OFFICIAL

A small number of major QTL clusters underpin cannabinoid and monoterpene diversity in *Cannabis sativa*

Rekhamani Das*[1], Erwin Tandayu*[2], Razlin Azman[3]; Nicolas Dimopoulos[1]; Paolo Miguel Siazon[1], Jos Mieog[1], Tobias Kretzschmar^[1]

*These authors contributed equally

^Corresponding author: tobias.kretzschmar@scu.edu.au

[1]Southern Cross University, Faculty of Science and Engineering, 1 Military Road, East Lismore, NSW 2480, Australia

[2]Agriculture Victoria, AgriBio, Centre for AgriBioscience 5 Ring Rd Bundoora VIC 3083, Australia

[3]CSIRO Agriculture & Food, 147 Underwood Ave, Floreat WA 6010, Australia

Keywords:

Medicinal Cannabis, Cannabinoids, Monoterpenes, Sesquiterpenes, QTL mapping, Trichome

Preprint not peer reviewed

OFFICIAL

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

A small number of major QTL clusters underpin cannabinoid and monoterpene diversity in *Cannabis sativa*

Rekhamani Das*[1], Erwin Tandayu*[2], Razlin Azman[3]; Nicolas Dimopoulos[1]; Paolo Miguel Siazon[1], Jos Mieog[1], Tobias Kretzschmar^[1]

*These authors contributed equally

^Corresponding author: tobias.kretzschmar@scu.edu.au

[1]Southern Cross University, Faculty of Science and Engineering, 1 Military Road, East Lismore, NSW 2480, Australia

[2]Agriculture Victoria, AgriBio, Centre for AgriBioscience 5 Ring Rd Bundoora VIC 3083, Australia

[3]CSIRO Agriculture & Food, 147 Underwood Ave, Floreat WA 6010, Australia

Keywords:

Medicinal Cannabis, Cannabinoids, Monoterpenes, Sesquiterpenes, QTL mapping, Trichome

Preprint not peer reviewed

OFFICIAL

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Abstract

Cannabis sativa is an ancient medicinal crop that has seen a recent resurgence in therapeutic applications. Its medicinal value is largely attributed to two chemical compound classes, cannabinoids and terpenes. Cannabinoids are unique to *Cannabis* and major cannabinoids, including delta-9-tetrahydrocannabinol (THC) and cannabidiol (CBD), provide benefit through direct interactions with the human endocannabinoid system. Terpenes are important to improve sensory aspects of *Cannabis* products, while some have further been shown to modulate the efficacy of cannabinoids. While the biochemical pathways of cannabinoid and terpene production are well characterised, knowledge on the natural variation and genetic control underpinning the respective chemotype diversity of *Cannabis* is only beginning to emerge. Here we report on quantitative trait loci (QTL) mapping for major cannabinoids and terpenes in a segregating biparental mapping population derived from a CBD-dominant and a THC-dominant landrace of contrasting monoterpene compositions. An F2 population of 155 female plants was clonally propagated and phenotyped for chemical composition of mature floral tissues in two contrasting protected cropping environments. We found that four major cannabinoids were controlled by a single locus on Chromosome 7, while six of the most abundant monoterpenes were controlled by a single locus on Chromosome 5. Expression analysis on secondary metabolite-producing glandular trichomes from both parental lines pinpointed several cannabinoid synthases and terpene synthases in proximity to the respective QTL peak markers as putative candidate genes.

Preprint not peer reviewed

OFFICIAL

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

1 Introduction

Cannabis sativa (*Cannabis*), a monotypic genus in the *Cannabaceae* family [1], has a long history as a medicinal and recreational plant [2]. *Cannabis* is predominantly dioecious and anemophilous, making it an outcrossing species of high heterozygosity [3]. Female flowers are characterized by presence of large glandular trichomes that produce an abundance of secondary metabolites [4]. The medicinal value of *Cannabis* is largely owed to the presence of cannabinoids, pharmacologically active terpenophenolic compounds that are unique to *Cannabis* [2] and exert their bioactivity through interactions with the human endocannabinoid system [5]. The aroma of *Cannabis* flowers and derived products on the other hand are attributed to the accumulation of a range of monoterpenes and sesquiterpenes in the glandular trichomes [6]. Unlike cannabinoids, terpenes are ubiquitously found in the plant kingdom and well known for their flavouring properties [2]. While differential profiles of cannabinoid and terpene composition form the cornerstone of the multibillion-dollar medicinal *Cannabis* industry [7], their genetic control is only beginning to be understood.

Cannabinoids, derived from a combination of polyketide and isoprenoid precursors [8], are the major secondary metabolite class of *Cannabis* [9]. Concentrations of up to 35% cannabinoids in dry female flowers have been reported [10, 11] and over 130 different cannabinoids have identified [12]. Major cannabinoids include C5 alkyl cannabinoids tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA) and respective decarboxylated forms tetrahydrocannabinol (THC) and cannabidiol (CBD). Common minor cannabinoids include C3 alkyl cannabinoids delta(9)-tetrahydrocannabivarinic acid (THCVA) and cannabidivarinic acid (CBDVA), as well as cannabigerolic acid (CBGA) and cannabichromenic acid (CBCA) [9, 13].

The polyketide pathway generates olivetolic acid (OLA) from hexanoyl-CoA and three molecules of malonyl-CoA through catalytic action of tetraketide synthase (TKS) [14] and olivetolic acid cyclase (OAC) [15]. OLA and geranyl pyrophosphate (GPP), produced by the plastidial methylerythritol phosphate (MEP) pathway, form the cannabinoid cannabigerolic acid (CBGA) via aromatic prenyltransferases [16, 17]. CBGA serves as a substrate for THCA synthase (THCAS), CBDA synthase (CBDAS) and CBCA synthase (CBCAS) to form THCA, CBDA and CBCA respectively, which can then be converted non-enzymatically to THC, CBD and CBC [18]. The ratio of THC to CBD has been used in clasifcation of *Cannabis*, with Type

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

I being THC-dominant drug types, Type III being CBD-dominant hemp types and Type 3 showing balanced THC and CBD content [19, 20].

Engineered yeast strains supplied with butanoyl-CoA have been reported to produce divarinic acid, which, when used as prenyltransferases substrate instead of OLA, resulted in cannabigerovarinic acid (CBGVA), along with THCVA and CBDVA [21]. While specific synthases for C3 alkyl cannabinoids THCVA and CBDVA have not been identified, genetic studies suggest a β-keto acyl carrier protein (ACP) reductase (BKR) gene to be associated with C3 alkyl cannabinoid content *in planta* [22].

Terpenes, the largest class of natural products with over 80,000 known compounds [23-25], are the other major group of secondary metabolites found in *Cannabis* [26, 27]. Terpenes, predominantly monoterpenes and sesquiterpenes, constitute 3-5% of dry-weight of female inflorescence [28]. Terpenes provide chemical defence against pathogens, predators and competitors, while also functioning as volatile messengers for conspecific and mutualist about presence of mates, food and enemies [25]. In *Cannabis* products, major monoterpenes such as myrcene, limonene, and α-terpinolene are often used in *Cannabis* product classification and labelling [6, 29].

The primary enzyme family responsible for low weight terpenes during terpene synthesis are Terpene Synthases (TPS) [30]. The plant TPS gene family comprises six subfamilies [31], of which the TPS-b subfamily is associated with monoterpene biosynthesis and the TPS-a subfamily with sesquiterpenes. Two MEP pathway-derived isoprenoid units condense into C10 geranyl diphosphate (GPP), the common subtrate for monoterpene synthases, while three mevalonate (MEV) pathway-derived C5 form C15 farnesyl diphosphate (FPP) the common substrate for sesquiterpene synthases [2]. Multi-substrate utilization by TPS can significantly alter terpene profiles in response to metabolic perturbations in stressed plants or during specific developmental stages [32], with many TPS capable of producing multiple products from a single substrate [2].

The CBD-type cultivar CBDRx chromosome-resolved genome assembly highlighted a complex array of cannabinoid synthases highlighting the introgression of CBDAS from hemp and lacks complete sequence for THCAS despite having marijuana ancestry which influences the cannabinoid expression [33, 34].

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Nine TPS-a and TPS-b subfamilies genes were identified in the 'Finola' assembly, which are responsible for the biosynthesis of sesquiterpene and monoterpene, respectively [2]. A study using 5 *Cannabis* cultivars with contrasting terpene profile characterised 33 different *Cannabis* terpene synthesis genes (CsTPS) suggesting distinct expression of cannabinoid and terpene pathway genes and identified 19 complete CsTPS gene model with isoprenoid and cannabinoid biosynthetic genes within the cv Purple Kush reference genome [35]. Seven QTLs related to cannabinoid traits, with 6 of them being major QTLs (> 15% PVE), and 28 QTLs related to terpene traits, with 14 of them being major QTLs (> 15% PVE), were identified in an F2 population created from two hemp cultivars [36]. A genome wide association study (GWAS) using 100 commercial *Cannabis* samples revealed that retail product labelling was driven by variation in key terpene contributors that were linked to tandem arrays of TPs on chromosomes 5 and 6 [29].

In this study, the contents of six cannabinoids and 10 terpenes were investigated in an F2 mapping population derived from cross between two parental lines of different biochemical profiles. Across two contrasting environments we identified two major quantitative trait loci (QTL), one affecting major monoterpenes and the other controlling major cannabinoids, alongside a few minor QTL for sesquiterpenes.

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

2 Materials and methods

2.1 Plant materials and growing conditions

Cannabis cultivation, sampling, storage and processing were performed in strict adherence to Sections 23(4)(b) and 41(b) of the NSW Drug Misuse and Trafficking Act 1985, held under the Authority granted to Prof. Bronwyn Barkla of Southern Cross University (SCU), issued by the New South Wales Ministry of Health, Australia. A diversity panel of *Cannabis* Germplasm was obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung, IPK), Germany). The "IPK collection" was imported under a federal Office of Drug Control (ODC) license to import No. 1820928 and handled under the FAO governed Standard Material Transfer Agreement (sMTA).

2.2 Phenotyping and genotyping for parental selection

Female individuals of the IPK collection were grown and harvested as previously described [37]. This diverse collection previously quantified for twelve cannabinoids [37], was quantified for ten terpenes and genotyped using the HASCH platform [38]. The aim was to generate a comprehensive dataset to guide selection of parental lines that were genetically distant and showed good variation for both target compound classes. Terpene contents of the IPK collection were quantified utilizing Gas Chromatography-Mass Spectrometry (GC-MS) as previously described [4]. Additionally, genotyping data for these diverse lines was obtained by genotyping with HASCH platform [38], providing data on 1,504 genome wide Single Nucleotide Polymorphism (SNP) markers. Methods related to chemical quantification and genotyping will be presented in more detail in sections 2.5 and 2.6 sections of materials and methods for the F2 populations.

Principal component analysis (PCA) biplots were used to explore phenotypic diversity among the 48 lines, visualize grouping patterns, and identify potential outlier lines based on terpene and cannabinoid profiles. This used z-standardized cannabinoid and terpene values.

SNP genotype data were filtered for missing rate < 0.9, heterozygosity < 0.5, and minor allele frequency of 0.05 in TASSEL (version 5.2.72) and NJ-tree was visualised using ITOLv6 (https://itol.embl.de/, accessed on 15 February 2024).

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

2.3 Establishment of the F2 population

Seeds from selected parents, Type III IPK_CAN_57 and Type I IPK_CAN_36, were germinated on paper towel using 2% hydrogen peroxide (H_2O_2) solution. MADC2 PCR primers were used for sex determination using previously published methodology [39]. A female IPK_CAN_57 individual and a male IPK_CAN_36 individual were transferred to 1.5 L pots filled with premix potting soil 70% coco, 30% perlite, 3.4 g/L osmocot exact 3-4, (NSW, Australia), and placed in airtight pollen-proof growth chamber for controlled pollination. The growth chamber is equipped with a 300W grow light (ViparSpectra©) as light source and an automated irrigation system. An initial 18:6 hour (light: dark) photoperiod was used to maintain plants in vegetative stage. After 4 weeks, the photoperiod was changed to 12:12 (light: dark) to initiate flowering. Fertilized flowers were harvested six to eight weeks after flowering, and seeds were collected by hand threshing. F1 plants were germinated and sex of the seedlings was determined as described above. Ten females were transferred to individual growth chambers for self-pollination. After 4 weeks of vegetative stage, two branches were treated with 3 mM Silver Thiosulfate Solution (STS) for induction of male flowers using previously published methodology Lubell and Brand [40]. Flowers were harvested after 6-8 weeks depending on seed maturity. Harvested flowers were tested for terpene content, and an intermediate phenotype between the parents were used to confirm hybridity. A total of 180 seeds from a single confirmed F1 was used to establish an F2 population.

