07	0.028936682	0.139115	0.339618	Zm00001eb194170	LOC103654247	TPK_repeat_containing_thioredoxin_TTL4	NA
Shoot_Mass_2018_Postflowering_Full_Irrigation	4_186948135	4	192358302	3.02E-08	0.000649373	0.237395	0.003156	Zm00001eb196380	LOC103654328	probable_calcium_transporting_ATPase_8_plasma_m	Zm00001eb196390
Root_Mass_2018_Postflowering_Limited_Irrigation	4_187603761	4	19302656	3.10E-10	0.00026712	0.201667	0.71712	Zm00001eb196590	LOC100281546	AMP_binding_protein	NA
Width_2018_Postflowering_Limited_Irrigation	4_193341102	4	199504085	7.08E-09	0.02032925	0.302941	0.92018	Zm00001eb198220	LOC100275453	uncharacterized	Zm00001eb198230
Root_Mass_2018_Postflowering_Limited_Irrigation	4_196928664	4	203027401	1.16E-08	0.004988518	0.447531	0.50234	Zm00001eb199080	LOC103654457	WEB_family_protein_chloroplast_like	Zm00001eb199070
RFP_2018_Earlyflowering_Limited_Irrigation	4_23899665	4	247718490	9.50E-08	0.01232527	0.055556	0.255737	Zm00001eb208840	LOC100191624	RuvB_like_protein_1	Zm00001eb208850
RFP_2018_Earlyflowering_Full_Irrigation	4_23947782	4	248172395	9.02E-09	0.001153863	0.137168	0.3115	Zm00001eb209190	LOC100501509	kat7_xylan_alpha_1_3_arabinofuranosyl_transferase	Zm00001eb209180
RFP_2018_Earlyflowering_Full_Irrigation	5_5023498	5	5336509	9.19E-07	0.04993413	0.433628	0.1600393	Zm00001eb213510	LOC113218479	uncharacterized	Zm00001eb213500
RFP_2018_Postflowering_Full_Irrigation	5_18699967	5	19629132	2.04E-07	0.029301865	0.332335	0.045001	Zm00001eb219240	LOC103626058	uncharacterized	Zm00001eb219250
Root_Mass_2018_Postflowering_Full_Irrigation	5_51435988	5	53136167	1.74E-10	0.000153003	0.22459	0.238844	Zm00001eb226000	LOC100274986	uncharacterized	Zm00001eb226010
RFP_2019_Midflowering_Full_Irrigation	5_61227311	5	63063966	8.07E-11	6.96E-05	0.071203	0.0112587	Zm00001eb228060	LOC100282640	protein BRE	Zm00001eb228070
RFP_2019_Midflowering_Full_Irrigation	5_66112461	5	68090202	7.22E-08	0.008888191	0.09019	0.262433	Zm00001eb229190	LOC100127010	a2_anthocyaninless2	NA
Root_Mass_2018_Postflowering_Limited_Irrigation	5_151631304	5	154800023	7.02E-08	0.02017197	0.195988	0.75382	Zm00001eb239530	LOC100280577	rps24_ribosomal_protein_S24	NA
RFP_2018_Earlyflowering_Limited_Irrigation	5_171562659	5	174940230	4.23E-09	0.003644646	0.346491	0.203737	Zm00001eb243430	NA	NA	NA
RFP_2018_Postflowering_Limited_Irrigation	5_199375500	5	20731813	2.10E-07	0.02897257	0.229532	0.4804843	Zm00001eb251400	LOC100193226	uncharacterized	Zm00001eb251410
RFP_2019_Midflowering_Full_Irrigation	5_204088245	5	212134701	3.58E-08	0.005034407	0.148734	0.681638	Zm00001eb253130	LOC100273249	uncharacterized	Zm00001eb253140
RFP_2018_Postflowering_Full_Irrigation	5_211939338	5	220320408	3.88E-07	0.03345518	0.088323	0.4012506	Zm00001eb256890	LOC100276172	uncharacterized	Zm00001eb256900
RFP_2019_Midflowering_Full_Irrigation	6_89809402	6	98631379	9.16E-08	0.009861784	0.254747	0.0170504	Zm00001eb273440	LOC541654	mir2_maize_insect_resistance2	Zm00001eb273450
