UNIVERSITÀ DI FOGGIA

Dipartimento di Scienze Agrarie, Alimenti, Risorse Naturali e Ingegneria

PhD in Management of Innovation in the Agricultural and Food Systems of the Mediterranean Region

XXXV cycle

Genetic and molecular studies on the accumulation of bioactive compounds in *Cannabis sativa* L.

Tutor

Dott. Pasquale De Vita

Dott.ssa Roberta Paris

Coordinator Prof. Giancarlo Colelli

Ph.D. Student

Dott. Flavia Fulvio

Table of content

List of abbreviations

CBC: Cannabichromene

CBCA: Cannabichromenic acid

CBCAS: Cannabichromenic acid synthase

CBD: Cannabidiol

CBDA: Cannabidiolic acid

CBDAS: Cannabidiolic acid synthase

CBG: Cannabigerol

CBGA: Cannabigerolic acid

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleoside triphosphate

HPLC: High-performance liquid chromatography

PCR: Polymerase chain reaction

SNP: Single nucleotide polymorphism

THC: Δ^9-Tetrahydrocannabinol

THCA: Δ^9-Tetrahydrocannabinolic acid

THCAS: Tetrahydrocannabinolic acid synthase

Extended abstract

Cannabis is a predominantly dioecious and phenotypically diverse, monotypic genus consisting of a single species, the *Cannabis sativa*. Its cultivation started in Eurasia, several thousand years ago, where the selective pressure that has led it to be cultivated around the globe today began. As a dioecious plant, *C. sativa* is an obligate outbreeder, and as such every individual plant is genetically unique, and producing offspring with desired traits can be challenging.

This genetic variation can be both a blessing and a curse for breeders, because it provides a rich genetic pool from which to select for desired traits, but also it can be difficult to achieve consistency and stability in breeding programs.

The current *C. sativa* constraints, such as its high level of heterozygosity, reside in the limited genetic improvement and poor conservation of genetic resources, resulting from prohibition against the plant as a source of one of the most widespread illicit drugs. Unlike major crops, *C. sativa* has suffered from a lack of progress in the knowledge of its physiology and has not benefit from advances in breeding technologies.

In recent years, there has been growing interest in the potential of *C. sativa* for both medicinal and industrial purposes. Its legalization in many countries has led to increased research and investment in the cultivation, processing, and marketing of hemp and medical *Cannabis* products. As a result, the plant is gaining recognition as a valuable crop with significant economic, environmental, and health benefits.

The plasticity and wide genetic variability represent the intrinsic functionality of this plant and emphasise the agricultural value of the species. The long stem fibres have been used for millennia as the predominant source of fibre in the textile and paper industry. The seeds are suitable for human and animal consumption and contain a balanced ratio of omega-6/omega-3 essential polyunsaturated fatty acids.

In addition to traditional uses, one of the main applications today is Cannabis as a source of countless bioactive compounds.

Indeed, *C. sativa* female flowers are rich in bioactive compounds belonging to the secondary metabolism with remarkable phytochemical potential. These compounds are mostly classified into three classes: phytocannabinoids, terpenoids and phenolics.

Phytocannabinoids are the most studied compounds, mainly due to their wide range of pharmaceutical effects in humans, with more than 150 constituents identified, of which THCA, CBDA, CBGA and CBCA are the most representative.

Over the last five years, significant advances in *C. sativa* genetic research have led to a deeper understanding of the plant's genetic background and its potential applications. With the increasing legalization and commercialization, genetic research has become a critical tool for developing new strains with desired traits and improving the quality and consistency of *C. sativa* products.

One major breakthrough in *C. sativa* genetic research has been the sequencing of the genome, providing valuable insights into the plant's complex genetics and the biochemical pathways that produce its various phytocannabinoids and terpenes.

Other recent studies have focused on identifying the genetic markers associated with specific traits, such as high THCA or CBDA content, flowering time, disease resistance, and yield.

Advances in new breeding techniques, such as CRISPR-Cas9, have also opened up new possibilities for editing genome for a desired trait.

In addition, genetic research has also shed light on the evolutionary history of the plant and its relationship with other members of the *Cannabaceae* family, helping to clarify the taxonomic classification of different genotypes.

Overall, the recent years have seen significant progress in *C. sativa* genetic research, providing new insights into the plant's biology and potential applications for human health and industry.

There is also a growing interest in the study of the plant's genetics and molecular biology. By identifying the genes and pathways involved in phytocannabinoids biosynthesis, as well as the regulatory mechanisms that control their expression, new strategies can be developed for manipulating their production, with important applications in medicine, industry, and agriculture.

In this Doctoral Thesis, *C. sativa* genetic was studied with the aim of better understanding the route leading to phytocannabinoids synthesis. Both industrial and medicinal varieties were used, tapping into the biodiversity available within a germplasm collection maintained at CREA-Research Centre for Cereal and Industrial Crops

in Bologna and Rovigo sites, where it was also possible to cultivate high-THC strains in outdoor and indoor facilities authorized with decree n. SP/041 on 13rd March 2017 according to art. 26 of the D.P.R. 309/90).

In the first chapter an investigation of the genetic and transcriptional variability of cannabinoid synthases is illustrated, performed on a set of Italian and French *C. sativa* genotypes with diverse chemotypes. The work focused also on the role of the *cannabichromenic acid synthase* (*CBCAS*), for a long time neglected, and considered a gene peculiar to fibre genotypes, considered a mutated, non-functional copy of *THCAS* for their extremely high sequence similarity.

The results showed that *CBCAS* is present and expressed in both drug-type and fiber-type genotypes, and several genetic variants of *CBCAS* genes were identified, which may contribute to differences in phytocannabinoid production between different chemotypes.

In the second chapter the transcription analyses of cannabinoid pathway genes, during early vegetative stages of plant development are reported. Besides cannabinoid synthases, also other genes responsible for early reaction in the synthesis of these terpeno-phenolic compounds were analysed in order to deepen knowledge for defining the main determinants of this metabolism in the early stages of plant life. Cannabicromenic acid was found the first cannabinoid accumulated in the seedlings, shortly after emergence, and hypotheses are given on the regulation of its synthesis *in planta*. Moreover, in this chapter is also reported the transcription of cannabinoid synthase genes during seed germination and in long-storage seeds, uncovering a new and important level of gene regulation during seed germination and providing an estimate of the importance of this metabolism for the plant.

The third chapter focused on the analysis of a cultivated population of the FINOLA variety, obtained from the certified commercial seed. As with many industrial varieties, FINOLA has a chemotype III and derives from years of breeding, however it still accumulates little amount of THC (residual THC) on inflorescence which can exceed the legal limit and cause seizure and losses to growers. The results show how the B1080/B1192 molecular marker was effective at identifying hemp plants with functional THCA synthase and total THC content above the legal limit. Biochemical analyses also demonstrated a 100% association between the chemotype predicted by molecular markers and the actual chemotype. This study suggests that molecular

markers could be used as effective tools for hemp growers and breeders, helping to ensure compliance with regulations and avoid legal issues related to hemp cultivation.

The possibility of combining the micropropagation with the maintenance and renewal of mother plants, deputed to the national production of medical cannabis was evaluated and reported in chapter four. Overall, the work provided a valuable protocol for the micropropagation of *C. sativa* involving the use of cytokinin and gibberellin-based media for the induction and proliferation of shoots, followed by rooting on a separate rooting medium. The results also demonstrated the successful acclimatization and transfer of *in vitro*-derived plants to *ex-vitro* conditions and the potential for using micropropagation as a tool for the genetic improvement and development of new *C. sativa* plants but also for the preservation of genetic diversity.

Chapter 1

Analysis of sequence variability and transcriptional profile of cannabinoid synthase genes in *Cannabis sativa* L. chemotypes with a focus on cannabichromenic acid synthase

This chapter has been extracted from "Analysis of Sequence Variability and Transcriptional Profile of Cannabinoid synthase Genes in Cannabis sativa L. Chemotypes with a Focus on Cannabichromenic acid synthase" from Flavia Fulvio [1,2], Roberta Paris [1], Massimo Montanari [1], Cinzia Citti [3,4], Vincenzo Cilento [1], Laura Bassolino [1], Anna Moschella [1], Ilaria Alberti [5], Nicola Pecchioni [6], Giuseppe Cannazza [3,4] and Giuseppe Mandolino [1]

[1] CREA—Research Centre for Cereal and Industrial Crops, Via di Corticella 133, 40128 Bologna, Italy;

[2] Department of Sciences of Agriculture, Food Natural Resources and Engineering, University of Foggia, Via Napoli 25, 71122 Foggia, Italy

[3] CNR NANOTEC—Institute of Nanotechnology, Via Monteroni, 73100 Lecce, Italy;

[4] Department of Life Science, University of Modena and Reggio Emilia, Via G. Campi 103, 41125 Modena, Italy

[5] CREA—Research Centre for Cereal and Industrial Crops, Via G. Amendola 82, 45100 Rovigo, Italy;

[6] CREA—Research Centre for Cereal and Industrial Crops, S.S. 673 Km 25,200, 71122 Foggia, Italy;

Published in Plants, doi: 10.3390/plants10091857.

Abstract: *Cannabis sativa* L. has been long cultivated for its narcotic potential due to the accumulation of tetrahydrocannabinolic acid (THCA) in female inflorescences, but nowadays its production for fiber, seeds, edible oil and bioactive compounds has spread throughout the world. However, some hemp varieties still accumulate traces of residual THCA close to the 0.20% limit set by European Union, despite the functional gene encoding for THCA synthase (THCAS) is lacking. Even if some hypotheses have been produced, studies are often in disagreement especially on the role of the cannabichromenic acid synthase (CBCAS). In this work a set of European *Cannabis* genotypes, representative of all chemotypes, were investigated from a chemical and molecular point of view. Highly specific primer pairs were developed to allow an accurate distinction of different *cannabinoid synthases* genes. In addition to their use as markers to detect the presence of *CBCAS* at genomic level, they allowed the analysis of transcriptional profiles in *hemp* or *marijuana* plants. While the high level of transcription of *THCAS* and *cannabidiolic acid synthase* (*CBDAS*) clearly reflects the chemical phenotype of the plants, the low but stable transcriptional level of *CBCAS* in all genotypes suggests that these genes are active and might contribute to the final amount of cannabinoids.

1. Introduction

Cannabis sativa L. is an annual, dioecious plant, characterized by the production of cannabinoids, terpenophenolic metabolites of great pharmaceutical interest, mainly synthetized and secreted in glandular trichomes of pistillate inflorescences.

Cannabinoids are produced by condensation of a phenolic moiety (usually olivetolic acid) with a terpenic one (geranylgeraniol-diphosphate). This reaction synthesizes cannabigerolic acid (CBGA), from which the other main cannabinoids are formed via enzymatic conversions.

Enzymatic cannabinoid biosynthesis is catalyzed by a number of oxidocyclases, among which the most prominent are the tetrahydrocannabinolic acid-, cannabidiolic acid- and cannabichromenic acid-synthase, leading to the accumulation of tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA) and cannabichromenic acid (CBCA) respectively (Gülck and Møller, 2020). These enzymes are thought to be poorly or completely non-functional in CBGA-accumulating plants (Mandolino and Carboni, 2004).

Based on both the amount (expressed as% weight/weight) of main cannabinoids and their ratio, the chemical phenotype (chemotype) of *Cannabis* plants can be classified from I to V (Mandolino and Carboni, 2004) .

THCA synthase (*THCAS*), *CBDA synthase* (*CBDAS*) and *CBCA synthase* (*CBCAS*) are members of the *Berberine Bridge Enzyme* (*BBE*)-*like* gene family, containing an N-terminal signal peptide and a Flavin Adenin Dinucleotide (FAD) binding domain (Sirikantaramas et al., 2004; Taura et al., 1996).

The genetics of cannabinoid synthesis has been studied for several years and different genes encoding these enzymes are known, each made of a single exon, with THCAS and CBCAS sharing 92% identity at amino acid level and 84% and 83% identity compared to CBDAS, respectively (Laverty et al., 2019; Shoyama et al., 2012; Sirikantaramas et al., 2004; Taura et al., 1996; van Velzen and Schranz, 2020). While *THCAS* and *CBDAS* have been widely studied at genetic and molecular level, little information is available on *CBCAS* genes.

Originally, Kojoma et al. (2006) obtained from hemp varieties gene sequences highly related to functional *THCAS*, putatively encoding for complete polypeptides, but differing for several single nucleotide polymorphisms (SNPs). They named these sequences fiber-type *THCAS* genes, under the assumption that they were not functional and unable to synthesize cannabinoids (Kojoma et al., 2006). Since then, a number of so-

called fiber-type *THCAS* sequences have been found in both hemp and high THCA varieties (Cascini et al., 2019; Weiblen et al., 2015).

In 2019, Laverty et al. (2019) demonstrated that a fiber-type *THCAS* gene coded for a 71 kDa CBCAS, capable of transforming the CBGA precursor into CBCA and that the accumulation of CBCA correlated with the transcriptional level of *CBCAS* in various *Cannabis* tissues, with the highest level observed in female floral tissue.

More recently, a comprehensive clade-based classification of all cannabinoid oxidocyclases proposed the naming of all fiber-type *THCAS* as *CBCAS* (van Velzen and Schranz, 2020) since they belonged to the same clade of the only functionally CBCAS characterized by Laverty et al. (2019). Based on these results today they are conventionally collectively referred to as *CBCAS* and, according to this definition, in the present paper we will refer to all fiber-type *THCAS* sequences with a complete open reading frame as *CBCAS*.

Despite the name, further verification of their functionality *in planta* is needed and many authors claim for a meta-analysis to be conducted to verify that hypothesis (Cascini et al., 2019; Hurgobin et al., 2020; Kojoma et al., 2006; van Velzen and Schranz, 2020; Weiblen et al., 2015) . Indeed, even though the CBCAS enzyme was biochemically characterized over 20 years ago (Morimoto et al., 1998) data on CBCAS are still scattered and somehow in disagreement. For example, one of the main inconsistencies with current evidence lies in previous analyses, where CBCA accumulation was demonstrated to be prominent during the juvenile stages of plant development, declining with maturity and irrespective of the chemotype at flowering (de Meijer et al., 2009). Both the physiology of this enzyme and its relationship with the more studied CBDAS and THCAS are poorly understood and need clarification, especially in view of the possible contribution to the final chemotype.

Understanding whether *CBCAS* sequences are present in each *Cannabis* variety or breeding material and the quantification of their transcription and potential translation *in planta*, would shed light on this important topic.

Several chemotype-associated PCR-based markers have been developed (Borroto Fernandez et al., 2020; Cascini et al., 2019; Pacifico et al., 2006; Staginnus et al., 2014)to detect the presence of functional *THCAS* sequences, allowing to unambiguously distinguish between drug-type and fiber type materials. None of these markers amplify the *CBCAS* sequences, despite their high similarity at nucleotide level and among them, the three-primer marker system *B1080/B1192* developed by Pacifico et al. (2006), is fully associated with the

chemotype, being able to discriminate chemotype I (THCA prevalent), II (both THCA and CBDA present in roughly similar amounts) and III (CBDA prevalent).