2.4 F2 population growing conditions

The F2 population was grown in a controlled environment (see below for details) under 18:6hour (light: dark) photoperiod for three weeks. Complete clonal replication of the population was then generated by taking cuttings from each individual, treating them with Clonex® rooting gel (3.0g/L indole-3-acetic-ccid) and transferring it onto coco peat plugs. The original plants were retained in controlled environment and used for controlled-environment growing, whereas the clonal replicates were used outdoor-environment growing. After cuttings were taken for cloning, the original plants were maintained under vegetative growth conditions in the controlled environment for an additional three weeks before the before the photoperiod was switched to 12:12 (light: dark).

For the controlled environment, seeds were germinated in July 2022, and harvesting was completed in October 2022 at SCU Lismore campus, 2480 NSW. Plants were grown in trays containing 15 0.5 L pots filled with premix potting soil comprised of 70% coco, 30%

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

perlite, and 3.4 g/L Osmocot Exact 3-4. A 300W LED grow lights (ViparSpectra©) were used as light source. Trays were placed on elevated benches equipped with an automated capillary watering system, and temperature was maintained at 27±2 ºC. Of the 180 plants, only 144 plants had harvestable seeds at 50% pistil maturity. Harvested flowers were dried and stored in the dark at 15% relative humidity and at 15 ˚C prior to extraction and quantification of secondary metabolites.

For the outdoor environment, plants were cultivated in a production shade-house at Cymra Life Sciences growing facility located at Rous, 2480 NSW. Cuttings were transferred to the site in August 2022, and harvesting was completed in November 2022. Rooted cuttings were transplanted into 1.5 L pots containing fresh organic potting mix (Searles Organic Potting Mix Advanced, Bunnings Australia). Plants were watered manually every two to three days and without temperature control with temperature ranging 25 ± 3 ˚C at daytime and 12 ± 4 ˚C at nighttime during the growth period. Natural light, supplemented with grow lights (300W Cultiv8 UFO Solar Sun LED) were used as light source and black-out shading was used to generate dark condition for flower initiation. The population was kept under 18:6 (light: dark) for three weeks for vegetative growth, followed by 12:12 photoperiod to induce flowering. At six to seven weeks at 50% pistil maturity, flowers from 144 plants were harvested, dried and stored under same conditions as controlled environment samples.

2.5 Cannabinoid and terpene quantification

Terpene and cannabinoid contents of the F2 populations were quantified as follows: For sample preparation, 1 g of homogenized plant material was extracted with 10 mL of HPLC grade ethanol with 0.022% tridecane as internal standard. Samples were sonication at 50/60 Hz (Soniclean®, Australia) for 30 min, then centrifuged in 3400 RFC for 10 min. One mL of the supernatant was transferred to an auto-sampler (screw cap glass vial) and stored at − 20 °C until gas chromatograph with flame ionization detection (GC-FID) analysis. The supernatant was further diluted 1/40 by adding HPLC grade ethanol and stored similarly until HPLC analysis.

For cannabinoids analysis liquid chromatographic analyses were performed using an high performance liquid chromatographic system (Agilent 1260 infinity II product No. M8413AA) with Diode Array Detector (DAD) with a Agilent infinity Lab Poroshell 120 column, column type- Poroshell HPH-C18 product No. 693775-702 (2.1 x 150 mm, 2.7 µm) (Agilent technologies, Santa Clara, California, United States). Mobile phases were water

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

+ 0.1% Formic acid (A), acetonitrile (B) and Methanol (C). Separation was achieved under gradient conditions: 35%A and 50%B at 0 min, 35% A and 50% B at 0.5 min, 28% A and 60%B at 6 min, 14% A and 80% B at 11 min, 1% A and 99% B at 12 min, 1% A and 99% B at 12.5 min, 35% A and 50% B at 13 min, 35% A and 50% B at 18 min. Samples were introduced with a 1 µL injection, with chromatographic data collected at 270 nm. Cannabinoid certified reference standards (Sigma-Aldrich) were used for peak identification and generation of calibration curves used for quantitation, and included: tetrahydrocannabinolic acid (THCA), tetrahydrocannabivarinic acid (THCVA), Cannabidiolic Acid (CBDA), Cannabidivarinic Acid (CBDVA), Cannabigerolic Acid (CBGA) and Cannabichromenenic Acid (CBCA). Data were recorded and processed using Openlab Software (Version 3.6).

Analysis of terpene profiles was performed by GC-FID Chromatography using Agilent 6850A gas chromatograph with Flame Ionization Detection (FID), with a BPX5 column (SGE054117, with a maximum temperature of 360°C, capillary dimensions of 50.0 m x 220 µm x 1.00 µm). (Phenomenex, Torrance, California, United States). The temperature for the inlet was at 280 ˚C for 5 min then increased to 320 ˚C at 15 ˚C /min and held for 50 min making the total run time per sample 59 min. Hydrogen was used as the GC carrier gas at a flow rate of @ 2.2 mL/min. Hydrogen at a flow rate of @ 30 mL/min and compressed air at a flow rate of @400 mL/min were used as the combustion gases. Terpene certified reference materials (Sigma Aldric CRMs #42262, 40448, 40431, 40921, 40339, 40339, 40748, 40339, 40378, 40433 40483, 40929 and 78492-1ml-F) were used for peak identification and generation of calibration curves used for quantitation. Data were recorded and processed using Openlab Software (Version 3.6). In the manuscript, the following abbreviations are used for terpenes: β-caryophyllene (BCAR), myrcene (MYR), α-pinene (APN), β-pinene (BPN), α-terpinolene (ATPO), α-terpinene (ATPE), γ-terpinene (GTP), humulene (HUM), limonene (LIM), and farnesene (FAR).

2.6 Genotyping

Genomic DNA (gDNA) was extracted from 0.2 gm leaf samples collected from each plant of the F2 population grown in the controlled environment. DNA was extracted by using Qiagen Plant Mini Kit (250) (Cas No.69106) from Qiagen Pty Ltd following the manufacturer protocol. Quantity of the DNA was checked using nanodrop and Quality of the DNA was check by running the DNA on 0.8% agarose gel. Genomic DNA (30 ng) of 48 diverse panel plants and

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

151 F2 plants were sent to Diversity arrays technology (DArT) for genotyping using the HASCH platform [38].

2.7 QTL mapping

The genetic map construction and QTL mapping were conducted using a combination of the R/qtl [41] and ASMap packages [42]. HASCH genotyping data for the F2 population was analysed for low call rates and pairs of unusually similar genotype data using R/qtl. Markers that were monomorphic or had sample genotyping call rates of less than 90% were excluded. The resulting genotype input was further refined in ASMap, where it was filtered for several parameters including segregation distortion, identical genotype data, evidence of genotyping error [43], and markers with duplicated information due to co-location on the same position.

The construction of the linkage map was performed with R/qtl, employing an estimated recombination fraction maximum of 0.35 and a minimum LOD score of 6.0 for placing two markers in the same linkage group. Inter-marker distances in centimorgans (cM) were calculated using the Kosambi function. The final genetic map was re-estimated using the Lander-Green algorithm [44] through the est.map function of R/qtl. This genetic map was then utilized for QTL mapping of cannabinoid content, analysing log-transformed percent values in an F2 population comprising 121 individuals with detectable cannabinoid and terpene contents.

The analysis included single QTL model using Haley-Knott regression, two-dimensional scans, and multiple QTL analyses. For single-QTL analysis, a density of 1 cM was used, while the two-dimensional scan utilized 2 cM. The final QTL model was derived from the "stepwiseqtl" analysis. The 95% Bayes credible intervals around the maximum likelihood estimate of the QTL location were estimated with the "bayesint" function. The proportion of phenotypic variance explained by the QTL was calculated using the "fitqtl" function. Finally, the genetic and QTL maps were illustrated using MapChart [45].

2.8 Transcriptome analysis of flower and trichome tissues

Clones of the female parent IPK_CAN_57 and clones of a female sister line of the male parent IPK_CAN_36 were established and grown as previously described [4]. Thirty (30) days after flower induction, flower samples were collected and trichome were extracted from four biological replicates (individual plants) of both lines, following Dimopoulos, Guo [4] method. Trichomes samples were flash frozen and stored at -80°C until required for RNA isolation.

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

RNA was extracted from 50 mg samples following the protocol provided by the Qiagen RNeasy Mini Kit, to prepare for downstream RNA-seq and qPCR applications. Purity of RNA samples was determined through Nanodrop quality check and Bioanalyzer 2100 to determine RNA integrity number (RIN).

Sequencing was outsourced to the Australian Genome Research Facility (AGRF), using the Illumina NovaSeq 6000 Single end (100bp) flow cell. Reads processing and mapping against the reference CBDRx genome was performed using CLC Genomics Workbench version 11.0.1. Expression values in transcript per million (TPM) were exported for comparison in four replicates of the trichome isolates of IPK_CAN_57 and IPK_CAN_36.

2.9 Blast-based identification of terpene synthase genes

Terpene synthase protein sequences previously characterised in the *Cannabis sativa* Finola cultivar were obtained (Booth et al 2017, Booth et al. 2020). These Finola TPS protein sequences were used as queries to identify corresponding genes in the *C. sativa* CBDRx genome. Sequence similarity searches were performed against the CBDRx protein dataset using DIAMOND (v2.1.16) in BLASTP mode (default parameters; e-value ≤ 1e-3). For each query, the top scoring hit was retained as the best-matching homolog in CBDRx. The corresponding CBDRx locus identifiers were used for downstream analyses.

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

3 Results

3.1 Chemotypic variation of the IPK collection and parental selection

A diverse set of 48 IPK lines previously quantified for twelve cannabinoids [37], was additionally quantified for ten terpenes and genotyped using the HASCH platform [38]. The aim was to generate a comprehensive dataset that allowed for selection of parental lines that were genetically distant and showed good variation for both target compound classes.

A Principal Component Analysis (PCA) biplot for cannabinoid content in the diverse set **(Figure 1a)** showed that THCA content had the most influence on PC1 while CBDA variation was captured by PC2, resolving the samples into chemotype clusters of high THC drug-type (Type I) and high CBD hemp-type (Type III). IPK_CAN_36 was identified as an outlier for the high THC drug-type in extreme right of first quadrant, while IPK_CAN_57 and IPK_CAN_64 were identified as two extremes for CBDA content in the second quadrant. Biplot trend is supported by hierarchical clustering **(Supplementary Figure 1)** with the IPK_CAN_36 and other high-THC accessions (IPK_CAN_23, IPK_CAN_58 and IPK_CAN_41) forming a distinct cluster separated at the first node attributable to their high THC and THCA contents. IPK_CAN_57 on the other hand demonstrated the highest CBDA content and was found on a different cluster together with IPK_CAN_26 and IPK_CAN_64.

PCA biplot using the ten main terpene **(Figure 1 b)** suggested that much of the clustering and variation in PC1 were attributable to α-pinene, β-pinene and myrcene, and α-terpinolene contents, the four dominant monoterpenes present in 90% of the diverse set. Sesquiterpenes β-caryophyllene and humulene exerted the greatest influence in delineating samples in PC2. For monoterpenes, IPK_CAN_59 was an outlier for α-pinene, while IPK_CAN_70 was an outlier for α-terpinolene. IPK_CAN_20 and IPK_CAN_68 were sesquiterpenes outliers accumulating high levels of β-caryophyllene and humulene. IPK_CAN_36 localized to the third quadrant indicating lower monoterpene content, while IPK_CAN_57 localized to the fourth quadrant associated with higher monoterpene content, particularly myrcene (MYR), α-terpinolene (ATPO), α-terpinene (ATPE) and γ-terpinene (GTP). Hierarchical clustering separated the monoterpene outliers, IPK_CAN_70 and IPK_CAN_59 from the rest of the samples at the first node **(Supplementary Figure 2),** while sesquiterpene outliers IPK_CAN_68 and IPK_CAN_20, formed a distinct cluster together with two other samples.

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

A neighbour-joining (NJ) tree broadly reflected the hierarchical clustering of the cannabinoid contents **(Figure 2)**. Extremes for THCA and CBDA clustered with their co-members in the phylogeny. IPK_CAN_36 and IPK_CAN_57, which showed most divergence in respect to cannabinoid content were located on separate clusters **(Figure 2).** In contrast, extremes for terpene classes were not visibly reflected in the NJ tree.