RFP_2018_Earlyflowering_Full_Irrigation	6_10012636	6	11033672	1.93E-09	0.000553035	0.136228	0.002993	Zm00001eb276250	LOC100383495	uncharacterized	NA
RFP_2018_Earlyflowering_Full_Irrigation	6_10686807	6	161845903	4.81E-07	0.031448089	0.29351	0.26961	Zm00001eb289310	LOC100274162	uncharacterized	Zm00001eb289320
RFP_2018_Earlyflowering_Full_Irrigation	6_165623621	6	167982260	1.21E-07	0.013079861	0.056047	0.0979338	Zm00001eb291820	NA	NA	Zm00001eb291810
RFP_2018_Earlyflowering_Full_Irrigation	7_5087243	7	5253849	5.11E-07	0.031448089	0.123894	0.188734	Zm00001eb299950	LOC103631941	WPP_domain_associated_protein	NA
RFP_2018_Postflowering_Limited_Irrigation	7_137339567	7	141950448	1.74E-08	0.003743774	0.160819	0.718042	Zm00001eb317490	LOC103632986	uncharacterized	NA
Area_2019_Midflowering_Limited_Irrigation	7_143397759	7	149546828	2.23E-07	0.02474692	0.070533	0.8931777	Zm00001eb319700	LOC100280671	uncharacterized	NA
RFP_2018_Postflowering_Limited_Irrigation	7_147349220	7	153752462	4.85E-08	0.008357043	0.067251	0.7754507	NA	NA	NA	NA
Shoot_Mass_2018_Postflowering_Full_Irrigation	8_14794756	8	15239785	9.91E-10	0.000426911	0.117363	0.0007953	Zm00001eb335970	LOC103634830	glutathione_hydrolase_3	NA
Width_2018_Postflowering_Limited_Irrigation	8_27634189	8	28400131	2.76E-10	0.00023756	0.445588	0.592249	Zm00001eb339070	LOC101027215	reticulon_like_protein_B1	Zm00001eb339080
RFP_2019_Midflowering_Full_Irrigation	8_130850637	8	134449811	3.67E-08	0.005394407	0.101266	0.212141	Zm00001eb354880	LOC541960	lgl_liguleless4	NA
RFP_2018_Earlyflowering_Limited_Irrigation	8_131560348	8	135586037	2.27E-08	0.009787305	0.111111	0.29				

Supplemental Table 15. List of significant SNPs for flowering time.

Trait	Marker_SNP_v2	Chromosome	B73_v5_Position	P.value	FDR_Adjusted_P.value	SNPxtreatment_pvalue
GDDTA_2018_Full_Irrigation	1_22384346	1	22570469	3.86E-08	0.006653651	0.99931
GDDTA_2018_Limited_Irrigation	1_34595396	1	34525723	4.82E-10	8.31E-05	0.709035
GDDTA_2018_Full_Irrigation	1_34595396	1	34525723	7.91E-08	0.00852531	0.709035
GDDTA_2018_Full_Irrigation	1_234059072	1	239779155	7.90E-07	0.044034128	9.39E-01
GDDTA_2018_Full_Irrigation	1_234059181	1	239779264	5.18E-07	0.034357417	0.639911
GDDTA_2018_Full_Irrigation	1_234122303	1	239841401	7.50E-07	0.044034128	0.855298
GDDTA_2018_Full_Irrigation	1_234122368	1	239841466	8.18E-07	0.044034128	0.869005
GDDTA_2018_Full_Irrigation	1_244469214	1	250149814	4.95E-08	0.007106611	0.86884
GDDTA_2019_Full_Irrigation	1_244469214	1	250149814	7.60E-08	0.007398732	0.92121
GDDTA_2018_Full_Irrigation	1_272275573	1	278827623	1.67E-12	1.43E-06	0.802781
GDDTA_2018_Full_Irrigation	2_37419164	2	38969540	1.36E-07	0.012322113	0.985472
GDDTA_2018_Full_Irrigation	2_45967533	2	47778943	2.10E-07	0.016421007	0.75916