A number of complete and partially annotated genomic sequences have been published in the latest years (Gao et al., 2020; Grassa et al., 2021; Laverty et al., 2019; McKernan et al., 2020; van Bakel et al., 2011)and information about the transcribed genes in different stages and tissues (Braich et al., 2019), as well as on putative regulators of secondary metabolites (Bassolino et al., 2020; Liu et al., 2021) is already known.

Despite the behavior as codominant alleles at a single locus observed in segregation analyses (Mandolino et al., 2003), *THCAS* and *CBDAS* genes have been recently mapped to separate loci in tight linkage on the same chromosome (Grassa et al., 2021; Laverty et al., 2019). Moreover, in these loci there are multiple copies of *THCAS*- and *CBDAS*-related sequences. A few functional genes are expressed and therefore contribute to the final chemotype, while others are pseudogenes or partially functional sequences (Braich et al., 2019; Onofri et al., 2015; Vergara et al., 2019).

Interestingly, many hemp inflorescences accumulate traces of residual THCA at concentrations close and sometimes above the limit of 0.20% of dry weight set by E.U., despite the lack of a functional *THCAS* gene in their genomes and years of breeding aimed at eliminating the accumulation of this cannabinoid. Some hypotheses have been proposed: the first one is that the CBDAS could produce with very low efficiency THCA from CBGA, due to its high similarity with THCAS (Zirpel et al., 2018); alternatively, sequences corresponding to putative *CBCAS* could be responsible for this residual production of THCA, small but apparently difficult to eliminate, since the substitutions at the protein level with functional THCAS are minimal and do not involve protein active sites or residues required for THCA synthesis (Shoyama et al., 2012; Sirikantaramas et al., 2004). Gaining a deeper knowledge of *CBCAS* sequences and of their transcription levels could help to better define the process behind the synthesis of THCA in hemp genotypes as an unexpected by-product and consequently guide future breeding strategies.

In the present work, several fiber- and drug-type genotypes were investigated for the presence of *CBCAS* sequences, assessing sequence variability and transcriptional levels. Markers were also developed to distinguish these sequences from the *THCAS* and *CBDAS* ones.

2. Materials and Methods

2.1. Plant Material and Samples Collection

Nine Cannabis sativa hemp varieties, two hemp selections and two medical varieties were used in this work. Among hemp, six were dioecious (Carmagnola, CS, Fibrante, Fibranova, Eletta Campana, Bernabeo) and five monoecious (Ermo, Santhica 27, Carmaleonte, Codimono and Futura 75). Medical varieties CINBOL and CINRO (CPVO registration numbers 50407 and 50406, respectively) were developed at CREA, cultivated in authorized indoor facilities and clonally propagated by cuttings.

For DNA isolation, seeds of the eleven hemp genotypes were sown in a peat-type TRAYSUBSTRATE (Klasmann-Deilmann Gmbh) and seedlings were grown in a plant growth chamber (Percival®AR-36LC8, Percival scientific, Inc. Perry, IA, USA), with a light/dark cycle of 18/6 h, a temperature of 24 °C and 60% relative humidity. About 14 days after sowing, young leaves were collected from 10 different seedlings of the same genotype, pooled and conserved at −80 °C until DNA extraction.

For RNA isolation, the seeds of hemp genotypes were sown on the 19 April 2019 and cultivated in open field in CREA-CI experimental station of Rovigo (GPS coordinates: 45.078722/11.766035). Three biological replicates were collected from three different plants for each genotype, each consisting of three female or monoecious 10 cm apical flowers (comprising apical flower and floral leaves) sampled 123 days after sowing from the monoecious varieties and 151 days after sowing from the dioecious ones.

Samples (Figure 1) were immediately frozen in liquid nitrogen and stored at −80 °C.

Medical varieties with high THCA CINRO and CINBOL were also grown outdoor starting from cuttings from mother plants. For DNA extraction expanded leaves were collected from young, rooted cuttings; for RNA extraction, inflorescences were collected 50 days after the start of flowering, in the attempt to prevent an accumulation of THCA be-yond the limit authorized for open field cultivation in Rovigo station. Samples were conserved as described above.

For chemical analysis, the same inflorescences were used after lyophilization.

Figure 1 (a) Cannabis samples collected for DNA extraction, molecular cloning and sequencing. (b) CINRO female in-florescence 50 days after onset flowering. (c) Chemotype III dioecious genotype female inflorescence 151 days from sowing.

2.2. LC-UV Analysis

For quantitative analysis of standard cannabinoids, namely CBDA, CBGA, THCA, CBD, CBG, Δ^9-THC, Δ8-THC and CBC, samples (500 mg inflorescence of each variety, finely powdered) were prepared and analyzed by liquid chromatography coupled with UV detection (HPLC-UV) according to the protocol of the German Pharmacopoeia as previously reported (Citti et al., 2016; Linciano et al., 2020a). Briefly, the extraction in analytical grade ethanol 96% (Carlo Erba, Milan, Italy) was carried out in three cycles with progressively decreasing volumes of solvent (20, 12.5 and 12.5 mL) under magnetic stirring for 15 min each. After collecting the liquid fraction in a volumetric flask, the extract was brought to 50 mL final volume with fresh ethanol. A 1 mL aliquot was filtered through a 0.45 µm syringe filter, diluted 1:10 with acetonitrile and injected into the chromatographic apparatus (5 µL). The chromatographic separation was performed on a Vanquish Core UHPLC system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a vacuum degasser, a binary pump, a thermostated autosampler at 4 °C, a thermostated column compartment set at 30 °C and a diode array detector. The analyses were carried out following the method reported in previous works (Baratta et al., 2019; Citti et al., 2019; Linciano et al., 2020b) with slight modifications and acquired with the Chromeleon 7.3 Data System software (Thermo Fisher Scientific) following the UV signal registered at 228 nm. The column employed was a Poroshell 120 C18 column (Poroshell 120 EC-C18, 3.0 × 150 mm, 2.7 µm) (Agilent, Milan,

Italy) and the mobile phase consisted of water (solvent A) and acetonitrile (solvent B) both with 0.1% formic acid. A linear gradient from 5 to 95% B was set over 20 min, then the column was washed with 95% B for 5 min and reconsti-tuted with the initial mobile phase (5% B) for other 5 min bringing the total run time to 30 min.

Calibration curves were individually prepared in acetonitrile for each cannabinoid standard from 1 mg/mL stock solutions (Cerilliant, Sigma Aldrich, Milan, Italy) and linearity was assessed in the range 0.05–5.00 µg/mL for all cannabinoids and also in the range 5.0–50 µg/mL for CBD in the case of CBD-rich, chemotype III, genotypes ($R2 > 0.999$). The total cannabinoid content was calculated according to Baratta et al. (Baratta et al., 2019). Three injections were performed for each sample and results are given as % (w/w), expressed as mean of replicates ($n = 3$). The limit of detection (LOD) was 0.001% and limit of quantification (LOQ) was 0.005% (Table 1).

2.3. DNA Isolation and Genotyping PCR Analysis

For DNA isolation, 100 mg of frozen pooled leaves were finely ground by physical treatment with steel beads in a Tissue Lyser II (Qiagen) at 30 Hz for five minutes and ex-tracted using the Invisorb® Spin Plant Mini Kit (STRATEC molecular Gmbh, Berlin, Ger-many) according to the manufacturer's instructions. DNA was eluted in 100 µL sterile water, quantified at the Infinite 200 PRO spectrophotometer (TECAN) and diluted to 10 ng/µL.

The three-primers multiplex system used to genotype the C. sativa plants at the B locus (Pacifico et al., 2006) is detailed in Appendix A and generated either a 1081 bp (presence of a functional CBDAS), a 1192 bp (presence of a functional THCAS) or both fragments (presence of both functional genes).

2.4. RNA Isolation and cDNA Synthesis

For RNA isolation, 100 mg frozen samples were finely ground and extracted using the Spectrum Plant Total RNA Kit (Sigma Aldrich, Merck Life Science S.r.l., Milan, Italy). Total RNA was eluted in 50 µL of DEPC-treated water and spectrophotometrically quantified. Five hundred ng of total RNA were treated with 1 unit of DNAse I Amplification Grade (Sigma-Aldrich, Merk Life Science S.r.l., Milan, Italy) and retrotranscribed with the High-Capacity RNA to cDNA kit (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's instructions.

2.5. PCR Amplification, Cloning and Sequencing

THCAS- and CBDAS-specific primers from Onofri et al. (2015), with some modifications as specified in Table 2, were used in this work to isolate the complete coding sequences of the genes.

For a specific amplification of the CBCAS genes, a search on Cannabis assemblies (available at: https://www.ncbi.nlm.nih.gov/assembly/?term=cannabis, accessed on 19 December 2019) was performed in order to identify conserved regions allowing the dis-crimination between active and "inactive" forms. Once identified, primers were designed using the software Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2007).

The amplification reactions were performed on 20 ng of template DNA using the Invitrogen™ Platinum™ SuperFi™ polymerase (Thermo Fisher Scientific, Waltham, MA, USA), which has a reported fidelity >300 times higher than Taq, in 50 µL reaction. PCR products, purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, D Düren, Germany), were cloned in pJET1.2 plasmid vectors (CloneJET PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA) and transferred in Escherichia coli DH5α cells in accordance with the manufacturer's protocol. Properly transformed cells were selected for ampicillin resistance (100 µg/mL final concentration) on LB -agar medium and colonies were screened by PCR, using the protocol and primers suggested in the cloning kit.

For each genotype, up to eight plasmids for each gene were isolated using the PureLink® Quick Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA, USA). The inserts were then sequenced according to the Sanger method (BMR Genomics Srl, Pa-dova, Italy), using the pJET 1.2 sequencing primers (pJET 1.2 forward: CGACTCAC-TATAGGGAGAGCGGC; pJET 1.2 reverse: AAGAACATCGATTTTCCATGGCAG).

Only complete coding sequences were further considered in this work. The variability of THCA and CBDA synthases was evaluated firstly by aligning the sequences from each individual genotype in order to find unique single sequences. These were aligned with the two reference sequences of THCA- and CBDA-synthase (E33090 and E55107) and, finally, compared across the different genotypes. The sequence comparison at nucleotide and amino acid level was carried out using MUSCLE 3.8 (available at: https://www.ebi.ac.uk/Tools/msa/muscle/ accessed on 8 March 2021). For phylogenetic analysis the evolutionary history was inferred using the Neighbor-Joining method in the MEGA-X software (Kumar et al.,

2018; Tamura et al., 2004). The final figure of NJ DIRs tree was obtained by using the iTOL tree editor (The iTOL Platform).

2.6. Transcriptional Analysis of Cannabinoid synthases

The transcriptional levels of the CBDA-, THCA and CBCAS and of three candidate reference genes CsActin, CsRAN (Ras-related Nuclear protein) and CsClathrin (Mangeot-Peter et al., 2016)were measured by reverse transcription quantitative real time PCR (RT-qPCR) using a Ro-tor-Gene 6000 (Corbett) and SYBR Green chemistry.

Each reaction contained 3 μL of a 1:9 dilution of cDNA, 5 μL of Power Up®SYBR master mix (Thermo fisher Scientific), highly specific primers and RNA-free water to a fi-nal volume of 10 μL. Primer pairs, optimized conditions and annealing temperatures are listed in Table 3. Primer design and the optimization steps performed to assure the high specificity of amplification were performed following Pagliarani et al. (Pagliarani et al., 2013)and a detailed description is reported in Appendix B. Further details of RT–qPCR conditions are reported in Appendices A and B following the Minimum Information for publication of Quanti-tative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2013; Combest et al., 2021).

Different amplification conditions were used depending on the annealing temperature of the primers used: when this was below 60 °C, a two-step method was used, consisting of 15 s at variable temperature for annealing and 1 min at 72 °C for extension; when this was above 60 °C, a one-step method was used, consisting of 60 s at the desired temperature in which both annealing and extension are performed. Finally, a heat dissociation protocol (from 60 °C to 99° C) was performed and a dissociation curve for each sample was generated. Three biological replicates were analyzed for each genotype, which in turn were tested in three technical replicates.

A standard curve was added in all assays, both for target and for reference genes. Standard curves were made of 5 points, prepared as four 1:4 serial dilutions of a 1:3 dilution of the cDNA (500 ng).

To verify the specificity of reaction, each assay included also several negative controls as specified in Appendix B.

The amplification efficiency (E) of each primer pair was estimated using the slope of the regression line, according to the equation: E = 10^(−1/slope)—1. The transcriptional level stability across samples of CsActin,

CsRAN and CsClathrin was verified by software RefFinder (Xie et al., 2012) and, according to results, CsClathrin and CsRAN were selected as reference genes.

Raw data for target and reference genes were transformed using the 'Standard Curve Method'; the transcripts level of target genes was normalized to the geometric mean of the transcripts level of CsClathrin and CsRAN and reported as Relative Quantitation (RQ) of transcriptional levels (Larionov et al., 2005) expressed in Arbitrary Units (A.U.). Finally, the standard error of the mean of three biological replicates was calculated and reported in the graphs as error bar. Comparison between genotypes vs CINRO for each gene transcript levels was done using Student's t-test with log-transformed expression data.

3. Results

3.1. Genotyping

Eleven hemp and two medical Cannabis genotypes (these last two currently used by pharmaceutical industry in Italy) were analyzed using the multiplex primer marker system B1080/B1192 (Pacifico et al., 2006).

A single amplification band at 1192 bp was obtained for CINBOL (chemotype I, the only THCA-predominant variety), while for CINRO (chemotype II) both bands of expected size were amplified (Appendix A). For all hemp genotypes, the multiplex primer system amplified the 1080 band, with no distinction between chemotype III (Fibrante, Fibranova, Carmagnola, CS, Eletta Campana, Codimono, Carmaleonte and Futura 75), chemotype IV (Santhica 27 and Bernabeo) and Ermo (chemotype V), as already previously reported (Pacifico et al., 2006).

3.2. Quantification of Major Cannabinoids

The total amount of the main cannabinoids CBD, CBG and THC was measured by HPLC in female or monoecious hemp inflorescences harvested as reported in Table 1, about 4–5 weeks after full bloom, which would correspond, at least for day-length-sensitive genotypes, to the highest concentration of CBD or THC. Inflorescences of CINBOL and CINRO were collected earlier, to prevent an excessive accumulation of THC beyond the authorized limit for open field cultivation; therefore, total cannabinoid content was expected to be lower than at maturity (up to 18% in CINBOL and 15% in CINRO).

The varieties with the highest cannabinoid content were CS and Carmagnola, with 5.49% and 5.32% CBD/d.w, respectively (Table 1). The other chemotype III genotypes (usually characterized by relatively high amounts of CBD) showed a CBD content ranging from 4.02% in Codimono to 1.92% in Carmaleonte. In Santhica 27, one of the lowest amounts of CBD (0.26%) was detected, followed by Ermo, which showed the lowest CBD content of all studied genotypes, as expected from its zero cannabinoids chemotype V. A high CBD content was found in CINRO, while CINBOL showed relatively very low amount of CBD in its female inflorescences.