To maximize both chemotypic and genetic divergence, IPK_CAN_36 and IPK_CAN_57 were selected as parents for a bi-parental mapping population. Direct comparison of cannabinoid and terpene contents between IPK_CAN_36 and IPK_CAN_57 further supported their chemotypic differences **(Figure 3)**. In addition to contrasting major C5 cannabinoids, they differed in minor C3 cannabinoids. IPK_CAN_36 contained no detectable CBDVA, while it was relatively abundant in THCVA. Contrastingly, IPK_CAN_57 contained no detectable THCVA, while containing low amounts of CBDVA.

IPK_CAN_36 and IPK_CAN_57 showed differences in the five most common monoterpenes within the IPK collection, namely myrcene (MYR), α-pinene (APN), α-terpinolene (ATPO), β-pinene (BPN) and limonene (LIM) **(Figure 3b)**. Interestingly IPK_CAN_36 was found to be completely devoid of α-terpinolene (ATPO), α-terpinene (ATPE) and γ-terpinene (GTP). Differences for the common sesquiterpenes β-caryophyllene and humulene were less pronounced **(Figure 3b)**.

3.2 Phenotyping of the F2 population

Comparison of phenotypes of the different compounds for both environments revealed greater range and variability for both terpene and cannabinoid content and composition in the outdoor shade-house environment **(Figure 4)**. Notably, plants grown in the outdoor environment had higher median for CBDA and CBCA concentration, while plants in the controlled-environment exhibited higher median levels of CBGA and to a lesser extent THCA. THCVA and CBDV remained low in concentrations and showed limited variation across growing conditions. In contrast, plants grown in outdoor shade-environment had higher median values for all terpenes. Furthermore, within-environment variability is notable in the terpene profiles. Correlation analysis using the controlled environment data **(Figure 5)** showed a strong positive correlation between ATPE, GTP, APN and BPN, while THCA and CBDA were strongly negatively correlated.

.

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

3.3 QTL mapping

QTL mapping for major cannabinoid and terpene was carried out for the population in both environments. Different methods employed in R/qtl, namely, single-QTL analysis, two-dimensional scan and multiple QTL analysis yielded similar results and most QTLs were consistently identified with data obtained from either environment **(Table 1, Supplementary Table 1)**.

Table 1 presents the QTL analysis based on data generated from the controlled environment. The optimal QTL models indicate that both cannabinoid **(Figure 6)** and terpene **(Figure 7)** accumulation are controlled by a small number of QTLs in this population. For the four main cannabinoids—CBDA, THCA, CBGA and THCVA, a single QTL was identified for each trait (qCDBA7, qTHCA7, qCBGA7 and qTHCVA7) with peak markers mapped at ~21-24 cM of the CBDRx genome. Peak LOD markers of CBDA and CBGA were mapped at ~21cM, whereas THCA and THCVA were mapped at 24cM. These QTLs explained a substantial proportion of phenotypic variation (PVE), with ~ 60% PVE for THCA and CBDA, and ~36% for THCVA and CBGA. Notably, these QTLs were consistently detected in the same genomic region using phenotype data from the outdoor- environment **(Table 1, Supplementary Table 1)**, although LOD scores for THC and CBD were higher in the outdoor shade-house environment. The primary difference between environments were the detection of additional four minor QTLs for CBGA in chromosomes 3, 8, and 10, spanning broader genetic intervals in the outdoor dataset. **(Supplementary Table 1)**. Among the cannabinoids measured, only CBCA exhibited a multiple QTL model, with loci mapped to chromosomes 3 (qCBCA3), 4 (qCBCA4) and 9 (qBCA9), collectively explaining ~42% CBCA variance. All the detected cannabinoid QTLs were additive with no interactions detected in our analysis.

For monoterpenes, a major QTL cluster was detected on chromosome 5 at ~16.9 cM. Six QTLs (qAPN5, qBPN5, qATPE5, qATPO5, qGTP5 and qLIM5) for the six of the ten terpenes measured— α-pinene, β-pinene, α-terpinene, α-terpinolene, γ-terpinene, and limonene had a major peak in this region **(Figure 7)**. This monoterpene QTL cluster was stable, as it was consistently detected for all of six terpenes across both environments **(Table 1, Supplementary Table 1)**. Except for β-pinene for which two QTLs were detected, the single QTLs in this cluster constituted the most optimal QTL models. These single QTL models returned the highest LODs (24.03 – 33.5) among terpene analyses and accounted for highest trait variances in the population ranging from 47% to 66% PVE. Additionally, this

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

monoterpene QTL cluster was only 2cM away from a major QTL detected for myrcene (qMYR5) at ~19cM. Myrcene, the most predominant among the terpenes measured, returned a four-QTL model. Aside from the major QTL for myrcene explaining ~43% of variance described on chromosome 5, three more minor QTLs were mapped at chromosome 3 (qMYR3), 4 (qMYR4) and 7 (qMYR7). These four QTLs for myrcene explain a total of ~63% of variance. While most of these QTLs were also detected in the outdoor shade-house environment, qBPN7, qCAR5, qFAR8, all humulene QTLs, qMYR3, qMYR4 and qMYR7 were only detected in the controlled indoor environment.

The sesquiterpene β-caryophyllene had the most complex QTL model with six QTLs mapped—two distantly located at chromosome 3 (qCAR3.1 and qCAR3.2), and one each on chromosome 5 (qCAR5), chromosome 6 (qCAR6), chromosome 8 (qCAR8), and chromosome 9 (qCAR9). Except for qCAR6, all these were minor QTLs (PVE < 15%) and together account ~55% of trait variance. A two QTL model was most optimal for the sesquiterpene farnesene, comprised of a major QTL (qFAR9) on chromosome 9 accounting for 18% PVE and a minor QTL (qFAR8) accounting for ~10% PVE. The sesquiterpene humulene also mapped a multiple QTL model, one major QTL at chromosome 3 (qHUM3) and three others minor QTLs at chromosome 5 (qHUm5), chromosome 6 (qHUM6), and chromosome 9 (qHUM9) accounting for ~45% humulene variation.

Marker trait association between the chromosome 5 peak marker NC044374.1_2380832 and the three monoterpenes absent in IPK_CAN_36, α-terpinene, α-terpinolene and γ-terpinene, suggested that the homozygous IPK_CAN_36 allele (BB) was indeed associated with absence of these compounds, while heterozygous (AB) and homozygous IPK_CAN_57 (AA) alleles associated with high concentrations **(Supplementary Figure 3)**.

3.4 QTL candidate genes for the major single locus for cannabinoid and terpene content

The cluster of single major QTLs for key cannabinoids (THCA, CBDA and THCVA) on chromosome 7 at 19.9-37.8Mb overlapped with the region harbouring the cannabinoid synthase genes and pseudogenes tandemly arrayed at 26 and 29Mb and a solitary synthase at 31Mb (Grassa et al. 2021). Cannabidiolic acid synthase LOC115697762 was highly expressed in IPK_CAN_57 trichomes, while being lowly expressed in IPK_CAN_36 trichomes (>360-fold lower). THCAs-like LOC115697880 was moderately expressed in IPK_CAN_36

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

trichomes, while being expressed 17-fold lower in IPK_CAN_57 trichomes. CBDAs-like LOC115696884 moderately expressed in IPK_CAN_36 trichomes, and lowly expressed in IPK_CAN_57 trichomes.

Similarly, the overlapping single monoterpene QTL region in chromosome 5 contained a cluster of annotated terpene synthases genes. Within 1Mb of peak marker of the QTL cluster, LOC115716405, LOC115716066, LOC115716604 and LOC115716605 were located. These four synthases were moderately expressed in trichomes of the IPK_CAN_57, while lowly expressed in IPK_CAN_36 (**Figure 8**). BLAST searches against characterised Cannabis Terpene Synthases (CsTPS) [2,7], identified LOC115716405 as highly homologous to the β-myrcene synthase CsTPS3FN (91.9% identity). LOC115716066 showed high homology with CsTPS1FN (77% identity) and CsTPS2FN (62.4%). LOC115716605 and LOC115716604 showed good homologies with CSTPS6FN, CsTPS11FN and CSTPS36.

Extending to further 2Mb up- and downstream found a genomic region enriched with four additional terpene synthases, terpene synthase 10-like isoform X1 LOC115715915, terpene synthases 10 LOC115715811 and LOC115716805, and terpene synthase 10-like isoform X4 LOC115716806. However, only LOC115715811 was lowly expressed in IPK_CAN_57, while LOC115716805 was expressed at basal levels. Neither were expressed at detectable levels in IPK_CAN_36 trichomes. Both LOC115715811 (98.3% identity) and LOC115716805 (90% identity) closely matched with CsTPS13PK, a β-ocimene synthase.

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

4 Discussion

The selection of parents for population development of our QTL mappings efforts were based on chemical phenotypes **(Figure 1a;1b)** and genetic distance **(Figure 2)**. Type I IPK_CAN_36 and Type III IPK_CAN_57 were on opposite extremes for major THCA and CBDA contents (**Figure 3a**), which was further reflected by their clear phylogenetic separation. While classification of Type I and Type III cultivars only depends on their THC/CBD content, selection and cultivation history tend to make them distinguishable by genetic relationships [38]. Moreover, IPK_CAN_36 and IPK_CAN_57 differed in respect to major monoterpenes. While IPK_CAN_36 had higher levels of MYR and LIM, it had lower levels of APN and BPN and did not contain any detectable levels of ATPO, GTP and ATPE **(Figure 1b and Figure 3b)**. Collectively this suggested them to be good parental lines for an F2 population that would be segregating for both major cannabinoids and monoterpenes.

The F2 phenotype data confirmed this hypothesis **(Figure 4)**, showing good variation for target compounds except for CBDVA, which was largely absent from the controlled environment data. Overall, the cannabinoid and terpene phenotype data were more variable for the outdoor shade-house setting than the controlled environment data, suggesting that secondary metabolite accumulation was responsive to environmental factors including abiotic and biotic stress, which were likely more pronounced in the shade-house. This aligns with a number of studies on other plant species showing that environmental effects like temperature fluctuation, water stress, as well as pest and disease pressure significantly impact the secondary metabolite production [46-49]. However, this difference in variation did not seem to bias QTL mapping as all major QTLs for monoterpenes and cannabinoids were found to be robust across both environments **(Table 1 and Supplementary table 1)**. Consistency of QTLs in both environments suggested a strong genetic control of cannabinoid and monoterpene accumulation in *Cannabis*, which aligns with previous findings [29, 33, 36] and increased confidence in these loci.

The QTLs for cannabinoid content detected on chromosome 7 were considered major (PVE > 15%), which was also the case for the monoterpene QTLs located on chromosome 5 **(Table 1 and Supplementary table 1).** We found single QTL models for CBGA, CBDA, THCA and THCVA that were mapped to a short genetic distance, highly likely the same genetic locus (**Figure 6**). The region is consistent with the single large effect QTL (92% PVE) for CBD/THC ratio [33] that contains a cluster of THCA and CBDA synthases. Differential

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

activity of THCAS and CBDAS within this single QTL can explain variation for CBGA, CBDA and THCA, as GBGA is the precursor to both THCA and CBDA. To explain control of variation of THCVA through this chromosome 7 cluster, it is reasonable to assume that THCAS is promiscuous, accepting both CBGA, derived from olivetolic acid and GPP, and CBGVA, derived from divarinic acid as co-substrates with GPP. Alternatively, a different cannabinoid synthase within the cluster on chromosome 7 could be responsible. CBDVA was not mapped as only 13 out of the 133 plants comprising the F2 population had detectable levels of CBDVA in the controlled indoor environment **(Figure 4a)**. Only CBCA QTLs did not overlap with any of the other cannabinoid QTLs (**Figure 6**). CBCA is produced from GBGA via CBCA synthase (CBCAS) [50], which was found in proximity to Cannabinoid synthase cluster on chromosome 7. A recent study suggested multiple CBCAS isoforms [51], of which two were proposed to be functional. However, our data suggested that CBCA content was associated with different linkage groups. CBCA QTLs were mapped to relatively large genetic distances on chromosomes 3, 4 and 9. While they explained lower trait variances **(Table 1 and Supplementary table 1)**, they were consistently detected in both environments showing similar size of effect sizes and position.