GDDTA_2018_Limited_Irrigation	2_152921504	2	157561349	3.45E-07	0.029717962	0.515189
GDDTA_2018_Full_Irrigation	2_152921504	2	157561349	1.43E-07	0.012322113	0.515189
GDDTA_2019_Full_Irrigation	2_194166189	2	200268685	5.45E-08	0.007398732	0.7564203
GDDTA_2018_Limited_Irrigation	3_18009555	3	18026393	2.77E-09	0.00039751	0.56572
GDDTA_2019_Full_Irrigation	3_176805743	3	181618571	3.71E-08	0.007398732	0.8431
GDDTA_2018_Full_Irrigation	3_202829071	3	208156072	6.59E-10	0.000283747	0.91432
GDDTA_2018_Limited_Irrigation	3_210522634	3	216109871	3.21E-11	1.39E-05	0.33557
GDDTA_2018_Limited_Irrigation	3_222671752	3	228285300	3.26E-09	0.00040083	0.2432276
GDDTA_2018_Full_Irrigation	3_222671752	3	228285300	1.45E-09	0.000415397	0.2432276
GDDTA_2019_Full_Irrigation	3_222671752	3	228285300	5.14E-08	0.007398732	0.9276
GDDTA_2019_Full_Irrigation	4_22329446	4	24042781	2.47E-07	0.017754335	0.86708
GDDTA_2019_Full_Irrigation	4_136058402	4	141143400	8.59E-08	0.007398732	0.89482
GDDTA_2018_Limited_Irrigation	4_142201504	4	147390958	2.02E-07	0.01930074	0.6881
GDDTA_2018_Limited_Irrigation	5_2811674	5	3128652	6.16E-07	0.048262206	0.6314
GDDTA_2019_Full_Irrigation	5_8454996	5	8970718	1.21E-08	0.003477176	0.4914411
GDDTA_2019_Limited_Irrigation	5_16851434	5	17615597	3.62E-10	0.000311745	0.5392
GDDTA_2019_Limited_Irrigation	5_54920199	5	56670866	1.82E-09	0.000410152	0.861649
GDDTA_2018_Limited_Irrigation	5_92354630	5	94465080	5.16E-11	1.48E-05	0.837323
GDDTA_2018_Limited_Irrigation	5_199138822	5	207087752	2.07E-12	1.78E-06	0.3566
GDDTA_2019_Limited_Irrigation	6_94761834	6	104558680	1.33E-09	0.000410152	0.9458
GDDTA_2019_Full_Irrigation	6_142661184	6	153795816	8.42E-08	0.007398732	0.99969
GDDTA_2019_Limited_Irrigation	6_148970518	6	160088684	2.84E-08	0.004888121	0.844
GDDTA_2019_Limited_Irrigation	6_153260064	6	164451679	5.49E-08	0.007883181	0.582
GDDTA_2018_Full_Irrigation	7_137162241	7	142051397	6.78E-08	0.008343274	0.70898
GDDTA_2019_Full_Irrigation	7_152441704	7	161442936	2.43E-07	0.017754335	0.9345
GDDTA_2018_Full_Irrigation	8_131779375	8	135384583	2.11E-09	0.000454472	0.675565
GDDTA_2018_Limited_Irrigation	8_132047205	8	135657401	3.79E-08	0.004077727	0.46263
GDDTA_2018_Full_Irrigation	8_161713363	8	168032736	5.05E-07	0.034357417	0.50173
GDDTA_2019_Full_Irrigation	8_161713363	8	168032736	7.86E-08	0.007398732	0.75648
GDDTA_2019_Full_Irrigation	9_50389650	9	54220507	1.08E-09	0.000463887	0.9994
GDDTA_2019_Full_Irrigation	9_121718186	9	127302804	3.29E-07	0.02179468	0.97423
GDDTA_2019_Full_Irrigation	9_144317637	9	150543792	5.69E-10	0.000463887	0.6514
GDDTA_2018_Limited_Irrigation	9_151977644	9	158369935	3.79E-10	8.17E-05	0.874404
GDDTA_2019_Limited_Irrigation	9_154893380	9	161266976	1.24E-07	0.015269731	0.542036
GDDTA_2019_Limited_Irrigation	10_14912510	10	15010034	1.90E-09	0.000410152	0.9725
GDDTA_2019_Full_Irrigation	10_14912510	10	15010034	4.07E-07	0.025041165	0.9725