Table 1. Quantitative results of total cannabinoid analysis expressed as the mean of 3 replicates. LOD (limit of detection): 0.001%. LOQ (limit of quantification): 0.005%. Information regarding inflorescences sampling period is expressed as days after sowing for hemp varieties, as days from onset flowering for CINBOL and CINRO. Results of genotyping obtained using the marker B1080/B1192 are given for each genotype. BD and BT refer to the allelic status as determined by the marker (Onofri et al., 2015). Data are expressed as percent of total inflorescence dry weight.

Genotype	Sampling	CBD (%)	CBG (%)	THC (%)	CBC (%)	Marker Phenotype
Santhica 27 [a]	123	0.26	1.55	0.04	<LOD	B_D/B_D
Carmagnola	151	5.32	0.28	0.16	<LOQ	B_D/B_D
Bernabeo	151	0.74	2.69	0.04	<LOQ	B_D/B_D
Carmaleonte	123	1.92	0.10	0.10	<LOQ	B_D/B_D
CS	151	5.49	0.20	0.17	<LOQ	B_D/B_D
Ermo	151	0.05	<LOD	<LOD	<LOD	B_D/B_D
Fibrante	151	3.61	0.09	0.13	<LOQ	B_D/B_D
Fibranova	151	2.34	0.10	0.07	<LOQ	B_D/B_D
Eletta Campana	151	3.85	0.17	0.11	<LOQ	B_D/B_D
Codimono	123	4.02	0.15	0.14	<LOQ	B_D/B_D
Futura 75 [a]	123	2.34	0.12	0.07	<LOD	B_D/B_D
CINBOL	50	0.01	0.37	3.71	0.01	B_T/B_T

CINRO	50	2.43	0.27	1.57	0.02	B_T/B_D

[a] French varieties.

Total CBG ranged between 0.10% in Carmaleonte to 2.69% w/w in Bernabeo; it was not detected in Ermo. All chemotype III genotypes had THC content below 0.20%, the legal limit for industrial hemp in E.U., while, as expected, the two medical varieties CINBOL and CINRO exceeded the limit up to 3.71% and 1.57%, respectively. Among hemp varieties, CS reached the highest residual THC content (0.17%), followed by Carmagnola and Eletta Campana. The lowest THC content and the highest CBG content was registered for Santhica 27 and Bernabeo (chemotype IV), while no trace of THC was detected in Ermo. CBC was under the limit of quantification for all but the two medical varieties, with a higher content in the CINRO inflorescences (0.02%).

3.3. Diversity in *CBCAS* Sequences

Specific primers were designed following a search for the "fiber-type *THCAS*" reference sequence, AB212830 (Kojoma et al., 2006), in the genome assemblies available on NCBI in 2019, listed in Supplementary Table S1.

Complete Open Reading Frames (ORFs), pseudogenes and fragments of a few hundred base pairs were found, but only the complete sequences, with at least 99% query cover and more than 99% identity were further considered. At least one *CBCAS* homolog putatively encoding for a complete protein was retrieved from each genome assembly, except for Chemdog 91 and LA Confidential.

The alignment of the complete ORFs to the AB212830 sequence revealed the presence of three conserved SNPs, in position 13, 18 and 1628 from the starting codon. The same SNPs were always confirmed aligning the ORFs derived from genome assemblies to the other available putative *CBCAS* genes deposited in public repositories. Based on this finding, a primer pair was designed (*CBCAS*-cds Fw and Rv, Table 2) including these three SNPs, which help to further differentiate these genes from *THCAS* and *CBDAS*. An alignment of one representative sequence for each genome against the reference sequence AB212830 is provided in Supplementary File S1, where these three polymorphisms are highlighted.

Table 2. List of primer pairs used for full length coding sequences (cds) amplification. Forward (Fw) and Reverse (Rv) primer sequences, amplicon length (bp) and references are given. Start and stop codons are underlined for each primer.

Gene	Primer Sequence (5′–3′)	Amplicon Length (bp)	Reference
CBCAS-cds	Fw: TGAAGAAAAATGAATTGCTCAACATTC Rv: ACATAGTATGGGTAGATAATTAATGATGAC	1666	This work
THCAS-cds	Fw: ATGAATTGCTCAGCATTTTCCTT Rv: ATGATGATGCGGTGGAAGA[a]	1635	(Onofri et al., 2015)
CBDAS-cds	Fw: ATGAAGTGCTCAACATTCTCCTT Rv: TTAATGACGATGCCGTGGAA	1635	revised from (Onofri et al., 2015)

[a] the stop codon is immediately after the end of the reverse primer.

In order to investigate the *CBCAS* sequence variability within the set of 13 genotypes chosen in this work, PCR amplicons obtained with specific primers, covering the entire length of the coding sequence, were cloned and Sanger sequencing was performed on recombinant plasmids. All the putative *CBCAS* sequences were obtained from a pool of genomic DNA from 10 individual plants for each genotype; sequencing led to the identification of both pseudogenes and full-length coding sequences. A total of 77 *CBCAS* sequences were obtained, among which 30 unique sequences with a new and complete ORF were submitted to GenBank with ID numbers: MW429515- MW429540, MW429551, MW561073- MW561075.

To properly classify these sequences, they were aligned to the sequences contained in Supplementary File S1 of the work by van Velzen and Schranz (2020) and a phylogenetic analysis based on the nucleotide sequences showed that they belong to the clade A2, comprising the only functionally characterized *CBCAS* and other putative *CBCAS* (Figure 2).

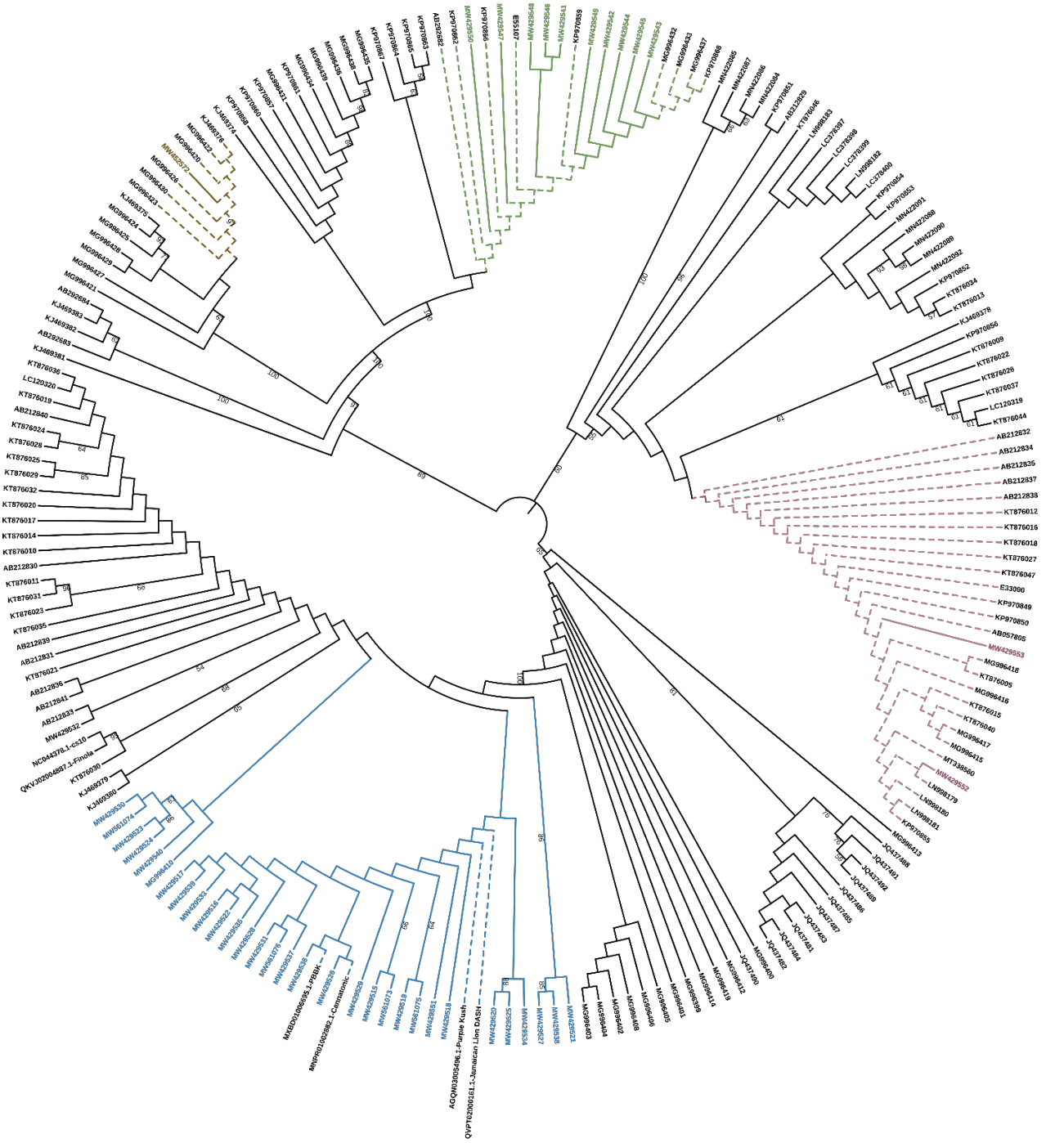


Figure 2. Phylogenetic tree analysis of 202 cannabinoid synthase genes from *C. sativa* at nucleotide level. The tree colors reflect the scheme presented by Van Velzen and Schranz in their Supplementary Figure S1. New *CBCAS* sequenced in the present work are in blue (clade A2), in pink new *THCAS* (clade A1) and in green new CBDAS sequences (clade, B, brown for clade B2-CBDAS). The optimal tree with the sum of branch length = 0.49420587 is shown. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered

together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985).The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. Codon positions included were 1st + 2nd + 3rd + Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1906 positions in the final dataset.

One to six *CBCAS* full cds were found in each Cannabis genotype analyzed in this work (Supplementary Table S2). The highest sequence variability was found in variety CS, with 6 different sequences, followed by Carmagnola, Carmaleonte and Fibrante, which have 5 different sequences each.

Several recurrent sequences were present in more than one genotype, such as MW429551 (Carmaleonte, Fibrante, Santhica 27, Bernabeo, Ermo, CINRO and CINBOL), MW429517 (CS, Futura 75, Carmaleonte), MW429518 (Carmagnola and Fibranova) and MW429528 (CS and Fibrante), whereas the other sequences were identified only in one out of 13 genotypes under investigation.

When compared with the AB212830 reference sequence, the alignment of the 30 new sequences revealed SNPs in 42 different positions of the cds, 28 of them resulting in either amino acid transitions or transversions. Sequences differed by a variable number of SNPs, from three (MW429517, MW429522, MW429528, MW429529, MW429534, MW429551) up to eight (MW429523).

The list of SNPs identified in the *CBCAS* sequences with the corresponding position and any amino acid substitutions is reported in Supplementary Table S3.

The similarity of the new cds with the AB212830 ranges from 99.51% (MW429520, MW429524) to 99.82% (MW429551), while the percentages of similarity with the *THCAS* (E33090) range from 95.72% (MW429523) to 96.02% (MW429534, MW429551).

3.4. Diversity in *THCAS* Sequences

Complete *THCAS* sequences were found only in the two medical Cannabis varieties, CINBOL and CINRO. All the eight sequences retrieved from CINRO (chemotype II) were identical to the reference functional *THCAS* sequence E33090. The variety CINBOL (chemotype I) showed higher variability, as three different sequences were found (E33090, MW429552, MW429553), out of seven sequenced complete cds.

The possible presence of these sequences in databases was checked by BlastN against the standard NCBI non-redundant database, filtered by organism (Cannabis sativa, taxid: 3483) and no sequence with 100% identity was found.

The two novel sequences found in CINBOL variety are respectively 99.94% and 99.88% identical to E33090. Only one SNP (A→986C) identified in MW429553 caused one amino acid change (N→328T) outside the three main domains of the putative translated protein.

3.5. Diversity in *CBDAS* Sequences

Among 65 sequenced clones, 10 unique sequences putatively encoding for functional proteins were identified. From one (Carmagnola, Futura 75, Santhica 27, Bernabeo, CIN-RO) up to four (Carmaleonte, Ermo) different complete coding sequences have been found in each genotype (Supplementary Table S2).

No complete CBDAS cds was identified in CINBOL (chemotype I), in agreement with other studies (Toth et al., 2020; Weiblen et al., 2015; Wenger et al., 2020).

Compared to the reference *CBDAS* gene (GenBank accession E55107), from one to four SNPs were found in 11 different positions. These SNPs differentiating each sequence are listed in Supplementary Table S4 and seven of them cause amino acid substitutions. All sequences share a common SNP respect to the reference sequence in position 105, where G was replaced by T, as already observed (Onofri et al., 2015). This change is synonymous and therefore cannot affect the functionality of the resulting protein.

Overall, the percentages of similarity between all identified sequences and the reference *CBDAS* sequence (E55107) are very high.

The most divergent sequence from E55107 is the *CBDAS* isolated from the chemotype V accession Ermo (MW429543) (99.76%), while the closest sequence is the MW429550 (99.94%), detected in 10 out of 13 genotypes.

3.6. Transcriptional Analysis of Cannabinoid synthase Genes

The transcriptional level of the three classes of cannabinoid biosynthetic genes has been estimated by RT-qPCR in Cannabis inflorescences of all genotypes under study.

Given the high similarity of the target genes, the specificity of the designed primer pairs was accurately validated in three steps. Firstly, an *in-silico* validation through blasting the primers to known reference

sequences ensured that they perfectly matched exclusively sequences corresponding to the target genes. Only primers that generated perfect matches passed through the second validation step. This step was performed by PCR on different templates: genomic DNA, one plasmid containing the specific gene targeted by the primer pair (positive control) and one or more plasmids containing other non-target cannabinoid synthase genes (negative controls). Only the primer pairs that did not produce a band for the negative controls were brought to the next step (Appendix B). The same controls were then used in RT-qPCR analyses. In some cases, the specificity was gained after decreasing the primer concentration. Finally, amplification specificity was validated through melting curve analyses. Primer pairs showing a single high peak, indicating the absence of either non-target amplifications or primer dimers, were selected for the RT-qPCR assay; the optimal settings ensuring only specific target amplification are re-ported in Table 3. The melting curves for each specific amplicon are reported in Appendix B.

Table 3. List of primer pairs used for RT-qPCR. Forward (Fw) and reverse (Rv) primer sequences, amplicon length ex-pressed in base pairs (bp), regression coefficient (R2), PCR efficiency (Eff), annealing temperature (Ta) and usage concentrations (μM) are given.