The single QTL models for CBDA and THCA were uncovered despite allowing for more QTLs to be detected by setting the argument max.qtl = 5 in our analysis. Grassa *et al.* (2021) [33] uncovered multiple QTLs for CBDA (chromosome 6 and 7) and THC (chromosomes 3,7,9) using a similar cross of a Type III - Carmen - and a Type I - Skunk#1 - derived population mapped against CBDRx genome. Similarly, using fibre cultivar Carmagnola and grain and fibre cultivar US031, Woods *et al.* (2021) Woods, Campbell [36] mapped multiple QTL models for CBDA and THCA with epistatic interactions accounting for smaller trait variances. Compared to these two studies, our single QTL models for CBDA and THCA accounted for higher trait variances than any of the biggest effect QTLs in their studies, or summation of the explained variances of their individual QTLs in the Grassa *et al.* (2021) [33] study and comparable total variances to that obtained by Woods *et al.* (2021) [36]. Thus, our data suggests that by optimizing parental combination, selecting greatest divergence ensured maximised variation in CBDA and THCA content, the single locus model for THCA and CBDA content was supported. This finding is consistent with legacy results of single locus of inheritance governing CBDA/THCA ratio [33, 52].

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Trichome head transcriptome data for IPK_CAN_57 and IPK_CAN_36 provided expression levels of cannabinoid synthase genes within the CBDA QTL relative to the CBDRx genome. The high expression of CBDAS LOC115697762 suggested that it is solely responsible for CBDA synthesis **(Figure 8a)** of our population. This result aligns with the findings of Hurgobin *et al.* (2021) [34] that the high CBD cultivar Canna Tsu had higher expression of LOC115697762 compared to high-THC cultivars counterparts, based on trichome transcriptome data of Zager *et al.* (2019) [53]. To account for the trace amount of THCA still present in IPK_CAN_57, we theorise that the expression of THCAs-like LOC115697880 may play a role. The considerable expression of LOC115697880 in the IPK_CAN_36 trichomes could be due to the lack of functional THCAs gene in the CBDRx genome [34], used in mapping our transcriptome. Hurgobin *et al.* (2021) [34] theorised that high-THC cultivars would contain one functional THCAs gene, but the reads for this functional THCA gene would map to the THCAs-like LOC115696986 and LOC115697880 in CBDRx genome.

Our terpene analysis mapped a QTL cluster for six of the monoterpenes, APN, BPN, ATPE, ATPO, GTP and LIM on chromosome 5 (~16.9cM) **(Figure 7, Table 1)**, indicating a shared TPS or TPS tandem array in their biosynthesis in proximity to the peak marker at 2.3Mb. Noting that IPK_CAN_36 was devoid of ATPE, ATPO, and GTP, phenotype plots against alleles at the peak marker and additive and dominance effect estimates demonstrated the IPK_CAN_57 allele to act dominantly for the three respective QTLs of these three terpenes **(Supplementary Figure 3)**. Strong positive correlations between APN, BPN, ATPE and GTP as well as weaker correlations of these with limonene **(Figure 5)** for the F2 population in the controlled environment further support co-inheritance of this locus. Interestingly, ATPO showed weak negative correlation with the other five monoterpenes in the F2 population.

The major QTL for myrcene (MYR), the predominant terpene in both parents, was tightly linked to this cluster at 19 cM and can be regarded as the same locus, an interpretation supported by the positive correlation of MYR with most of the others in this group although not with LIM, where there was a moderate negative correlation (**Figure 5**). The other myrcene QTLs on chromosomes 3, 4 and 7 were considered less reliable owing to large genetic distances with minor effects.

In the CBDRx genome, a cluster of eight TPS gene models were found in proximity to the monoterpene multi-QTL peak marker at 2.3 Mb (NC_0044374.1). This consistent multi-QTL peak marker located to the genomic region of the CBDRx genome that had previously

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

been associated with aromas in *Cannabis* in a GWAS [29]. Watts *et al.* (2021) [29] suggested that the two identified GWAS peaks at NC_004374.1_1348048 and NC_004374.1_2546403 represented two independent LD blocks possessing two independent TPS gene clusters. Indeed, we found a cluster of four TPSs at 1.3 Mb and a cluster four TPS at 2.0-2.5 Mb **(Figure 8b)**. In the 1.3 Mb cluster, two terpene synthase 10 LOC115715811 and LOC115716805 were expressed in IPK_CAN_57, while none were expressed in IPK_CAN_36. Both gene models closely matched with CsTPS13PK (98.3% and 90% AA identity Supplementary table x)) which mainly produced β-ocimene in a heterologous system [2]. However, β-ocimene was not found at detectable levels in either parent making CsTPS13PK an unlikely orthologue of LOC115715811 and LOC115716805. CsTPS13PK is part TPS b2 clade and closely related to CsTPS33PK. Both LOC115715811 and LOC115716805 shared relatively high sequence identity with CsTPS33PK as well, which was shown to generate α-terpinene (61%) and γ-terpinene (39%) from GPP [2]. Given the location of LOC115715811 and LOC115716805 within *qATE5* and *qGTE5,* the absence of expression of LOC115715811 and LOC115716805 in IPK_CAN_36 and the absence of α -terpinene and γ-terpinene in IPK_CAN_36 it is likely that LOC115715811 and/or LOC115716805 code for CsTPS33PK orthologues.

All the TPSs at 2.0-2.5Mb cluster were expressed in both parental lines, although IPK_CAN_57 expressions were ten-fold to thirty-fold higher than IPK_CAN_36 for all four TPSs **(Figure 8b)**. These collective difference in the two clusters of TPSs between the parental may explain the variation in terpene contents and profile in the two lines. Woods *et al.* (2021) [36], using the Finola as reference assembly in his QTL mapping for terpenes in *Cannabis*, found QTLs for all five terpenes measured converged to a single region in chromosome 9. However, according to Hurgobin *et al.* (2021) [34], Finola chromosome 9 is not corresponding to CBDRx chromosome 5 and thus the Woods *et al.* [36] QTL is likely independent of our monoterpene multi QTL.

LOC115716066, annotated as a putative limonene synthase, was the most expressed in IPK_CAN_57 and the highest differentially expressed TPS between both parents (**Figure 8b**). However, limonene was of low abundance in IPK_CAN_57 and significantly more abundant in IPK_CAN_36 (Figure 3b), making it an unlikely product of the protein corresponding to LOC115716066 expression. While it did share highest homology of 77% AA identity with the limonene synthase CsTPS1FN, it also showed 62.4% AA identity with CsTPS2FN, an α -

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

pinene synthase. With α -pinene the most abundant monoterpene in IP_CAN_57 at nearly twice the abundance as observed for IPK_CAN_36, α -pinene is the more likely product for this TPS.

The second highest expressed, LOC115716405, was annotated as a myrcene synthase, which was supported by high homology (92.9% AA identity) with the validated myrcene synthase CsTPS3FN. Myrcene was further found in high abundance in both parents. The higher abundance of myrcene in IPK_CAN_36 despite an about nine-fold lower expression of LOC115716405 in IPK_CAN_36 can be explained by the presence of other myrcene QTLs in this population.

While the differentially expressed LOC115716604 and LOC115716605 were also annotated as myrcene synthases in CBDRx, they are more likely responsible for the observed differences in α–terpinolene. Both show high homologies to CsTPS6FN, CsTPS11FN and CsTPS36. CsTPS6FN and CsTPS11FN are validated as β-ocimene synthases [2], but β-ocimene was not present in either parent. While the product of CsTPS36 remains unknown it shares the same active AA residues as CSTPS37 [54], which had previously been characterised as an α–terpinolene synthase [55].

Taken together, it is likely that expression differences for these gene models are responsible for the observed presence/absence of α–terpinolene, α -terpinene and γ-terpinene in the IPK_CAN_36 parent **(Figure 3 b)** and progeny carrying the homozygous IPK_CAN_36 allele for the peak marker **(Supplementary Figure 3)**. Furthermore, differences in LOC115716066 expression could be responsible for differences in α-pinene abundance explained by qAPN.

Single QTL governing over the abundance of multiple terpenes is a common occurrence in plants. In a GWAS for tea tree (*Melaleuca alternifolia*), a single long-LD genomic region harbouring a large tandem cluster of TPS genes was significantly associated to a range of terpenes [56]. Similarly, in eucalyptus (*Eucalyptus polybractea*) a monoterpene synthase cluster was found linked to variations of various monoterpene concentrations [57]. Tandem duplication with succeeding sub- or neo-functionalization are thought to be responsible for these large TPS clusters [58, 59] and we speculate this to be the case in our study as well. These gene clusters have limited recombination breakpoints due to their close proximity as well as structural features inhibiting recombination [56].

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

β-pinene (BPN) had a second major QTL peak on chromosome 7 at 23.6 cM, albeit mapped to a long genetic distance (**Figure 7**). It overlapped with the cannabinoid QTLs on chromosome 7 (**Figure 6**), with the peak LOD marker more tightly linked to the location of the cannabinoid THCA and THCVA QTL. This may explain the moderate to strong positive correlation of β-pinene (BPN) with THCVA and THCA, respectively **(Figure 5)**. β-pinene is among the four most predominant terpene in both the parental lines and its possible association with the THCA and THCVA could indicate co-selection in the high THC parent.

As opposed to the single major QTL region for most of the monoterpenes, the sesquiterpene QTLs were numerous for each compound, scattered across five chromosomes and did not overlap with each other. Except for one humulene QTL at chromosome 3, and one caryophyllene QTL at chromosome 6, sesquiterpene QTLs were mostly of minor effect, explaining 10% or less of the observed variation and were not stable across environments. Collectively, this suggested that genetic control of sesquiterpene accumulation is more complex. However, it must be kept in mind that sesquiterpenes accumulate early on in trichome development [55], while monoterpenes accumulate later closer to maturity and harvest time. It is therefore likely that sesquiterpene contents vary considerably due volatilization throughout reproductive development, creating a bias that might have negatively affected QTL mapping precision. Thus, for precise mapping of sesquiterpene content, it should be considered to quantify them several weeks earlier than monoterpenes or cannabinoids.

In conclusion, we found that, in our population derived from the Type I landrace IPK_CAN_36 and the Type III landrace IPK_CAN_57, four major cannabinoids were controlled by a single major multi-QTL cluster on c 7 and six major monoterpenes were controlled by a single major multi-QTL cluster on chromosome 5. In contrast, the three major sesquiterpenes were controlled by multiple small effect QTLS that are not overlapping and scattered across five chromosomes.

Authorship contribution

Rekhamani Das: Conceptualization, Methodology, Visualization, Writing—original draft; Erwin Tandayu: Conceptualization, Methodology, Visualization, Writing—original draft; Razlin Mohd Azman Halimi: Methodology; Nicolas Dimopoulos: Methodology; Paolo Miguel Siazon: Methodology, Visualization; Jos Mieog: Conceptualization, Supervision,

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Writing—review & editing; Tobias Kretzschmar: Conceptualization, Methodology, Supervision, Writing—review & editing.

Funding

This work was co-funded by Southern Cross University and Cymra Life Sciences.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Simon Pettinger from Cymra Life Sciences for supporting this work through industry co-supervision of Rekhamani Das, providing space in the Cymra Life Sciences shade house growth facilities and value feedback on the experimental work. The authors would further like to thank Adam Burn for always being available to help out with controlled environment and laboratory work.

References

1. Small, E. and A.J.T. Cronquist, *A practical and natural taxonomy for Cannabis.* 1976: p. 405-435.
2. Booth, J.K., J.E. Page, and J. Bohlmann, *Terpene synthases from Cannabis sativa.* PLoS ONE, 2017. **12**(3): p. 1-20.
3. Barcaccia, G., et al., *Potentials and challenges of genomics for breeding cannabis cultivars.* Frontiers in plant science, 2020. **11**: p. 573299.
4. Dimopoulos, N., et al., *From dawn 'til dusk: daytime progression regulates primary and secondary metabolism in Cannabis glandular trichomes.* Journal of Experimental Botany, 2024: p. erae148.
5. Zou, S. and U. Kumar, *Cannabinoid Receptors and the Endocannabinoid System: Signaling and Function in the Central Nervous System.* Int J Mol Sci, 2018. **19**(3).
6. Oswald, I.W.H., et al., *Minor, Nonterpenoid Volatile Compounds Drive the Aroma Differences of Exotic Cannabis.* ACS Omega, 2023. **8**(42): p. 39203-39216.
7. Booth, J.K. and J.J.P.S. Bohlmann, *Terpenes in Cannabis sativa–From plant genome to humans.* 2019. **284**: p. 67-72.
8. Fellermeier, M., et al., *Biosynthesis of cannabinoids: incorporation experiments with 13C-labeled glucoses.* European Journal of Biochemistry, 2001. **268**(6): p. 1596-1604.
9. Stack, G.M., et al., *Cannabinoids function in defense against chewing herbivores in Cannabis sativa L.* Horticulture Research, 2023. **10**(11): p. uhad207.
10. Smith, C.J., et al., *The phytochemical diversity of commercial Cannabis in the United States.* PLoS one, 2022. **17**(5): p. e0267498.
11. Geweda, M.M., et al., *Evaluation of dispensaries' cannabis flowers for accuracy of labeling of cannabinoids content.* Journal of cannabis research, 2024. **6**(1): p. 11.