Gene	Primer Sequence (5′–3′)	Amplicon Length (bp)	R^2	Eff (%)	Ta	[μM]
CBDAS-RTqPCR	Fw: GCAATACACACTTACTTCTCTTCAGTTTTC Rv: ACGTAGTCTAACTTATCTTGAAAGCAC	241	0.99	1.04	61.5	0.075
THCAS-RTqPCR	Fw: AAAACTTCCTTAAATGCTTCTCAA Rv: TAAAATAGTTGCTTGGATATGGGAGTT	198	0.94	0.8	58	0.175
CBCAS-RTqPCR	Fw: GCTCACGACTCACTTCAGAACTAG Rv: GTAGAAGATGGTTGTATCAATCCAGCTC	198	0.98	1.01	62	0.1

The transcriptional level of *THCA synthase* was analyzed in all 13 genotypes, but it could be detected and quantified only in CINBOL and CINRO varieties. The highest amount of THCA synthase transcripts was found in CINBOL (chemotype I, Figure 3a), where it was about 15 times higher than the transcription level

detected in CINRO (chemotype II). On the contrary, CBDA synthase gene was transcribed in almost all samples. The highest transcription level was found for CINRO (R.Q. = 4.9, Figure 3b), while it was undetectable in CINBOL, where no functional gene encoding for a CBDAS was indeed found in this work. The CBDAS transcript levels varied among hemp varieties (R.Q. be-tween 0.28 and 1.81), except for Santhica 27 where a very low transcript level was detected (R.Q. = 0.02).

The transcription level of the putative *CBCAS* genes has been investigated using the primers designed to avoid the amplification of non-target genes as *THCAS* and *CBDAS* (Table 3; Figure 3c) while amplifying all putative isoforms isolated in this work. In general, the CBCAS sequences were transcribed in all genotypes under examination, with a lower transcription level compared to *THCAS* and *CBDAS* genes. Similar levels were observed for Carmagnola, CS, Fibrante, Codimono and Bernabeo; lower ones were detected in Ermo, Santhica 27, Carmaleonte, Futura 75, Eletta Campana, Fibranova, CINRO and CINBOL. Sequencing of RT-qPCR products further confirmed that the expressed genes of CINBOL and CINRO corresponded to the *CBCAS* genes (data not shown).

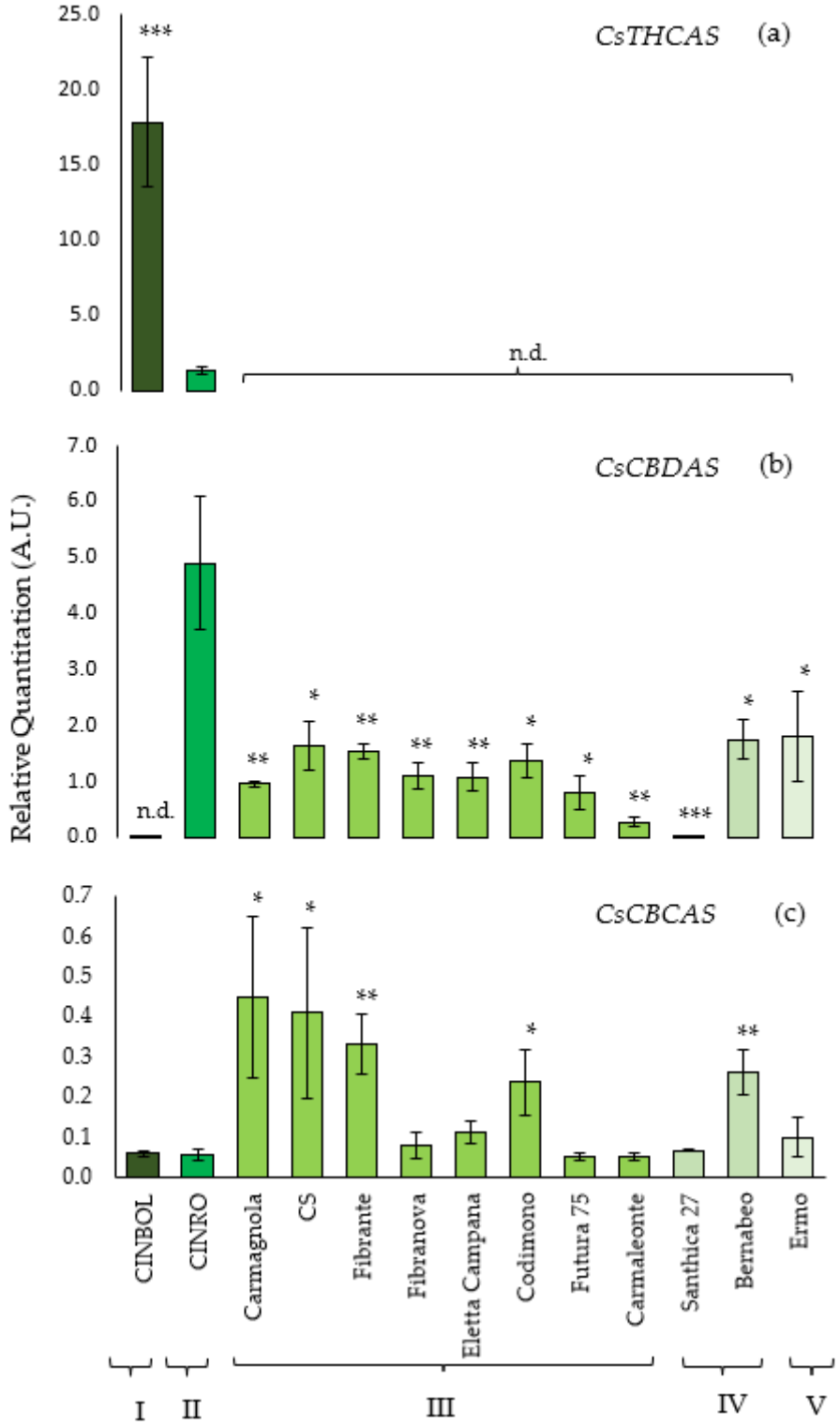


Figure 3. Relative Quantitation (R.Q.) of *CsTHCAS* (a), *CsCBDAS* (b), *CsCBCAS* (c) transcriptional levels in monoecious or female inflorescences. Y axis reports R.Q. expressed in Arbitrary Units (A.U.). Bars

represent the standard errors of the mean of three biological replicates (n = 3). No detectable transcription levels are indicated as n.d. Below, for each genotype the corresponding chemotype is indicated (from I to V). Asterisks indicate Student's t-test statistically significant differences: * $p < 0.05$; ** $p < 0.01$ and *** $p < 0.001$.

In CINRO, the only variety where all three genes are expressed, *THCAS* transcription levels (R.Q.= 1.32) are about four times lower than *CBDAS* levels (R.Q. = 4.90) and 24 times higher than *CBCAS* (R.Q.=0.05). In CINBOL the expression of *THCAS* (R.Q. = 17.80) is about 300× the one of *CBCAS* (R.Q.=0.06).

In general, in industrial hemp genotypes, where *THCAS* could not be detected, *CBDAS* transcript levels were higher than CBCAS ones (from 2× in Carmagnola up to 18× in Ermo). *CBCAS* was transcribed more than *CBDAS* only in Santhica 27.

To the best of our knowledge, this is the first report about the transcription of putative *CBCAS* genes related to the cannabinoid synthases and likely involved in CBCA synthesis and about a low but consistent presence of their mRNA in mature inflorescences in all tested samples, drug- but also hemp genotypes.

4. Discussion

The current knowledge of Cannabis sativa has been shaped by a growing number of studies regarding its therapeutic potential and the wide applications of its main bioactive compounds, mostly cannabinoids, turning this plant from a criminalized drug to a multibillion-dollar business in only a few years (Schluttenhofer and Yuan, 2017) . Recently the United Nations Commission on Narcotic Drugs agreed on the removal of Cannabis from Schedule I of the Con-trolled Substances Act and this is thought to boost research into medical Cannabis, giving the plant a well-deserved attention and recognizing the therapeutic properties of its bioactive metabolites.

As more studies on Eurasian landraces are required to unveil the genetic heterogeneity of the species (Booth and Bohlmann, 2019), the present work focused on a set of Italian and French registered or patented *Cannabis* cultivars. The sequence variability and transcriptional levels of three cannabinoid synthases were investigated, focusing on *CBCAS,* reported for the first time in 2006 as fiber-type *THCAS* (Kojoma et al., 2006) and for many years neglected from the functional and genetic point of view.

In this work, many complete and putatively functional *CBCAS* sequences were retrieved from both fiber- and drug-types genotypes, using a primer pair newly developed after *in-silico* analysis of different publicly available *Cannabis* assemblies.

Interestingly, a variable number of *CBCAS* sequences was obtained from the genotypes under study, ranging from one in Codimono, CINRO and CINBOL to up to six different sequences in the case of CS, some of them recurring (e.g., MW429551, MW429517, MW429518, MW429528) in more than one variety, while others specific to one genotype suggesting a wide distribution and variability among genotypes.

Overall, these putative *CBCAS* sequences confirmed a great conservation of the SNPs differentiating this class of genes from the THCAS genes, as previously reported (Cascini et al., 2019; Kojoma et al., 2006; Weiblen et al., 2015). This diversification from *THCAS* and the high degree of conservation among genotypes would support the hypothesis that these genes could exert a specific role in planta (Laverty et al., 2019).

A role in the biosynthesis of CBCA was demonstrated only for one sequence (Laverty et al., 2019). Whether (and if) the expression of these sequences correlates with the ability of the different varieties to produce CBCA in some stage of their development and consequently if they should all be considered as actually involved in

CBCA biosynthesis in planta, is still to be determined. This is true not only for the sequences identified in this work, but also for all other putative *CBCAS* sequences released so far (van Velzen and Schranz, 2020).

A recent phylogenetic analysis and subsequent classification of cannabinoid synthases has led to the definition of three different clades (A–C) (van Velzen and Schranz, 2020). The *CBCAS* retrieved in this work are included in the A2 clade, comprising full coding CBCAS genes and sequences defined as "incomplete", "inactive" or "obscure" *THCAS*.

As expected, our *CBCAS* are in a separate branch compared to other Sanger sequenced cds, due to the presence of three SNPs at positions 13, 18 and 1628 bp, which de-serve to be added to those previously identified as highly distinctive from *THCAS* (Cascini et al., 2019; Kojoma et al., 2006; Weiblen et al., 2015).

Here, an assay was specifically developed to identify and study the transcription of *CBCAS*, allowing to detect the low expression of these genes in all the analyzed genotypes.

Weakly expressed *CBCAS* were found also in an RNA-seq experiment from female flower buds of nine Cannabis strains, where no detectable amounts of cannabichromene were found (Zager et al., 2019). In addition, Onofri et al., reported a small, yet detectable expression in pure-CBDA accessions (Onofri et al., 2015).

According to Laverty et al. (2019), the highest expression levels of *CBCAS* were found in female flowers and trichomes rather than in leaves. However, the primers used in Laverty's work to analyze the *CBCAS* transcription by RT-qPCR are not specific and can actually target a *CBDAS*, as verified after BlastN vs. non redundant nucleotide database. On the other hand, de Meijer et al. (2009) determined that CBCA accumulates predominantly in juvenile leaves and not in flowers. Taken together, these data made us sceptic about actuality and completeness of knowledge on *CBCAS*, that strongly encourages further analyses and verification of the function of these genes.

Grassa et al. (2021) analyzed cannabinoid synthase genes expression by sequencing full length cDNA transcripts from high-CBD cultivars CBDRx (in leaves) and First Light (in flowers). No open reading frame for *CBCAS* was found, probably due to the low transcriptional level of the gene in the analyzed samples, compared to other cannabinoid synthases.

Braich et al., using an RNA-seq approach, did not reveal *CBCAS* expression; however, this could be probably due to the misleading bioinformatics analyses of the short fragments deriving from paralogues and/or pseudogenes, all mapped to a same locus (Braich et al., 2019).

The analyses of transcriptional levels by using highly specific primers, as done in the present work, avoid the creation of misleading artifacts sometimes resulting by using short reads and coupled bioinformatics methods to analyze Cannabis transcriptomic massive outputs.

Different studies suggested the possibility that residual THCA in hemp plants is an apparently unavoidable by-product of CBDA synthesis by CBDAS; THCAS and CBDAS, in fact, have very similar kinetic characteristics (Vmax, KM, turnover) and this might explain the simultaneous formation of these two cannabinoids from the common precursor CBGA (Zirpel et al., 2018) , though with different efficiencies. However, assuming this hypothesis was true, the issue about the existence of highly inbred lines accumulating practically only one single cannabinoid (Mandolino and Carboni, 2004) would remain unexplained.

An alternative explanation is that, given their ubiquity in *Cannabis* germplasm, *CBCAS* sequences likewise could be responsible for the low levels of THCA accumulation in hemp varieties at maturity, with occasional overcoming of the 0.20% d.w. THC thresh-old in some cultivations. The availability of specific markers, such as those presented in this work, could allow the counter-selection of these sequences during breeding, or their silencing by biotechnological strategies in a gene-specific and copy number-independent way, in order to definitively clarify the issue.

Comparing protein sequences of THCAS with CBCAS found in this work showed that there are no amino acid substitutions at residues necessary for the protein activity (from Arg110 to His114 and Cys176) or in residues whose substitution could affect its functionality, as demonstrated before (Shoyama et al., 2012; Sirikantaramas et al., 2004). Therefore, the possibility cannot be ruled out that the resulting CBCAS proteins, when translated, synthetize some THCA from CBGA precursor.

Based on our results, it appears that *CBCAS* is transcribed at low levels in all genotypes and that the THCA amount is also low in hemp varieties; it cannot be overlooked that varieties accumulating higher residual THC (Carmagnola, CS, Fibrante, Codimono) also share higher amounts of CBCAS transcripts (Table 1, Figure 3c).

Opposite to *CBCAS*, functional *THCAS* have been isolated in this work only from the medical varieties CINRO and CINBOL. In CINRO, only a THCAS sequence was retrieved from the genome, suggesting the presence of a single copy gene. Three different complete sequences of THCAS have been obtained from chemotype I CINBOL, suggesting the presence of at least two and up to three distinct copies of the gene. Both new sequences belong to the Subclade A1 from van Velzen and Schranz, with Type 1/1 (MW429552) and Type 1/2 (MW429553) differing by a single conservative substitution and therefore probably encoding for a functionally equivalent protein (van Velzen and Schranz, 2020) .

It has been suggested that reported variations in *THCAS* copy number had to be at-tributed to the use of promiscuous primers or to the unspecific inclusion of *CBCAS* sequences instead of only functional *THCAS* (van Velzen and Schranz, 2020). In the present study, the selective amplification and sequencing of cds from clones and their phylogenetic analysis represent stronger evidence of the presence of more than one copy of THCAS gene per single genotype (CINBOL).

Moreover, a Copy Number Variation (CNV) between CINBOL and CINRO was sup-ported by preliminary RT-qPCR data, showing that the amplification plot from the genomic DNA of CINBOL crossed the threshold line earlier than those from CINRO (data not shown), as expected if CINBOL had more copies of the target gene.