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

12. Lange, B.M. and J.J. Zager, *Comprehensive inventory of cannabinoids in Cannabis sativa L.: Can we connect genotype and chemotype?* Phytochemistry Reviews, 2022. **21**(4): p. 1273-1313.
13. Tanney, C.A., et al., *Cannabis glandular trichomes: A cellular metabolite factory.* Frontiers in Plant Science, 2021: p. 1923.
14. Kearsey, L.J., et al., *Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases.* The FEBS Journal, 2020. **287**(8): p. 1511-1524.
15. Kovalchuk, I., et al., *The Genomics of Cannabis and Its Close Relatives.* 2020. **71**: p. 713-739.
16. Kajima, M. and M. Piraux, *The biogenesis of cannabinoids in Cannabis sativa.* Phytochemistry, 1982. **21**(1): p. 67-69.
17. Fellermeier, M. and M.H. Zenk, *Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol.* FEBS letters, 1998. **427**(2): p. 283-285.
18. van Bakel, H., et al., *The draft genome and transcriptome of Cannabis sativa.* Genome Biology, 2011. **12**(10): p. R102-R102.
19. McPartland, J.M. and E. Small, *A classification of endangered high-THC cannabis (Cannabis sativa subsp. indica) domesticates and their wild relatives.* PhytoKeys, 2020. **144**: p. 81-112.
20. Birenboim, M., et al., *Multivariate classification of cannabis chemovars based on their terpene and cannabinoid profiles.* Phytochemistry, 2022. **200**: p. 113215.
21. Luo, X., et al., *Complete biosynthesis of cannabinoids and their unnatural analogues in yeast.* Nature, 2019. **567**(7746): p. 123-126.
22. Welling, M.T., et al., *An extreme-phenotype genome-wide association study identifies candidate cannabinoid pathway genes in Cannabis.* Sci Rep, 2020. **10**(1): p. 18643.
23. Christianson, D.W., *Structural and chemical biology of terpenoid cyclases.* Chemical reviews, 2017. **117**(17): p. 11570-11648.
24. Powder-George, Y., *Terpenoids*, in *Pharmacognosy*. 2024, Elsevier. p. 253-294.
25. Gershenzon, J. and N. Dudareva, *The function of terpene natural products in the natural world.* Nature Chemical Biology, 2007. **3**(7): p. 408-414.
26. Jin, D., et al., *Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes.* Scientific Reports, 2020. **10**(1): p. 3309.
27. Chacon, F.T., et al., *Secondary Terpenes in Cannabis sativa L.: Synthesis and Synergy.* Biomedicines, 2022. **10**(12).
28. Nuutinen, T., *Medicinal properties of terpenes found in Cannabis sativa and Humulus lupulus.* European Journal of Medicinal Chemistry, 2018. **157**: p. 198-228.
29. Watts, S., et al., *Cannabis labelling is associated with genetic variation in terpene synthase genes.* 2021. **7**(10): p. 1330-1334.
30. Tholl, D., *Terpene synthases and the regulation, diversity and biological roles of terpene metabolism.* Current Opinion in Plant Biology, 2006. **9**(3): p. 297-304.
31. Jiang, S.Y., et al., *A Comprehensive Survey on the Terpene Synthase Gene Family Provides New Insight into Its Evolutionary Patterns.* Genome Biol Evol, 2019. **11**(8): p. 2078-2098.
32. Pazouki, L. and Ü. Niinemets, *Multi-substrate terpene synthases: their occurrence and physiological significance.* Frontiers in Plant Science, 2016. **7**: p. 1019.
33. Grassa, C.J., et al., *A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana.* New Phytologist, 2021. **230**(4): p. 1665-1679.
34. Hurgobin, B., et al., *Recent advances in Cannabis sativa genomics research.* New Phytologist, 2021. **230**(1): p. 73-89.
35. Booth, J.K., et al., *Terpene synthases and terpene variation in Cannabis sativa.* 2020. **184**(1): p. 130-147.
36. Woods, P., et al., *Quantitative trait loci controlling agronomic and biochemical traits in Cannabis sativa.* Genetics, 2021. **219**(2).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

37. Gloerfelt-Tarp, F., et al., *Using a global diversity panel of Cannabis sativa L. to develop a near InfraRed-based chemometric application for cannabinoid quantification.* Scientific reports, 2023. **13**(1): p. 2253.
38. Mansueto, L., et al., *HASCH-A high-throughput amplicon-based SNP-platform for medicinal cannabis and industrial hemp genotyping applications.* BMC genomics, 2024. **25**(1): p. 818.
39. Mandolino, G., et al., *Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.).* Theoretical and applied genetics, 1999. **98**: p. 86-92.
40. Lubell, J.D. and M.H. Brand, *Foliar Sprays of Silver Thiosulfate Produce Male Flowers on Female Hemp Plants.* HortTechnology hortte, 2018. **28**(6): p. 743-747.
41. Broman, K.W., et al., *R/qtl: QTL mapping in experimental crosses.* bioinformatics, 2003. **19**(7): p. 889-890.
42. Taylor, J. and D. Butler, *R package ASMap: efficient genetic linkage map construction and diagnosis.* arXiv preprint arXiv:1705.06916, 2017.
43. Lincoln, S.E. and E.S. Lander, *Systematic detection of errors in genetic linkage data.* Genomics, 1992. **14**(3): p. 604-610.
44. Lander, E.S., et al., *MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations.* Genomics, 1987. **1**(2): p. 174-181.
45. Voorrips, R., *MapChart: software for the graphical presentation of linkage maps and QTLs.* Journal of heredity, 2002. **93**(1): p. 77-78.
46. Block, A.K., et al., *Biosynthesis and function of terpenoid defense compounds in maize (Zea mays).* Planta, 2019. **249**: p. 21-30.
47. Kopaczyk, J.M., J. Warguła, and T. Jelonek, *The variability of terpenes in conifers under developmental and environmental stimuli.* Environmental and Experimental Botany, 2020. **180**: p. 104197.
48. Kostanda, E. and S. Khatib, *Biotic stress caused by Tetranychus urticae mites elevates the quantity of secondary metabolites, cannabinoids and terpenes, in Cannabis sativa L.* Industrial Crops and Products, 2022. **176**: p. 114331.
49. Nolan, M., et al., *Bigger is better: modern Cannabis trichomes are larger and more productive than their landrace ancestors.* Plant and Cell Physiology, 2025. **66**(10): p. 1477-1492.
50. Laverty, K.U., et al., *A physical and genetic map of Cannabis sativa identifies extensive rearrangement at the 1 A physical and genetic map of Cannabis sativa identifies extensive rearrangement at the 2 T.* Genome Research, 2019. **29**: p. 146-156.
51. Song, Y., et al., *RNATACs: Multispecific small molecules targeting RNA by induced proximity.* Cell chemical biology, 2024. **31**(6): p. 1101-1117.
52. de Meijer, E.d., H. Van der Kamp, and F. Van Eeuwijk, *Characterisation of Cannabis accessions with regard to cannabinoid content in relation to other plant characters.* Euphytica, 1992. **62**: p. 187-200.
53. Zager, J.J., et al., *Gene networks underlying cannabinoid and terpenoid accumulation in cannabis.* Plant physiology, 2019. **180**(4): p. 1877-1897.
54. Wiles, D., et al., *Integrated platform for structural and functional analysis of terpene synthases of Cannabis sativa.* PeerJ, 2025. **13**: p. e19723.
55. Livingston, S.J., et al., *Cannabis glandular trichomes alter morphology and metabolite content during flower maturation.* Plant Journal, 2020. **101**(1): p. 37-56.
56. Voelker, J., R. Mauleon, and M. Shepherd, *A terpene synthase supergene locus determines chemotype in Melaleuca alternifolia (tea tree).* New Phytologist, 2023. **240**(5): p. 1944-1960.
57. Kainer, D., et al., *High marker density GWAS provides novel insights into the genomic architecture of terpene oil yield in Eucalyptus.* New Phytologist, 2019. **223**(3): p. 1489-1504.
58. Külheim, C., et al., *The Eucalyptus terpene synthase gene family.* BMC genomics, 2015. **16**: p. 1-18.

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

765 59. Butler, J.B., et al., *Annotation of the Corymbia terpene synthase gene family shows broad*
766 *conservation but dynamic evolution of physical clusters relative to Eucalyptus.* Heredity, 2018.
767 **121**(1): p. 87-104.

768

769

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Tables and Figure Legends

Table 1. Putative QTLs, locations, effect sizes identified for cannabinoids and terpenes using an all-female F2 mapping population derived from a Type III IPK_CAN_57 and Type I IPK_CAN_36 cross and grown in a controlled indoor environment.

Trait	QTL name	Chr	Marker position	Location (cM)	CI (cM)	LOD	PVE (%)
Cannabinoids:							
CBCA	*qCBCA3*	3	NC_044372.1_91261032	43.0	39.0-49.0	8	18.75
CBCA	*qCBCA4*	4	NC_044373.1_67890639	10.7	7.3-20.0	3.98	8.66
CBCA	*qCBCA9*	9	NC_044376.1_35501335	26.6	16.0-32.0	6.57	15
CBDA	*qCBDA7*	7	NC_044378.1_63326767	21.0	21.0-24.0	26.5	57.13
CBGA	*qCBGA7*	7	NC_044378.1_60852813	21.3	21.0-24.0	13.57	35.84
THCA	*qTHCA7*	7	NC_044378.1_25696363	24.0	24.0-25.7	28.9	60.32
THCVA	*qTHCVA7*	7	NC_044378.1_25696363	24.0	21.0-25.7	14.35	36.8
Terpenes:							
α-Pinene	*qAPN5*	5	NC_044374.1_2380832	16.9	16.0-17.0	33.5	66.0
α-Terpinene	*qATPE5*	5	NC_044374.1_2380832	16.9	14.0-17.0	18.87	49.55
α-Terpinolene	*qATPO5*	5	NC_044374.1_2380832	16.9	15.0-17.0	24.03	58.16
β-Pinene	*qBPN5*	5	NC_044374.1_2380832	16.9	16.9-17.0	10.32	24.18
β-Pinene	*qBPN7*	7	NC_044378.1_37830480	23.6	13.0-30.0	6.91	15.27
β-Caryophyllene	*qCAR3.1*	3	NC_044372.1_1993167	11.9	6-.0-14.0	8.96	10.65
β-Caryophyllene	*qCAR3.2*	3	NC_044372.1_87923072	39.0	36.0-44.0	5.61	6.27
β-Caryophyllene	*qCAR5*	5	NC_044374.1_81259952	31.0	31.0-42.0	6.7	7.66
β-Caryophyllene	*qCAR6*	6	NC_044377.1_77687128	64.0	62.0-66.0	12.69	16.07
β-Caryophyllene	*qCAR8*	8	NC_044379.1_47702906	8.9.0	7.0-13.0	6.2	7.06
β-Caryophyllene	*qCAR9*	9	NC_044376.1_49699970	27.8	26.0-34.0	6.71	7.68
Farnesene	*qFAR8*	8	NC_044379.1_47702906	8.9	0-15.0	3.7	9.5
Farnesene	*qFAR9*	9	NC_044376.1_49699970	27.76	15.9-34.0	6.71	18.0
γ-Terpinene	*qGTP5*	5	NC_044374.1_2380832	16.9	16.0-17.0	19.49	46.62
Humulene	*qHUM3*	3	NC_044372.1_1993167	11.9	5.0—14.0	9.65	19
Humulene	*qHUM5*	5	NC_044374.1_62754453	26.0	22.0--30.0	4.91	8.93
Humulene	*qHUM6*	6	NC_044377.1_49948370	41.0	26.0-49.0	4.94	8.99
Humulene	*qHUM9*	9	NC_044376.1_4058908	17.0	11.0--25.3	4.65	8.47
Limonene	*qLIM5*	5	NC_044374.1_2380832	16.9	16.0-17.0	30.49	62.55
Myrcene	*qMYR3*	3	NC_044372.1_91261032	44.0	33.0-64.0	3.2	4.89
Myrcene	*qMYR4*	4	NC_044373.1_3829198	4.2	1.0-10.33	4.62	8.27
Myrcene	*qMYR5*	5	NC_044374.1_3037253	19.0	18.0-20.0	20.88	43.01
Myrcene	*qMYR7*	7	NC_044378.1_68480444	8.2	0-16.9	4.2	6.5