Whether or not this hypothetical CNV has a direct effect on the higher transcriptional level of *THCAS* and the higher amount of THCA in CINBOL in respect to CINRO is still to be determined. Other players like transcription factors or environmental conditions may be responsible for differences in transcriptional levels of these genes and cannabinoid accumulation (Liu et al., 2021).

Regarding the *CBDA synthases* class, a *CBDAS* gene has been sequenced (MW429549), with a SNP at position 1426, responsible for an amino acid substitution at position 476 of the enzyme (Pro→Ser). This substitution was already observed in Ermo and in another CBG-prevalent accession by Onofri et al. (2015) and might result in a minimally functional CBDAS, unable to synthesize CBDA and leading to the accumulation of CBGA. This sequence probably accounts for chemotype IV when there are no other functional CBDAS, as in the case of Santhica 27 and Bernabeo.

5. Conclusions

This work describes for the first-time sequence variability of cannabinoid synthases coding sequences and their transcriptional profiles in a set of *Cannabis* genotypes, belonging to five different chemotypes. Highly specific primers were developed, to be used as robust markers for univocal identification and analysis of *THCAS*, *CBDAS* and *CBCAS* in different *Cannabis* tissues. While the higher levels of transcription detected for *THCAS* in respect to *CBDAS* can be related to the different end use of THCA-rich varieties compared with fiber varieties, the meaning of the low but almost constant level of transcription of *CBCAS* still remains to be fully understood. The existence of a cluster of *CBCA synthases* separated from functional *THCA synthases* suggest a different role in the plant as evolution and/or selection has created a whole set of highly conserved sequences. According to our data, *CBCAS* could play a role for producing residual THCA in CBD-predominant genotypes, but more research is still needed to confirm this hypothesis. Altogether, the results reported here and the assay developed will pave the way for more accurate functional studies in planta of these gene families and of how their reciprocal relationships can influence the quantitative component of chemotype.

6. Funding

This research was funded by MAIDET research project "Metodi analitici innovativi per la determinazione del THC in matrici vegetali", funded by the Ministero delle Politiche Agricole Alimentari e Forestali (MIPAAF), grant decree D.M.34176/7303/2017 and UNIHEMP research project "Use of iNdustrIal Hemp biomass for Energy and new biocheMicals Production" (ARS01_00668) funded by European Regional Development Fund (within the PON R&I 2017-2020–Axis 2—Action II—OS 1.b). Grant decree UNIHEMP prot. no. 2016 of 27/07/2018; CUP B76C18000520005.

7.References

Baratta, F., Simiele, M., Pignata, I., Enri, L.R., Torta, R., de Luca, A., Collino, M., D'Avolio, A., Brusa, P., 2019. Development of standard operating protocols for the optimization of cannabis-based formulations for medical purposes. Front Pharmacol 10, 1–10. https://doi.org/10.3389/fphar.2019.00701

Bassolino, L., Buti, M., Fulvio, F., Pennesi, A., Mandolino, G., Milc, J., Francia, E., Paris, R., 2020. In Silico Identification of MYB and bHLH Families Reveals Candidate Transcription Factors for Secondary Metabolic Pathways in Cannabis sativa L. Plants 9, 1540. https://doi.org/10.3390/plants9111540

Booth, J.K., Bohlmann, J., 2019. Terpenes in Cannabis sativa – From plant genome to humans. Plant Science 284, 67–72. https://doi.org/10.1016/j.plantsci.2019.03.022

Borroto Fernandez, E., Peterseil, V., Hackl, G., Menges, S., de Meijer, E., Staginnus, C., 2020. Distribution of Chemical Phenotypes (Chemotypes) in European Agricultural Hemp (Cannabis sativa L.) Cultivars. J Forensic Sci 65, 715–721. https://doi.org/10.1111/1556-4029.14242

Braich, S., Baillie, R.C., Jewell, L.S., Spangenberg, G.C., Cogan, N.O.I., 2019. Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis. Sci Rep 9, 1–12. https://doi.org/10.1038/s41598-019-53023-6

Bustin, S.A., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G., Wittwer, C.T., Schjerling, P., Day, P.J., Abreu, M., Aguado, B., Beaulieu, J.F., Beckers, A., Bogaert, S., Browne, J.A., Carrasco-Ramiro, F., Ceelen, L., Ciborowski, K., Cornillie, P., Coulon, S., Cuypers, A., de Brouwer, S., de Ceuninck, L., de Craene, J., de Naeyer, H., de Spiegelaere, W., Deckers, K., Dheedene, A., Durinck, K., Ferreira-Teixeira, M., Fieuw, A., Gallup, J.M., Gonzalo-Flores, S., Goossens, K., Heindryckx, F., Herring, E., Hoenicka, H., Icardi, L., Jaggi, R., Javad, F., Karampelias, M., Kibenge, F., Kibenge, M., Kumps, C., Lambertz, I., Lammens, T., Markey, A., Messiaen, P., Mets, E., Morais, S., Mudarra-Rubio, A., Nakiwala, J., Nelis, H., Olsvik, P.A., Pérez-Novo, C., Plusquin, M., Remans, T., Rihani, A., Rodrigues-Santos, P., Rondou, P., Sanders, R., Schmidt-Bleek, K., Skovgaard, K., Smeets, K., Tabera, L., Toegel, S., van Acker, T., van den Broeck, W., van der Meulen, J., van Gele, M., van Peer, G., van Poucke,

M., van Roy, N., Vergult, S., Wauman, J., Tshuikina-Wiklander, M., Willems, E., Zaccara, S., Zeka, F., Vandesompele, J., 2013. The need for transparency and good practices in the qPCR literature. Nat Methods 10, 1063–1067. https://doi.org/10.1038/nmeth.2697

Cascini, F., Farcomeni, A., Migliorini, D., Baldassarri, L., Boschi, I., Martello, S., Amaducci, S., Lucini, L., Bernardi, J., 2019. Highly predictive genetic markers distinguish drug-type from fiber-type cannabis sativa L. Plants 8, 1–12. https://doi.org/10.3390/plants8110496

Citti, C., Ciccarella, G., Braghiroli, D., Parenti, C., Vandelli, M.A., Cannazza, G., 2016. Medicinal cannabis: Principal cannabinoids concentration and their stability evaluated by a high performance liquid chromatography coupled to diode array and quadrupole time of flight mass spectrometry method. J Pharm Biomed Anal 128, 201–209. https://doi.org/10.1016/j.jpba.2016.05.033

Citti, C., Linciano, P., Russo, F., Luongo, L., Iannotta, M., Maione, S., Laganà, A., Capriotti, A.L., Forni, F., Vandelli, M.A., Gigli, G., Cannazza, G., 2019. A novel phytocannabinoid isolated from *Cannabis sativa* L. with an in vivo cannabimimetic activity higher than Δ9-tetrahydrocannabinol: Δ9-Tetrahydrocannabiphorol. Sci Rep 9, 1–13. https://doi.org/10.1038/s41598-019-56785-1

Combest, M.M., Moroz, N., Tanaka, K., Rogan, C.J., Anderson, J.C., Thura, L., Rakotondrafara, A.M., Goyer, A., 2021. StPIP1, a PAMP-induced peptide in potato, elicits plant defenses and is associated with disease symptom severity in a compatible interaction with Potato virus y. J Exp Bot 72, 4472–4488. https://doi.org/10.1093/jxb/erab078

de Meijer, E.P.M., Hammond, K.M., Micheler, M., 2009. The inheritance of chemical phenotype in Cannabis sativa L. (III): Variation in cannabichromene proportion. Euphytica 165, 293–311. https://doi.org/10.1007/s10681-008-9787-1

Felsenstein, J., 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution (N Y) 39. https://doi.org/10.2307/2408678

Gao, S., Wang, B., Xie, S., Xu, X., Zhang, J., Pei, L., Yu, Y., Yang, W., Zhang, Y., 2020. A high-quality reference genome of wild Cannabis sativa. Hortic Res. https://doi.org/10.1038/s41438-020-0295-3

Grassa, C.J., Weiblen, G.D., Wenger, J.P., Dabney, C., Poplawski, S.G., Motley, S.T., Michael, T.P., 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

Gülck, T., Møller, B.L., 2020. Phytocannabinoids: Origins and Biosynthesis. Trends Plant Sci 25, 985–1004. https://doi.org/10.1016/j.tplants.2020.05.005

Hurgobin, B., Tamiru-Oli, M., Welling, M.T., Doblin, M.S., Bacic, A., Whelan, J., Lewsey, M.G., 2020. Recent advances in Cannabis sativa genomics research , New Phytologist. https://doi.org/10.1111/nph.17140

Kojoma, M., Seki, H., Yoshida, S., Muranaka, T., 2006. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" Cannabis sativa L. Forensic Sci Int 159, 132–140. https://doi.org/10.1016/j.forsciint.2005.07.005

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35. https://doi.org/10.1093/molbev/msy096

Larionov, A., Krause, A., Miller, W.R., 2005. A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 6. https://doi.org/10.1186/1471-2105-6-62

Laverty, K.U., Stout, J.M., Sullivan, M.J., Shah, H., Gill, N., Holbrook, L., Deikus, G., Sebra, R., Hughes, T.R., Page, J.E., van Bakel, H., 2019. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Res 29, 146–156. https://doi.org/10.1101/gr.242594.118

Linciano, P., Citti, C., Luongo, L., Belardo, C., Maione, S., Vandelli, M.A., Forni, F., Gigli, G., Laganà, A., Montone, C.M., Cannazza, G., 2020a. Isolation of a High-Affinity Cannabinoid for the Human CB1 Receptor from a Medicinal Cannabis sativa Variety: Δ9-Tetrahydrocannabutol, the Butyl Homologue of Δ9-Tetrahydrocannabinol. J Nat Prod 83, 88–98. https://doi.org/10.1021/acs.jnatprod.9b00876

Linciano, P., Citti, C., Russo, F., Tolomeo, F., Lagan, A., Capriotti, A.L., Luongo, L., Iannotta, M., Belardo, C., Maione, S., Forni, F., Vandelli, M.A., Gigli, G., Cannazza, G., 2020b. Identification of a new cannabidiol n hexyl homolog in a medicinal cannabis variety with an antinociceptive activity in mice : cannabidihexol 1–11. https://doi.org/10.1038/s41598-020-79042-2

Liu, Y., Zhu, P., Cai, S., Haughn, G., Page, J.E., 2021. Three novel transcription factors involved in cannabinoid biosynthesis in Cannabis sativa L. Plant Mol Biol 106, 49–65. https://doi.org/10.1007/s11103-021-01129-9

Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., de Meijer, E., 2003. Qualitative and quantitative aspects of the inheritance of chemical phenotype in cannabis. Journal of Industrial Hemp 8, 51–72. https://doi.org/10.1300/J237v08n02_04

Mandolino, G., Carboni, A., 2004. Potential of marker-assisted selection in hemp genetic improvement. Euphytica 140, 107–120. https://doi.org/10.1007/s10681-004-4759-6

Mangeot-Peter, L., Legay, S., Hausman, J.F., Esposito, S., Guerriero, G., 2016. Identification of reference genes for RT-qPCR data normalization in Cannabis sativa stem tissues. Int J Mol Sci 17. https://doi.org/10.3390/ijms17091556

McKernan, K.J., Helbert, Y., Kane, L.T., Ebling, H., Zhang, L., Liu, B., Eaton, Z., McLaughlin, S., Kingan, S., Baybayan, P., Concepcion, G., Jordan, M., Riva, A., Barbazuk, W., Harkins, T., 2020. Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. bioRxiv. https://doi.org/10.1101/2020.01.03.894428

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., n.d. PURIFICATION AND CHARACTERIZATION OF CANNABICHROMENIC ACID SYNTHASE FROM CANNABIS SATIVA.

Onofri, C., de Meijer, E.P.M., Mandolino, G., 2015. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. Phytochemistry 116, 57–68. https://doi.org/10.1016/j.phytochem.2015.03.006

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., Mandolino, G., 2006. Genetics and marker-assisted selection of the chemotype in Cannabis sativa L. Molecular Breeding 17, 257–268. https://doi.org/10.1007/s11032-005-5681-x

Pagliarani, G., Paris, R., Arens, P., Tartarini, S., Ricci, G., Smulders, M.M.J., van de Weg, W.E., 2013. A qRT-PCR assay for the expression of all Mal d 1 isoallergen genes. BMC Plant Biol 13, 1. https://doi.org/10.1186/1471-2229-13-51

Schluttenhofer, C., Yuan, L., 2017. Challenges towards Revitalizing Hemp: A Multifaceted Crop. Trends Plant Sci 22, 917–929. https://doi.org/10.1016/j.tplants.2017.08.004

Shoyama, Yoshinari, Tamada, T., Kurihara, K., Takeuchi, A., Taura, F., Arai, S., Blaber, M., Shoyama, Yukihiro, Morimoto, S., Kuroki, R., 2012. Structure and function of Δ1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of Cannabis sativa. J Mol Biol 423, 96–105. https://doi.org/10.1016/j.jmb.2012.06.030

Sirikantaramas, S., Morimoto, S., Shoyama, Yoshinari, Ishikawa, Y., Wada, Y., Shoyama, Yukihiro, Taura, F., 2004. The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. Journal of Biological Chemistry 279, 39767–39774. https://doi.org/10.1074/jbc.M403693200

Staginnus, C., Zörntlein, S., de Meijer, E., 2014. A PCR marker Linked to a THCA synthase polymorphism is a reliable tool to discriminate potentially THC-rich plants of cannabis sativa L. J Forensic Sci 59, 919–926. https://doi.org/10.1111/1556-4029.12448

Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences 101. https://doi.org/10.1073/pnas.0404206101

Taura, F., Morimoto, S., Shoyama, Y., 1996. Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of

cannabigerolic acid to cannabidiolic acid. Journal of Biological Chemistry 271, 17411–17416. https://doi.org/10.1074/jbc.271.29.17411

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K.C., Smart, L.B., 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. GCB Bioenergy 12, 213–222. https://doi.org/10.1111/gcbb.12667

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2007. Enhancements and modifications of primer design program Primer3. Nucleic Acids Res 40, 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., Page, J.E., 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biol 12. https://doi.org/10.1186/gb-2011-12-10-r102

van Velzen, R., Schranz, M.E., 2020. Origin and evolution of the cannabinoid oxidocyclase gene family. bioRxiv. https://doi.org/10.1101/2020.12.18.423406

Vergara, D., Huscher, E.L., Keepers, K.G., Givens, R.M., Cizek, C.G., Torres, A., Gaudino, R., Kane, N.C., 2019. Gene copy number is associated with phytochemistry in Cannabis sativa. bioRxiv. https://doi.org/10.1101/736181

Weiblen, G.D., Wenger, J.P., Craft, K.J., ElSohly, M.A., Mehmedic, Z., Treiber, E.L., Marks, M.D., 2015. Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytologist 208, 1241–1250. 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. Am J Bot 107, 1423–1432. https://doi.org/10.1002/ajb2.1550

Xie, F., Xiao, P., Chen, D., Xu, L., Zhang, B., 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol 80, 75–84. https://doi.org/10.1007/s11103-012-9885-2

Zager, J.J., Lange, I., Srividya, N., Smith, A., Lange, B.M., 2019. Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis. Plant Physiol 180, 1877–1897. 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. J Biotechnol 284, 17–26. https://doi.org/10.1016/j.jbiotec.2018.07.031

Chapter 2

Assessing expression of genes involved in the cannabinoid pathway during early vegetative stages to gain insight into cannabichromenic acid biosynthesis

This chapter will be submitted to Phytochemistry by Flavia Fulvio[1,2], Giuseppe Mandolino[1], Cinzia Citti[3,4], Nicola Pecchioni[5], Giuseppe Cannazza[3,4], Roberta Paris[1]

1 CREA—Research Centre for Cereal and Industrial Crops, Via di Corticella 133, 40128 Bologna, Italy;

2 Department of Sciences of Agriculture, Food Natural Resources and Engineering, University of Foggia, Via Napoli 25, 71122 Foggia, Italy;

3 CNR NANOTEC—Institute of Nanotechnology, Via Monteroni, 73100 Lecce, Italy;

4 Department of Life Science, University of Modena and Reggio Emilia, Via G. Campi 103, 41125 Modena, Italy

5 CREA—Research Centre for Cereal and Industrial Crops, S.S. 673 Km 25,200, 71122 Foggia, Italy.

Abstract: *Cannabis sativa* (L.) is characterized by great genetic and phenotypic diversity, also expressed in the array of bioactive compounds synthesized. Despite its great potential economic interest, knowledge of the biology and genetics of this crop is incomplete, and still many efforts are needed for a complete understanding of the molecular mechanisms regulating its key traits. To better understand the synthesis of these compounds, we analysed the transcription levels of cannabinoid pathway genes, during early phases of plant development, then comparing the transcriptional results with a chemical characterization of the same samples.

The work was conducted on both industrial and medicinal C. sativa plants, using samples belonging to three different chemotypes. Genes coding for the cannabinoid synthases, involved in the last step of the cannabinoid biosynthetic pathway, were found to be already expressed in the seed, providing a measure of the importance of this metabolism for the plant. Cannabichromenic acid is known as the first cannabinoid accumulating in the seedlings, shortly after emergence, and it was found that there is a good correspondence between transcript accumulation of the cannabichromenic acid synthase gene and accumulation of the corresponding metabolite.

1.Introduction

The potential of medical cannabis has been known for millennia and its therapeutic applications have been made legal by an increasing number of countries around the world (Kumar et al., 2021). As a result, the scientific community is exploring the full potential of this plant and of its wealth of bioactive compounds, starting from phytocannabinoids. These molecules are unique to the *Cannabis* genus; more than 150 phytocannabinoids are currently known (Hanuš et al., 2016), but to date most clinical research has been focused on the two cannabinoids, accumulated in greater amounts in the plant female inflorescences: cannabidiolic acid (CBDA) and Δ^9-tetrahydrocannabinolic acid (Δ^9-THCA), and their two neutral derivatives CBD and Δ^9-THC.

Their biosynthesis was recently well summarised by Tahir et al. (2021): the synthesis starts from the hexanoyl-CoA which is condensed with three molecules of malonyl-CoA by the action of a tetraketide synthase (TKS) (Kearsey et al., 2020), a type III polyketide synthase (PKS). The cyclization is catalyzed by the olivetolic acid cyclase (OAC) and the loss of coenzyme A leads to the formation of olivetolic acid, the phenolic portion of phytocannabinoids. The aromatic prenyltransferase (PT) then inserts a terpenoid prenyl group to form cannabigerolic acid (CBGA). This compound is the substrate for several oxidocyclases, to generate the different final cannabinoids (Tahir et al., 2021) .

Non-enzymatic decarboxylation reactions (e.g. long-term storage and heating driven) transform the acidic, native forms of cannabinoids into their neutral bioactive forms, capable of interaction with receptors of the human and animal endocannabinoid system (Piscitelli and Marzo, 2015).

After Δ^9-THCA and CBDA, cannabichromenic acid (CBCA) is one of the most abundant cannabinoids reported in *C. sativas* and, together with its neutral form, cannabichromene (CBC), is currently being evaluated in preclinical research for its pharmacological properties (Walsh et al., 2021).

Anti-inflammatory effects have been reported for CBC (Izzo et al., 2012; Maione et al., 2011) and its positive action on brain health exerted on neural stem cells (Shinjyo and di Marzo, 2013).

CBC showed anti-inflammatory potential in activated macrophages (Romano et al., 2013) and exhibited strong effects against acne *in vitro* (Oláh et al., 2016) .

Recent studies demonstrated that CBC shows remarkable brain penetration and anticonvulsant efficacy in

mouse models, suggesting that the anticonvulsant action of medicinal cannabis preparations may be mediated by the presence of this compound in the phytocomplex (Anderson et al., 2021) . In addition, CBCA has shown antibacterial efficacy against the methicillin-resistant bacterium *Staphylococcus aureus* (Galletta et al., 2020) .

Although some molecular targets of CBC and CBCA have been proposed (CB2 receptors, transient receptor potential vanilloid (TRPV) channels), the characterization of their pharmacological actions is still ongoing (Etchart et al., 2022) .

Up to now, very little research on CBCA *in planta* has been conducted, and detailed information regarding its biosynthesis and regulation are still lacking. The CBCA Synthase enzyme was identified for the first time in 1996 and isolated from young leaves of 2-weeks-old plants (Morimoto et al., 1997). Its purification allowed the determination of its features as a soluble, homodimer enzyme consisting of two monomers of 71 kDa each, with higher affinity for CBGA substrate and lower catalytic capacity than THCA and CBDA synthases (Morimoto et al., 1998).

The lack of information on CBC metabolism is mostly due to the inherently transient accumulation of this cannabinoid in the plant. In fact, according to previous studies, the highest amounts of CBC can be detected in the juvenile stages of *C. sativa* seedlings (de Meijer et al., 2009; Morimoto et al., 1997, 1998); however, no studies on the time course of synthesis and accumulation in different plant organs during early stages of development and in different *C. sativa* chemotypes have so far been carried out.

In a previous work, we described novel putative *CBCAS* sequences isolated from high-THC and hemp genotypes, discussing their transcription in mature inflorescences, and suggesting that the transcription of these genes might contribute to the final chemotype (Fulvio et al., 2021).

The development of *C. sativa* genotypes oriented towards the production of minor cannabinoids, is one of the challenges needing to be addressed in *C. sativa* breeding (Backer et al., 2020) . Since a demand for high-CBC producing plants is expected as the pharmacological research on the properties of this cannabinoid proceeds, understanding its mechanism of synthesis will allow the development of faster and more efficient systems to increase its availability and production efficiency through advanced breeding.

Moreover, knowledge about the cannabinoid composition of different plant tissues and organs, along with their correlation with age, sex, developmental stage, and interaction with biotic and abiotic factors, needs to be deepened in order to fully understand the biological significance of these compounds in plant physiology.

In this work, different plant tissues at different stages of development were analysed by assaying their cannabinoid content and profile together with the expression of key cannabinoid synthesis genes to better understand the cannabinoid biosynthesis and with a specific emphasis on the synthesis of CBCA.

2. Materials and methods

2.1 Plant material

Four different genotypes were used in this work, chosen from a *C. sativa* germplasm collection owned and maintained at the CREA-Research Centre for Cereals and Industrial Crops in Rovigo (Italy): 1. a chemotype I accession (V9S) with high THC content, 2. a chemotype III accession, selected for a high CBDA accumulation (V19); 3. a chemotype III hemp registered variety (Fibrante); and 4. a chemotype V genotype (Ermo) with total cannabinoids <0.3% w/dw. Even if from a botanical point of view, *C. sativa* fruits are achenes; in this work, from here on, for simplicity and uniformity with the BBCH scale, they will be referred to as seeds (Mishchenko et al., 2017; Small, 1975). The seeds of each genotype were stored at 4°C for different periods before utilization. They were sterilized with a solution of 5 % sodium hypochlorite for 15 min with intermittent agitation, followed by two consecutive rinsing in sterile water. A batch of seed for each genotype was sown in a peat-type tray substrate (Klasmann-Deilmann Gmbh) and seedlings were grown in a growth chamber (Percival®AR-36LC8, Percival scientific, Inc. Perry, IA, USA), with 18-h light, 6 h dark photoperiod and a temperature of 24 °C.

For each genotype, samples were collected at six phenological stages of growth and development, defined based on the BBCH scale (Mishchenko et al., 2017) during germination, sprouting and leaf development main stages as illustrated in Figure 1: 00: dry seeds; 05: radicle emerged from seed; 09: cotyledons and hypocotyls; after emergence, the last expanded leaf couple were collected when the seedling was at stages: 11: one true leaf pair; 12: two leaf pairs; 13: three leaf pairs (Fig. 1). Three biological repetitions of each sample were collected, each consisting in a pool of samples taken from 5 to 10 different individuals at the same stage,

immediately frozen and stored at -80°C in order to preserve the chemical composition and the transcriptional profile, for HPLC-UV and RT-qPCR analyses, respectively. For chemical analysis, mature inflorescences, collected at the BBCH-67 stage, were also analyzed for each genotype to confirm chemotype classification; for cannabinoid quantification, samples were used after lyophilization.

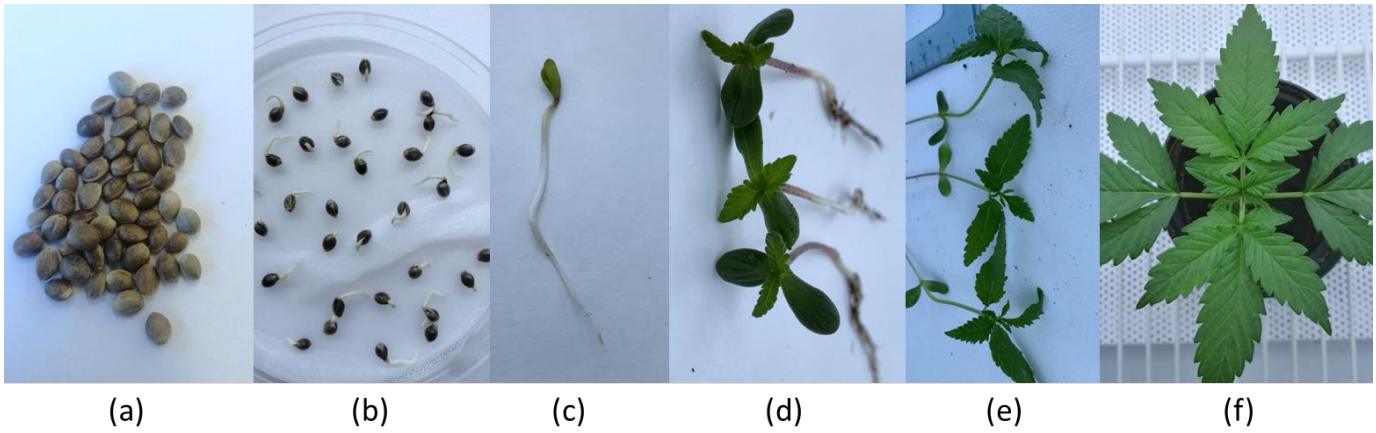


Figure 1. *C. sativa* samples collected for molecular and biochemical analyses. A) BBCH stage 00, B) BBCH stage 05; C) BBCH stage 09; D) BBCH stage 11; E) BBCH stage 12; F) BBCH stage 13.

2.2 HPLC-UV analysis

For quantitative analysis of cannabinoids samples were prepared and analyzed by liquid chromatography coupled with UV detection (HPLC-UV) according to the protocol of the German Pharmacopoeia as already reported (Citti et al., 2016; Linciano et al., 2020; Linciano et al., 2021). Samples were injected in triplicate after dilution in the mobile phase at the concentration of 1 mg of biomass per mL. The limit of detection (LOD) was 0.001% and the limit of quantification (LOQ) was 0.005%. Analytical standards of CBDA, cannabigerolic acid (CBGA), THCA, CBD, cannabigerolo (CBG), Δ^9-THC, Δ^8-THC, CBCA, and CBC were purchased from Cayman Chemical (Ann Arbor, Michigan, USA). Reag. Ph Eur ethanol 96 % and LC-MS grade acetonitrile (ACN) and formic acid (FA) were all bought from Merck (USA). Ultrapure water was obtained with a water purification system (Direct-Q 3UV, Merck Millipore, Milan, Italy).

The analyses of cannabinoids were carried out on a Vanquish Core System (Thermo Fisher Scientific, Waltham, Massachusetts, USA) equipped with a binary pump, a vacuum degasser, thermostated autosampler

(4 °C) and column compartment (30 °C) and a diode array detector (DAD) set at 228 nm. Chromatographic separation was achieved on a Poroshell 120 EC C18 (100 × 3.0 mm, 2.7 µm, Agilent Technologies (Santa Clara, California, USA) using a previously validated method for cannabinoids with a gradient elution of ACN from 5% to 95% in 20 min, followed by an isocratic step at 95% ACN for 3 min and a washing step at 98% ACN for 7 min (Tolomeo et al., 2022). The column was re-equilibrated at 5% ACN in 6 min for a total run time of 36 min. The flow rate was 0.5 mL/min and the injection volume was 5 µL.

Results of the concentrations of cannabinoids are reported as the sum of carboxylated (A) and decarboxylated (B) species according to the following formula: Total = A*0.877+B.

2.3 RNA isolation and cDNA synthesis

Total RNA was isolated starting from 100 mg frozen samples as previously described (Fulvio et al., 2021). After treatment of 100 ng RNA with 1 unit of DNAse I Amplification Grade (Sigma-Aldrich, Merck Life Science S.r.l., Milan, Italy), the cDNA was synthesized using High-Capacity RNA to cDNA kit (Thermo Fisher Scientific,Waltham, MA, USA) according to manufacturer's instructions in 20 µL final volume.

The cDNA was finally diluted 1:9 and 2 µL were used in RT-qPCR reactions.

2.4 Transcriptional analysis of cannabinoid synthases

Transcriptional analyses were performed by RT-qPCR using the Rotor-Gene 6000 (Corbett) and the SYBR Green chemistry, on three biological replicates and two technical replicates. Each reaction contained 2 µL of cDNA, 5 µL of Power Up®SYBR master mix (Thermo Fisher Scientific), specific primers (0.3 Mm) and sterile water to a final volume of 10 µL. A standard curve made of 5 points was added in all assays, prepared as four 1:4 serial dilutions of a 1:8 dilution of the Cdna (100 ng). Amplification reactions were performed using an annealing temperature of 60°C. Raw data for target and reference genes were transformed using the 'Standard Curve Method' (Larionov et al., 2005) ; the transcripts level of target genes was normalized to the geometric mean of the transcripts level of glyceraldehyde 3-phosphate dehydrogenase (*CsGAPDH*) and *CsF-box* (Mangeot-Peter et al., 2016), selected after RefFinder results as the most stable genes across samples.