Chr = Chromosome, CI = Confidence Interval, cM = centi Morgan, LOD = Logarithm Of Odds, PVE = Percent Variation Explained,

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823
Preprint not peer reviewed

OFFICIAL

Figure 1 Principal Component Analysis (PCA) biplots for cannabinoid (a) and terpene (b) contents in a diverse panel of *Cannabis sativa*

The numbers 17-100 correspond to their respective IPK_CAN_# identifiers. The distinction between Drug types (peach squares) and Hemp types (green circles) is based on a THC threshold of 0.3%. Sample scores for the first and second principal components were plotted represented by 50% confidence for cannabinoids and 60% confidence for terpene from all lines. The loading of each group on the biplot was represented by arrows. The length of the arrow approximated the variance of each group, and the angle approximated their correlations. Red circles highlight parental accessions.Cannabidiolic acid (CBDA), Cannabidiol (CBD), Δ9-Tetrahydrocannabinol acid (THCA), Δ9-Tetrahydrocannabinol (THC), Cannabigerolic Acid (CBGA), Cannabigerol (CBG), Tetrahydrocannabivarinic acid (THCVA), Tetrahydrocannabivarin (THCV), Cannabidivarinic acid (CBDVA), Cannabidivarin (CBDV), Cannabinol (CBN), Cannabichromene (CBC). β- caryophyllene (BCAR), myrcene (MYR), α -pinene (APN), β- pinene (BPN), α -terpinolene (ATPO), α -terpinene (ATPE), γ- terpinene (GTP) humulene (HUM), limonene (LIM) and farnesene (FAR).

Figure 2 Neighbourhood Joining Tree (NJ) for a diverse panel of *Cannabis sativa* based on 1504 genome-wide markers.

Accessions were genotyped with the HASCH genotyping platform. High THC accessions (identified in Figure 1 a) are represented in green, and High CBD accessions are represented in red, while sesquiterpene outliers (identified in Figure 1 b) are represented in blue.

Figure 3 Major cannabinoid (a) and terpene (b) contents from female individuals of the parental accessions IPK_CAN_36 (n=2) represented by green and IPK_CAN_57 (n=3) represented by red.

Cannabidiolic acid (CBDA), Δ9-Tetrahydrocannabinol acid (THCA), Cannabigerolic Acid (CBGA), Tetrahydrocannabivarinic acid (THCVA), Cannabidivarinic acid (CBDVA), Cannabichromenic acid (CBCA).

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Figure 4 Comparison of major cannabinoid (a) and terpene (b) contents in the F2 population (n=133) derived from an THC-dominant IPK_CAN_36 (Male) and CBD-dominant IPK_CAN_57 (Female) cross for two contrasting growing environments.

Outdoor shade-house environment represented by orange bars (N=133). Controlled indoor environment is represented by turquoise bars (N=133). The boxes depict the interquartile range, the horizontal line within each box represents the median, and the whiskers show the range of the data, excluding outliers and black dots represent the outliers. Cannabidiolic acid (CBDA), Cannabidiol (CBD), Δ9-Tetrahydrocannabinol acid (THCA), Δ9-Tetrahydrocannabinol (THC), Cannabigerolic Acid (CBGA), Cannabigerol (CBG), Tetrahydrocannabivarinic acid (THCVA), Tetrahydrocannabivarin (THCV), Cannabidivarinic acid (CBDVA), Cannabidivarin (CBDV), Cannabinol (CBN), Cannabichromene (CBC). β- caryophyllene (BCAR), myrcene (MYR), α -pinene (APN), β- pinene (BPN), α -terpinolene (ATPO), α -terpinene (ATPE), γ- terpinene (GTP) humulene (HUM), limonene (LIM) and farnesene (FAR).

Figure 5 Correlations between major cannabinoid and terpene contents in the F2 population (n=133) derived from an THC-dominant IPK_CAN_36 (Male) and CBD-dominant IPK_CAN_57 (Female) cross for the controlled growing environment using Pearson's correlation.

The colour scale on the right indicates the correlation strength, ranging from -1 (red) indicating a strong negative correlation to +1 (blue) indicating a strong positive correlation. Lighter colour represents weaker correlations. Each cell in the heatmap represents the correlation coefficient between the corresponding pair of compounds. Cannabidiolic acid (CBDA), Cannabidiol (CBD), Δ9-Tetrahydrocannabinol acid (THCA), Δ9-Tetrahydrocannabinol (THC), Cannabigerolic Acid (CBGA), Cannabigerol (CBG), Tetrahydrocannabivarinic acid (THCVA), Tetrahydrocannabivarin (THCV), Cannabidivarinic acid (CBDVA), Cannabidivarin (CBDV), Cannabinol (CBN), Cannabichromene (CBC). β- caryophyllene (BCAR), myrcene (MYR), α -pinene (APN), β- pinene (BPN), α -terpinolene (ATPO), α -terpinene (ATPE), γ- terpinene (GTP) humulene (HUM), limonene (LIM) and farnesene (FAR).

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

836

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

Figure 6 Genetic map and quantitative trait locus (QTL) for cannabinoid contents of a segregating F2 population (n=133) derived from an THC-dominant IPK_CAN_36 (Male) and CBD-dominant IPK_CAN_57 (Female) cross based on 313 SNP markers for the controlled growing environment.

Segregation patterns represented as horizontal hash marks on genetic map. QTL for six cannabinoids detected by composite interval mapping ($P < 0.05$; 1000 permutations) scored over 133 F2 plants indicated as vertical bars. Genetic distance (centimorgans) scale bar to left of panel. Cannabidiolic acid (CBDA), Cannabidiol (CBD), Δ9-Tetrahydrocannabinol acid (THCA), Δ9-Tetrahydrocannabinol (THC), Cannabigerolic Acid (CBGA), Cannabigerol (CBG), Tetrahydrocannabivarinic acid (THCVA), Tetrahydrocannabivarin (THCV), Cannabidivarinic acid (CBDVA), Cannabidivarin (CBDV), Cannabinol (CBN), Cannabichromene (CBC)

Figure 7 Genetic map and quantitative trait locus (QTL) for terpene contents of a segregating F2 population (n=133) derived from an THC-dominant IPK_CAN_36 (Male) and CBD-dominant IPK_CAN_57 (Female) cross based on 313 markers for the controlled growing environment.

Segregation patterns represented as horizontal hash marks on genetic map. QTL for ten terpenes detected by composite interval mapping ($P < 0.05$; 1000 permutations) scored over 133 F2 plants indicated as vertical bars. Genetic distance (centimorgans) scale bar to left of panel. β- caryophyllene (BCAR), myrcene (MYR), α -pinene (APN), β- pinene (BPN), α -terpinolene (ATPO), α -terpinene (ATPE), γ- terpinene (GTP) humulene (HUM), limonene (LIM) and farnesene (FAR).

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

OFFICIAL

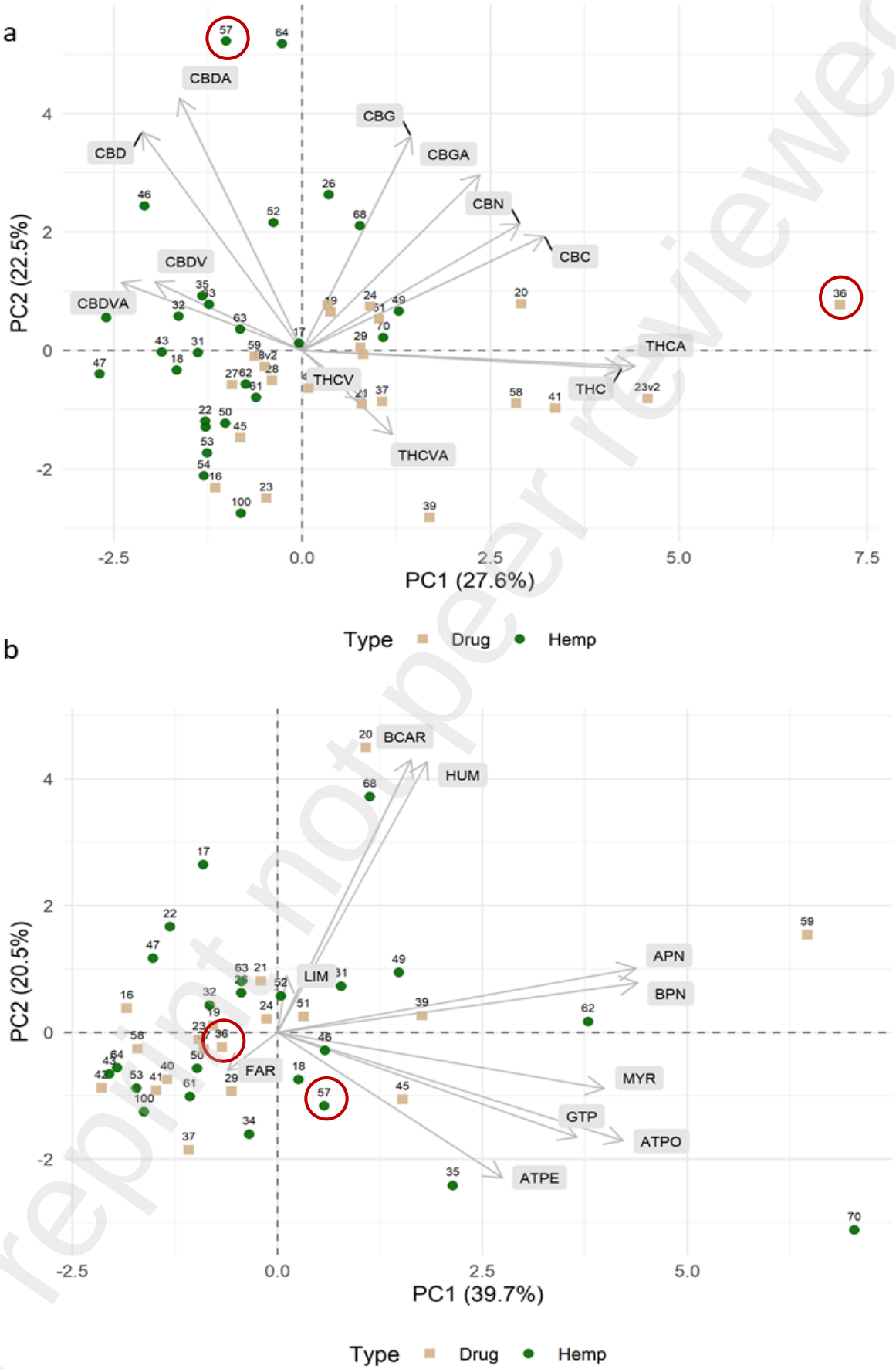
Figure 8 Expressed genes in female individuals of the parental accessions IPK_CAN_36 and IPK_CAN_57 relating to cannabinoid content in proximity to cannabinoid multi QTL cluster on Chromosome 7 (a) and relating to monoterpene contents in proximity to terpene multi QTL cluster on Chromosome 5 (b).

Locus identifiers are based on the CBDRx genome annotation and numbers indicate transcripts per million (TPM). R1-R4 indicate biological replicates.