Transcription data were reported as Relative Quantitation (R.Q.) of transcriptional levels expressed in

Arbitrary Units (A.U.). Primers for target genes were designed using the software Primer3 and reported in Supplemental Table 1.

3.Results

3.1 Phytocannabinoids profile in seedling stages of *C. sativa*

The amount of the main cannabinoids was measured by HPLC-UV. As expected, in seeds and in seeds with just an emerging radicle, no phytocannabinoids were detectable in any of the *C. sativa* genotypes. The first BBCH stage in which phytocannabinoids (except CBC) were detectable was found to differ slightly depending on the genotype. CBC was constantly the first cannabinoid to be produced, irrespective of the final chemotype of the plant (Table 1, Fig. 2), whereas it was the least abundant cannabinoid in mature inflorescences in all genotypes at stage 67, when, with the remarkable exception of Ermo, its amount stabilizes close to 0.3 % w/dw. The total amount of each cannabinoid in the different stages and genotypes, intended as the sum of carboxylated and decarboxylated species and calculated as described in the Materials and Methods section, is reported in Table 1.

Table 1. Quantitative results of total cannabinoid content of C. sativa genotypes at different growth stages, mean of 3 replicates. LOD (limit of detection): 0.001%. LOQ (limit of quantification): 0.005%. Data are expressed as a percent of total inflorescence dry weight (g/100 g biomass). SD was <0.01. The last column (PCBC) shows the percentage of CBC over the total cannabinoid content for each sample. Phytocannabinoid quantification at maturity (BBCH-67) is reported solely to confirm the chemotype of each genotype under investigation.

Sample	CBG	CBD	THC	CBC	Total	P CBC
V9-00	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
V9-05	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
V-09	<LOQ	<LOQ	<LOQ	0.05	0.05	100%
V9-11	0.09	0.06	0.08	0.43	0.66	65%
V9-12	0.08	0.1	0.14	0.52	0.84	62%
V9-13	0.57	0.69	0.56	0.34	2.16	16%
V9-67	0.74	0.16	6.85	0.28	8.03	3%
V19-00	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
V19-05	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
V19-09	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
V19-11	0.07	<LOQ	<LOQ	0.65	0.72	90%
V19-12	0.22	0.25	0.05	1.73	2.25	77%
V19-13	0.4	1.3	0.1	0.37	2.17	17%
V19-67	0.12	9.10	0.36	0.35	9.93	4%
Fibrante-00	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
Fibrante-05	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
Fibrante-09	<LOD	<LOD	<LOD	0.01	0.01	100%
Fibrante-11	0.02	0.02	<LOD	0.07	0.11	64%
Fibrante-12	0.08	0.12	<LOD	0.07	0.28	25%
Fibrante-13	0.34	0.54	0.02	0.32	1.22	26%
Fibrante-67	0.19	4.88	0.28	0.30	5.65	5%

Ermo-00	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
Ermo-05	<LOQ	<LOQ	<LOQ	<LOQ	0	0%
Ermo-09	<LOD	<LOD	<LOD	<LOD	0	0%
Ermo-11	0.02	<LOD	0.04	0.08	0.13	62%
Ermo-12	<LOD	<LOD	0.07	0.07	0.13	54%
Ermo-13	<LOD	<LOD	<LOD	0.01	0.01	100%
Ermo-67	<LOD	0.05	<LOD	<LOD	0.05	0%

Total cannabinoid content per plant ranged from 0% to >9% w/dw, depending on the stage examined. V9 (chemotype I) and V19 (chemotype III), as expected, are the genotypes in which the maximum cannabinoid accumulation is observed. For all genotypes except Ermo, the highest cannabinoid content was observed in mature flowers, determined mainly by the accumulation of CBD or THC, depending on the chemotype.

Excluding the inflorescences, known as the preferred site for phytocannabinoid accumulation, the following sections will discuss in more detail the quantification of cannabinoids in the early vegetative stages of plant growth (from BBCH stage 00 to BBCH stage 13).

In V9, the early vegetative stage with the highest cannabinoid accumulation is BBCH-13 (three-leaf pairs), where a total of 2.16 % w/dw is observed. In V19, on the other hand, a strong cannabinoid accumulation is reached already in BBCH stage 12 (two-leaf pair) and remains quite stable in the following stage (2.17 w/dw %).

For samples belonging to these two genotypes, an increase in cannabinoid content is observed in association with the plant stage, with an increasing tendency during growth.

In V9 genotype, CBC was the only cannabinoid detected at BBCH stage 9; in V19, CBC was detected at stage 11 together with other phytocannabinoids, accounting for 90% of the total cannabinoids; CBC reached the maximum amount at stage 12, with 1.7 % w/dw. In the following stages, it begins to decline both in V9 and V19 samples, while the other phytocannabinoids increased. The highest CBG content was detected in the BBCH- 13 stage of V9, similar to that of THC, whereas in V19 its content is always lower than CBD.

Similarly, the highest THC content was detected in V9, and increased with seedling age. However, it can be observed in V9 stage 11 and 12 that CBG, CBD and THC accumulated in comparable proportions in the early stages of plant development, regardless of the plant chemotype; indeed, only from stage 13 on the THC content becomes prevalent over the others.

The genotype accumulating the highest amount of CBD is V19, where a preferential accumulation of CBD over CBG and THC is evident, although this latter cannabinoid is detected at concentrations near 0.10 % w/dw. Compared to V9 and V19 the two industrial hemp genotypes tested had significantly lower cannabinoid content in all the stages considered, not exceeding 6% w/dw in Fibrante (chemotype III) and remaining below 0.05 % w/dw in Ermo (chemotype V) in mature inflorescences. In Fibrante, the juvenile stage with the highest total content is the BBCH-13, while in Ermo the stages with the highest quantity of cannabinoids are the BBCH-11 and BBCH-12. Very low levels of metabolites were found in type V samples, in which a THC level above 0.3% was never detectable. In Fibrante stage 9, only CBC was detected, resulting in the 100% of cannabinoids, while in Ermo, CBC is quantified only in stage 11, when other cannabinoids are also present, which is why it accounts for about 60% of the total.

In seedling stages of Fibrante an increasing trend of accumulation was observed for all phytocannabinoids.

3.2 Transcription of biosynthetic genes

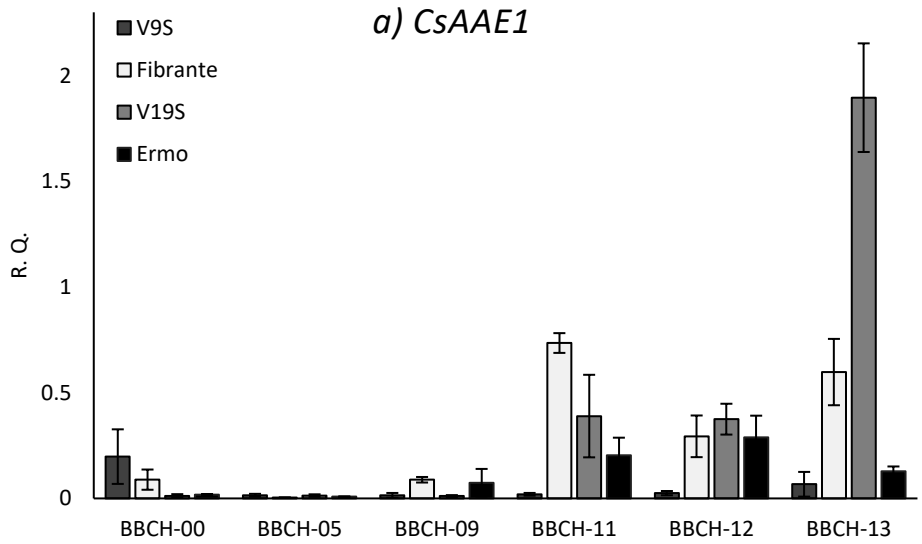
To correlate the amount of the different cannabinoids during the early stages of plant development with the time course of transcription of the different genes involved in their biosynthesis, we analysed the transcription of the genes upstream CBGA synthesis and of the main cannabinoid synthases. Transcription levels were monitored from the dry seed to the three leaf pairs stage (BBCH-13), in the four genotypes considered (Fig. 3).

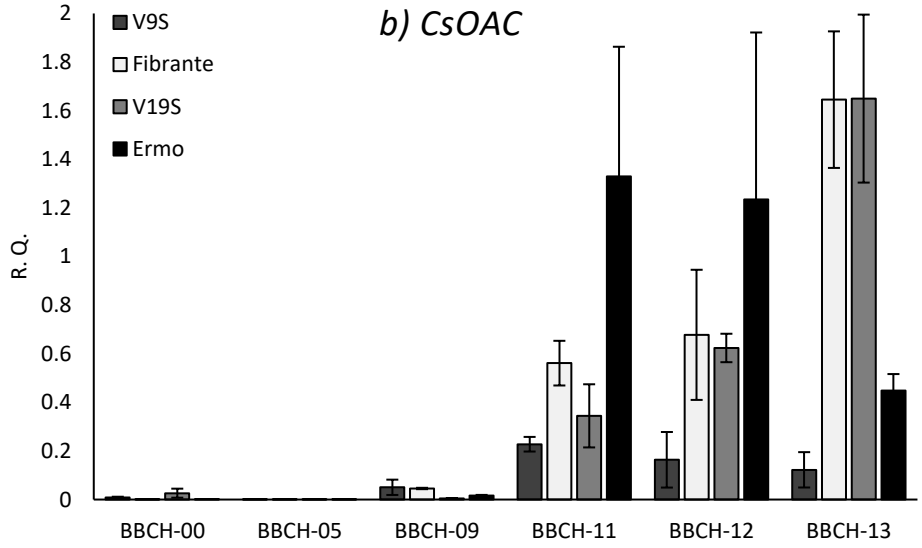
In general, transcription levels of all genes coding for enzymes involved in steps upstream of the CBGA formation were exceedingly low until the BBCH-11 (one leaf) stage was reached. After this stage, a strong increase was usually observed, dependent on the genotype (Fig. 2). However, it was also observed that mRNA levels for the genes *AAE1* and *OLS* were low but significant in the dry seed, decreasing in the subsequent stages until BBCH-11, when a general recovery of transcription was evident in all genotypes.

Acyl-activating enzyme 1 (*AAE1*) encodes for the last enzyme of the hexanoate pathway, that synthesizes the hexanoyl-CoA. This gene is transcribed in almost all growth stages analysed, including the dormant seed, at very low levels until the stage of one true leaf (BBCH-11) and in all genotypes, with a preferential expression

in the green leaves of CBD-prevalent seedlings. Its transcription increases from stage BBCH-11 as the plantlet grows.

The *olivetol synthase OLS* or *tetraketide synthase TKS* gene encodes a polyketide synthase catalysing the synthesis of olivetolic acid in concert with Olivetolic acid cyclase (see below). This gene was transcribed at all growth stages in all genotypes, starting from aged seeds to the stage BBCH-13. In general, also for this gene, there is a burst of transcription starting from BBCH-11, but this sharp increase is not the same for all genotypes, being especially marked for V19 and Fibrante (chemotypes II and III respectively). However, for most genotypes, the stage with the highest transcription was the stage BBCH-11, which is also the stage from which a significant amount of total cannabinoids is accumulated.


a) CsAAE1
V9S
Fibrante
V19S
Ermo
R. Q.
2
1.5
1
0.5
0
BBCH-00
BBCH-05
BBCH-09
BBCH-11
BBCH-12
BBCH-13


b) CsOAC
V9S
Fibrante
V19S
Ermo
R. Q.
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
BBCH-00
BBCH-05
BBCH-09
BBCH-11
BBCH-12
BBCH-13

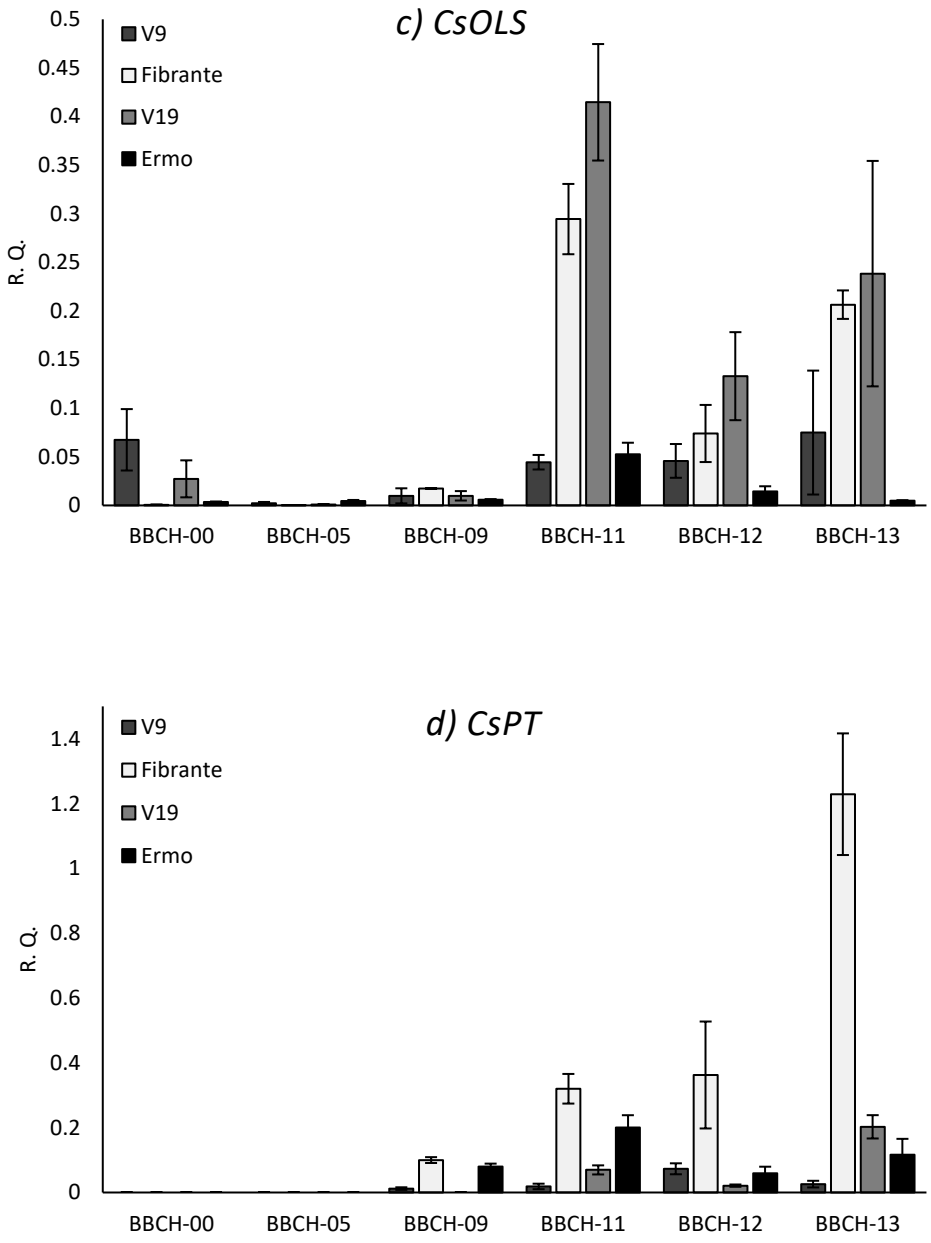


Figure 2. Relative Quantitation (R.Q.) of CsAAE1 (a), CsOAC (b), CsOLS (c) and CsPT (d) transcriptional levels in BBCH stages from 0 to 13. Y axis reports R.Q. expressed in Arbitrary Units (A.U.). Bars represent the standard errors of the mean of three biological replicates (n = 3).