Preprint not peer reviewed

OFFICIAL
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

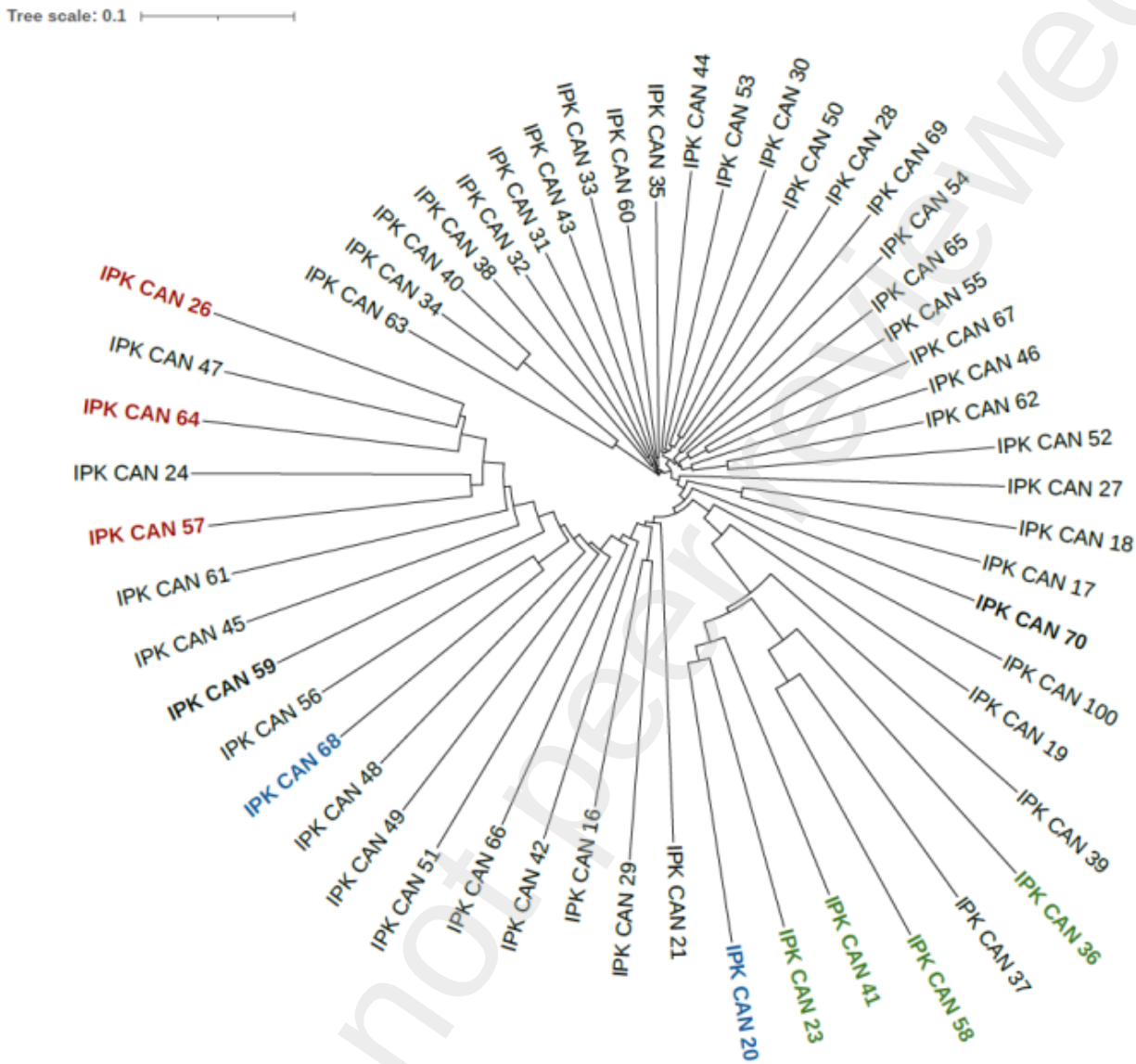
Figure 1



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

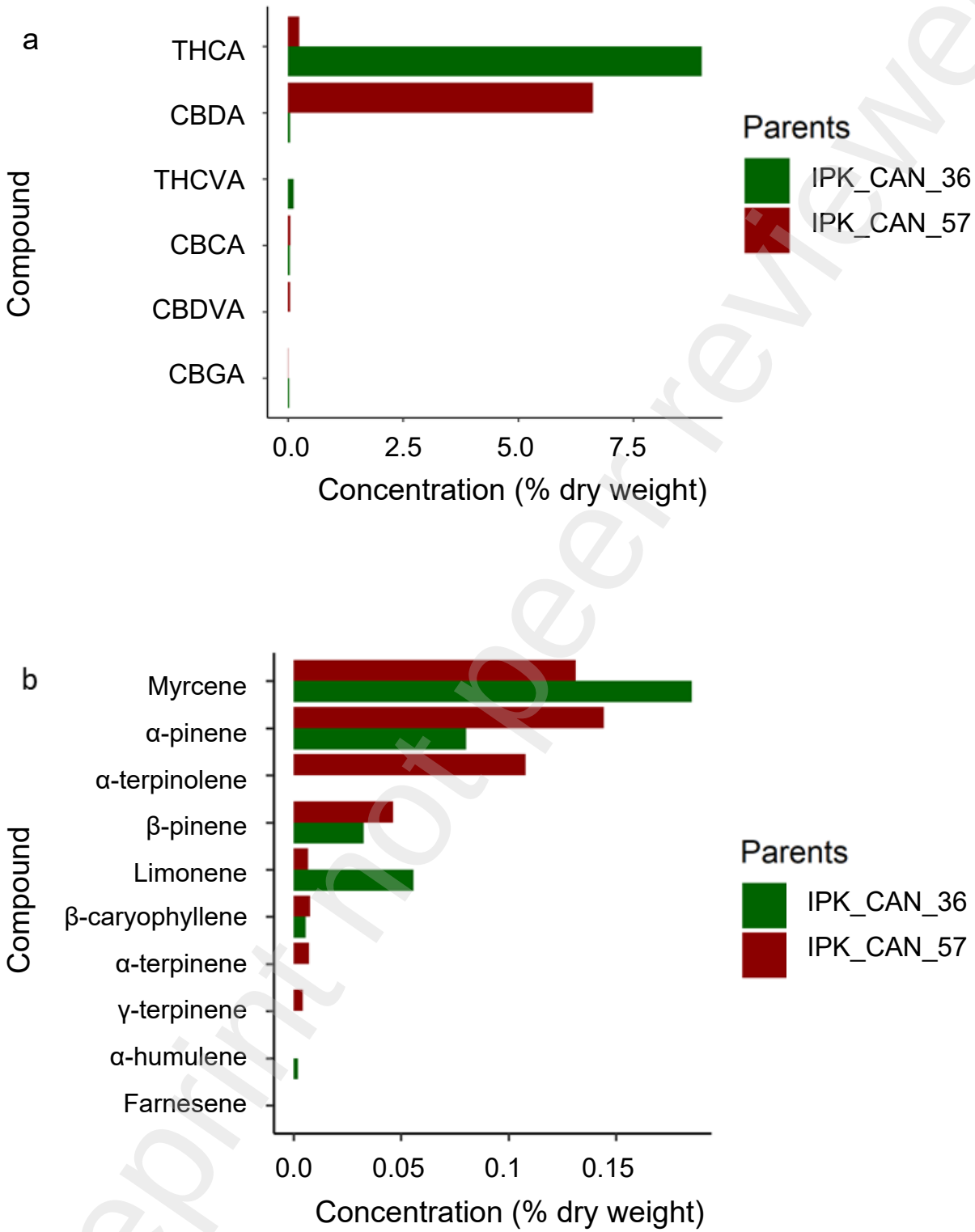
Figure 2



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

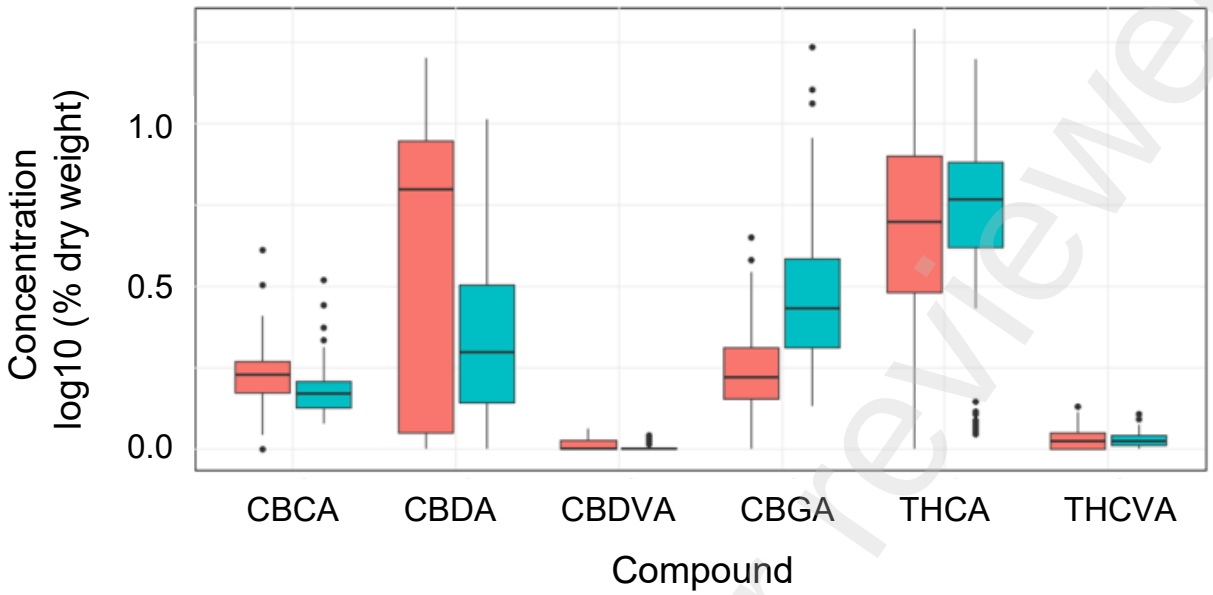
Figure 3

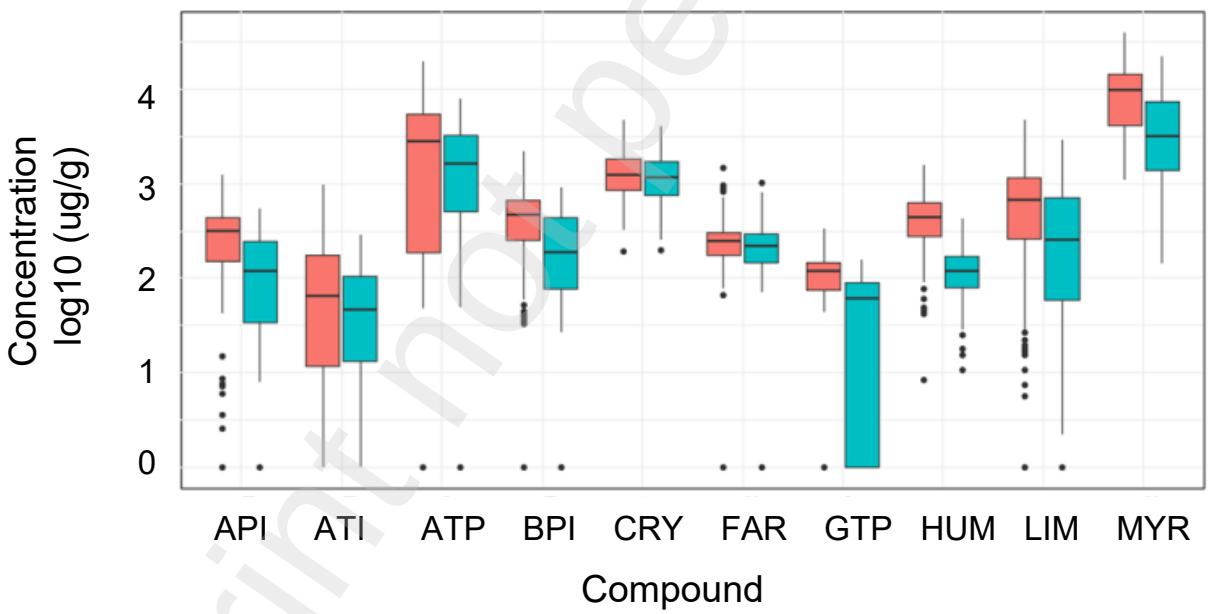


Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

Figure 4

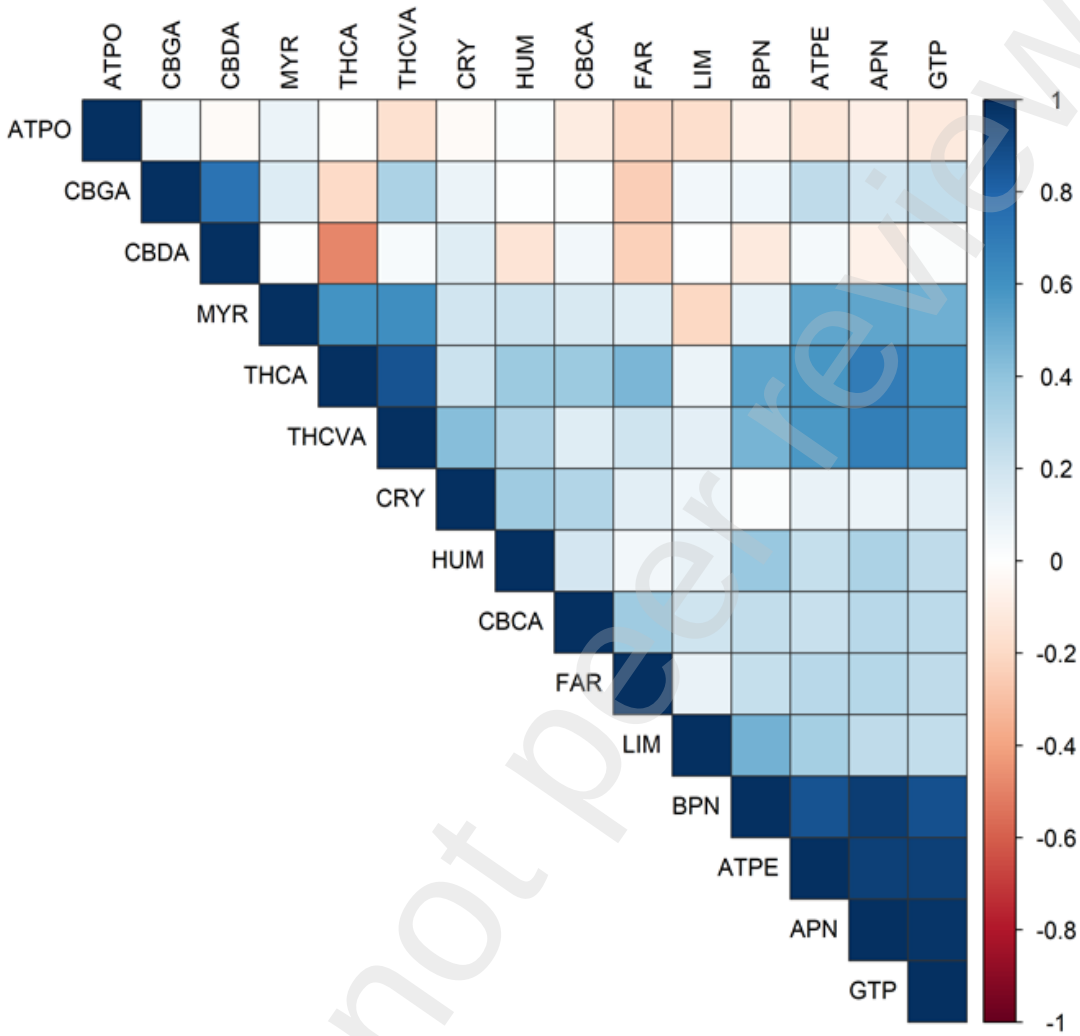




Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

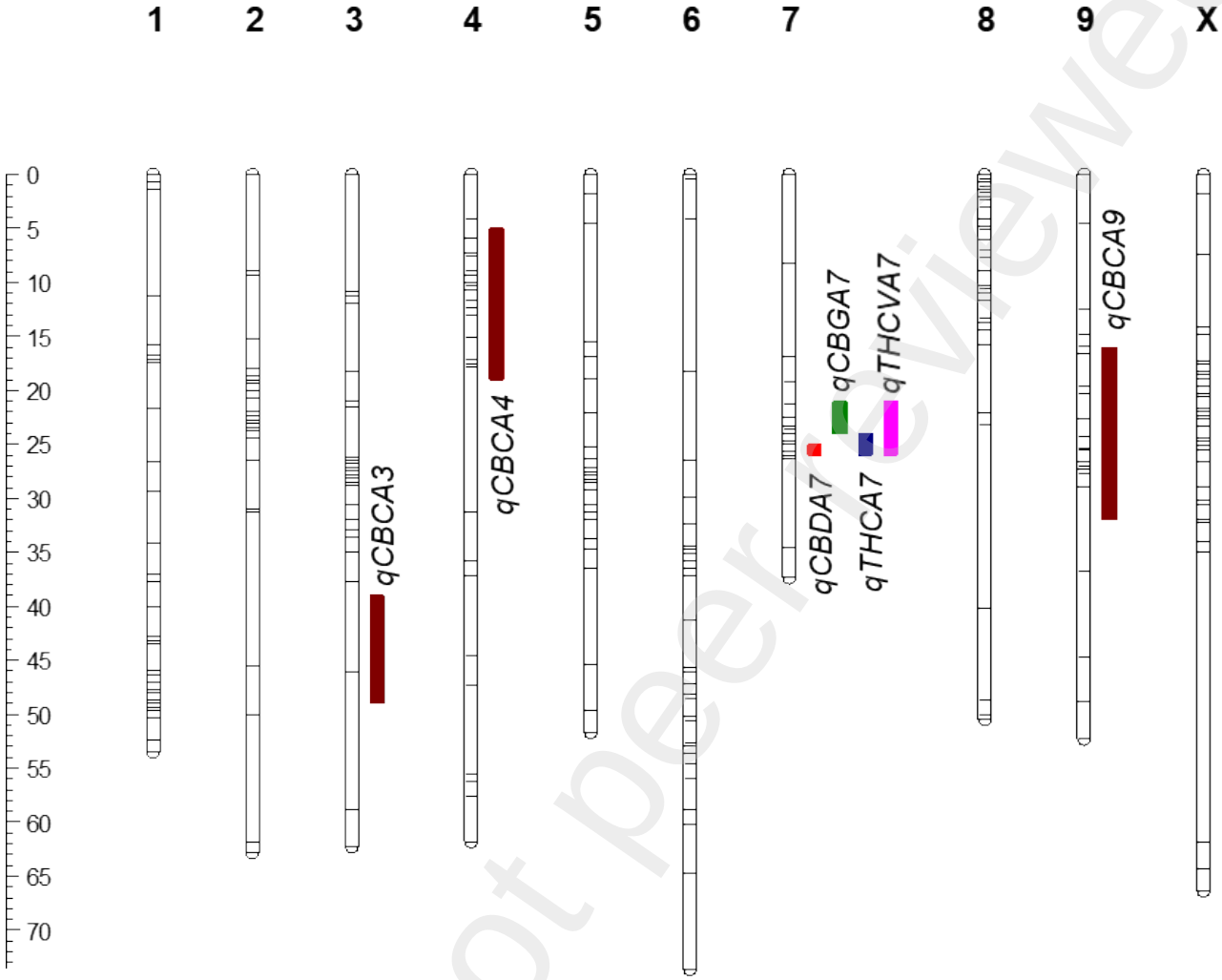
Figure 5



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

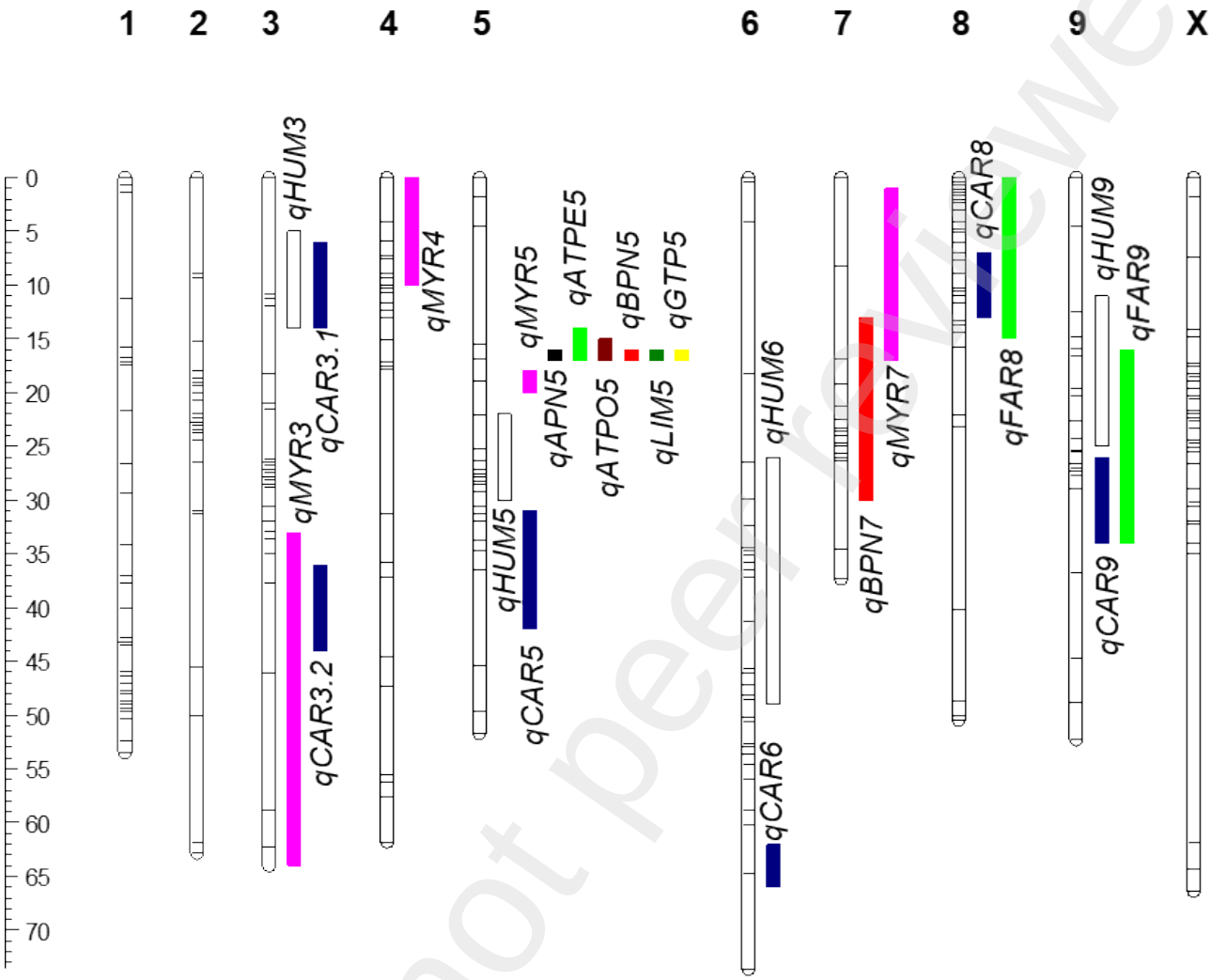
Figure 6



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

Figure 7



Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823

Figure 8

a

Locus Identifier	Annotation	Position		IPK_CAN_57 Trichome				IPK_CAN_36 Trichome			
		Chr	Mb	R1	R2	R3	R4	R1	R2	R3	R4
LOC115697880	THCAS-like	7	25.8	44	43	53	69	720	1065	991	819
LOC115696909	THCAS-like	7	25.9	0	0	0	0	14	14	0	19
LOC115696884	CBDAS-like	7	29.6	13	12	19	19	84	115	108	127
LOC115697762	CBDAS	7	30.9	4590	4602	5162	4838	14	11	20	6
LOC115696987	CBAS-like	7	34.1	1	1	0	3	6	12	1	10
LOC115696986	THCAS-like	7	34.1	0	1	1	2	1	1	0	1
LOC115697126	CBDAS-like	7	34.1	0	0	1	0	0	0	0	0

b

Locus Identifier	Annotation	Position		IPK_CAN_57 Trichome				IPK_CAN_36 Trichome			
		Chr	Mb	R1	R2	R3	R4	R1	R2	R3	R4
LOC115715915	TPS 10-like isoform X1	5	1.3	0	1	1	0	0	0	0	0
LOC115715811	TPS 10	5	1.3	38	42	35	20	1	0	1	1
LOC115716805	TPS 10	5	1.3	6	5	4	3	0	0	0	0
LOC115716806	TPS10-like isoform X4	5	1.3	0	0	0	0	0	0	0	0
LOC115716604	Myrcene synthase	5	2.0	139	123	107	95	7	14	6	13
LOC115716605	Myrcene synthase	5	2.0	89	80	74	62	8	7	6	7
LOC115716066	Limonene synthase	5	2.4	670	578	530	409	13	28	15	17
LOC115716405	Myrcene synthase	5	2.5	432	329	363	278	27	50	28	48

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7163823