As for the *OAC* gene, which encodes for an enzyme that cyclizes the substrate resulting from the condensation of hexanoyl-CoA with 3 malonyl-CoA units, it was found more transcribed in the latest stages analysed. In this case, however, the peak of expression is shifted toward stage BBCH-13 in chemotype III genotypes (V19 and Fibrante), tissues in which the total cannabinoid content varies between approximately 1.2 and 2.2 % w/dw. Low levels of transcription are also observed in the early stages, of long-stored seeds. In Ermo a burst of transcription of OAC gene is observed at the first true leaves stage (BBCH-11), with a subsequent decline visible at stage 13 (Fig. 2).

In most genotypes, the *PT1* gene, encoding for an aromatic prenyltransferase which inserts a prenyl group into the olivetolic acid, is transcribed, though at low levels, from the stage 09. However, no expression was detected for this gene in stage 09 of V19. In Fibrante, a general trend of increasing expression during plant growth can be observed, whereas for the other genotypes transcription of *PT1* remains almost constant throughout the different developmental stages. No transcription of *PT1* was detected in seeds.

The *PT* transcription is related to the accumulation of CBG in all the different samples; in fact, increased transcription usually corresponds to increased accumulation of this metabolite.

In *C. sativa*, CBGA is cyclized by three homologous monooxygenases, THCA, CBDA and CBCA synthases, encoded by single, specific genes which were also analysed in this work. The results of their transcription are reported in comparison with the accumulation of the respective cannabinoid.

The transcriptional level of *THCA synthase* was detected and quantified only in V9 (chemotype I) and followed a trend strongly correlated to THC content accumulated at the different stages, starting at BBCH-09 and increasing during growth up to the BBCH-13 stage (Fig.3). No amplification of *THCAS* was detectable in chemotypes III and V samples analysed (V19, Fibrante and Ermo).

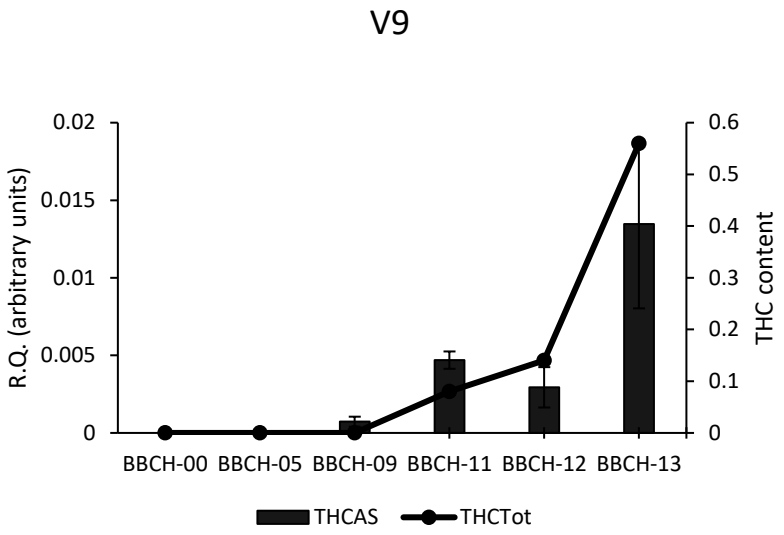
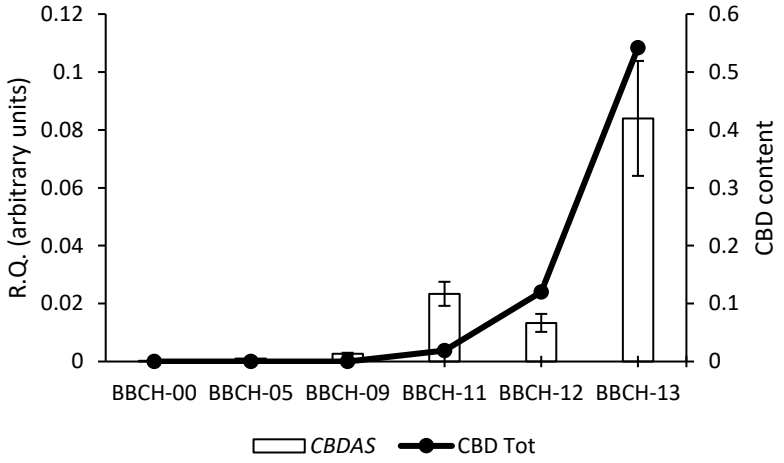


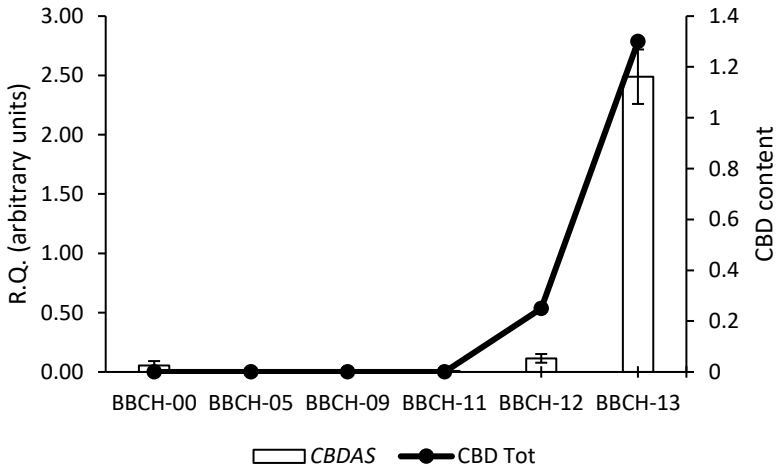
Figure 3. Relative Quantitation (R.Q.) of CsTHCAS transcriptional levels in V9 in BBCH stages from 0 to 13. Y axis reports R.Q. expressed in Arbitrary Units (A.U.). Bars represent the standard errors of the mean of three biological replicates (n = 3). The different graphs show the superimposition of THC content of the genotype in the analysed samples.

The transcriptional levels of *CBDA Synthase* (*CBDAS*) were quantifiable only in V19, Fibrante and Ermo, while no transcript was detected for V9 genotype samples (Fig. 4). Again, *CBDAS* transcription levels correlate very well with the amount of CBD accumulated at the different stages; interestingly, in Ermo (chemotype V, i.e. no cannabinoid accumulation), transcription levels of *CBDAS* were perfectly similar to those of the other two genotypes (Fig. 4).

a) Fibrante



b) V19



c) Ermo

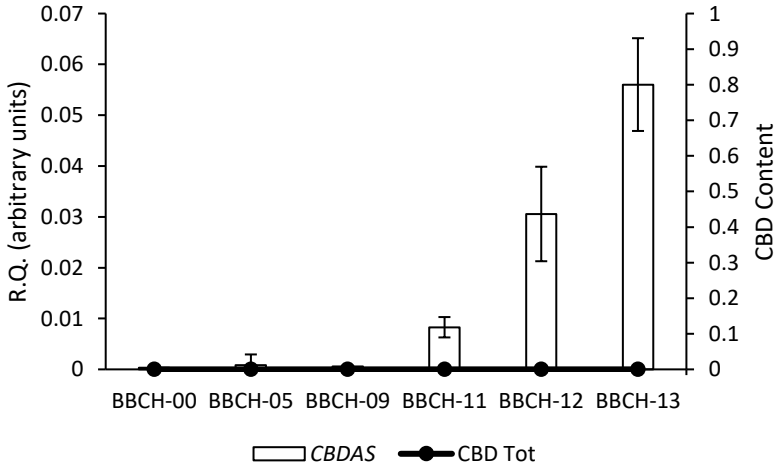
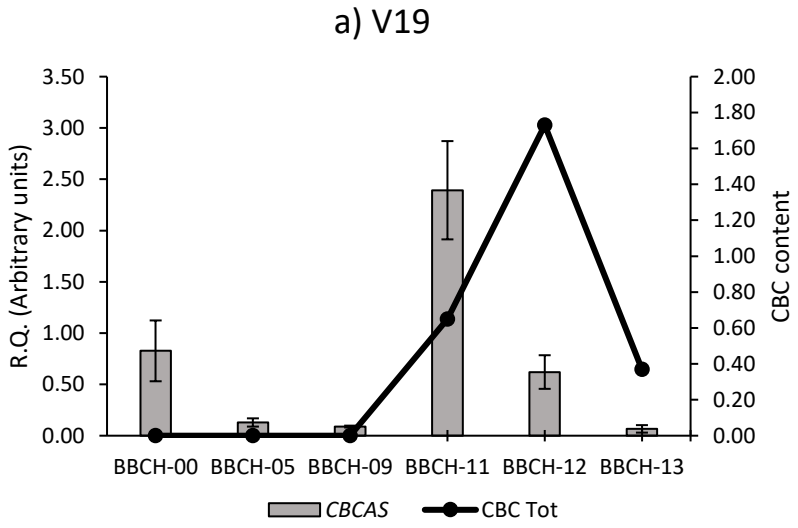


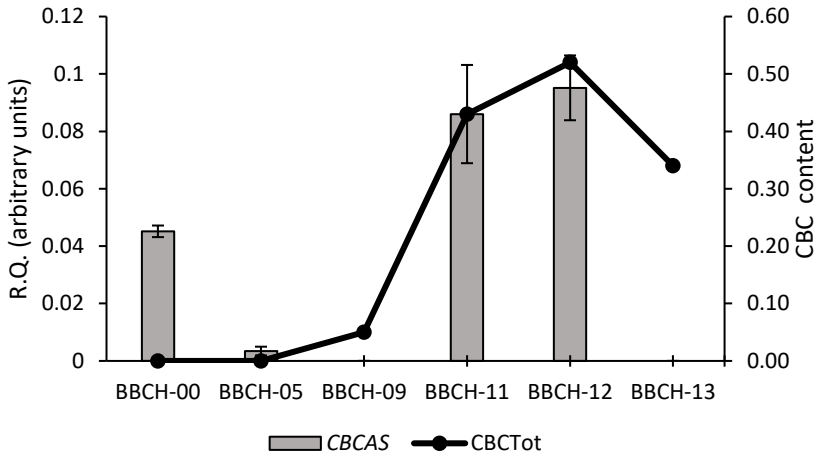
Figure 4. Relative Quantitation (R.Q.) of CsCBDAS transcriptional levels in Fibrante (a), V19 (b) and Ermo (c) in BBCH stages from 0 to 13. Y axis reports R.Q. expressed in Arbitrary Units (A.U.). Bars represent the standard errors of the mean of three biological replicates (n = 3). The different graphs show the superimposition of CBD content in each genotype in the analysed samples.

Albeit at very low levels, the *CBDA* Synthase is transcribed in seeds, and has an increasing trend during vegetative growth, reaching its peak at the last stage analysed, BBCH-13.

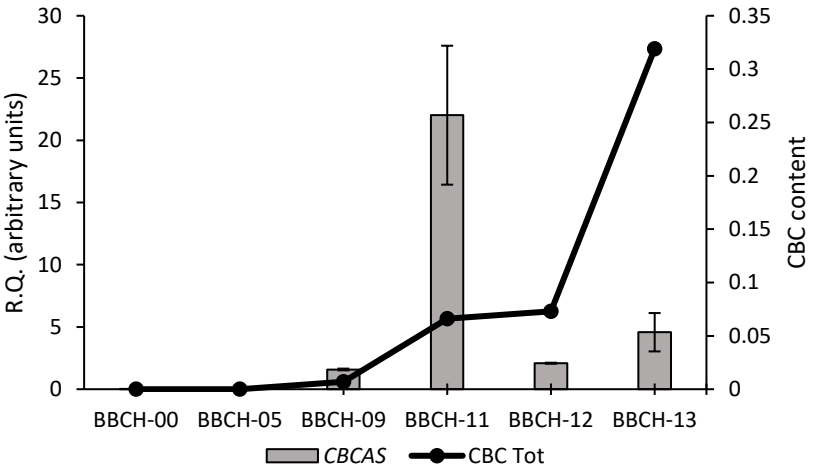
The transcription levels of the CBCA synthase (*CBCAS*) gene were also investigated, and they were detected and quantified in all four genotypes (Fig. 5).



b) V9



c) Fibrante



d) Ermo

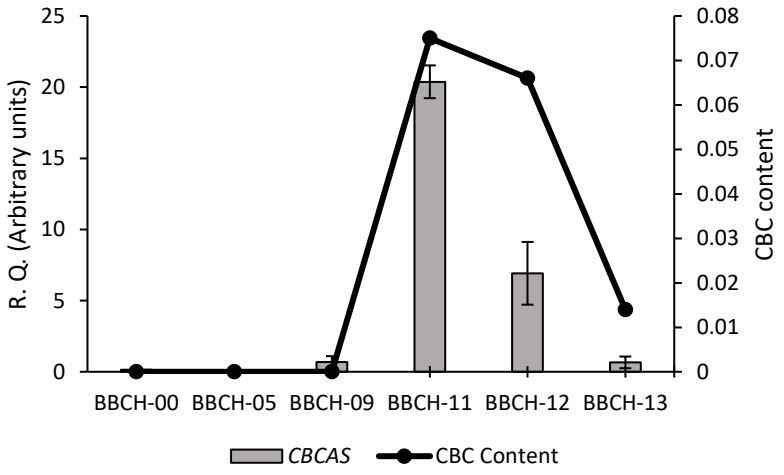


Figure 5. Relative Quantitation (R.Q.) of CsCBCAS transcriptional levels in V19 (a), V9 (b), Fibrante (c) and Ermo (d) in BBCH stages from 0 to 13. Y axis reports R.Q. expressed in Arbitrary Units (A.U.). Bars represent the standard errors of the mean of three biological replicates (n = 3). The different graphs show the superimposition of CBC content in each genotype in the analysed samples.

The transcription of this gene is detected in the seeds of all genotypes studied, especially in V9 and V19, then in the early stages of germination it undergoes a considerable decline, but it resumes and reaches its highest level between stage 11 and stage 12, to decline again at BBCH-13 in all genotypes, except Fibrante.

Whereas in V9 and Ermo transcript accumulation and CBC accumulation show a good correlation, for the two chemotype III genotypes (Fibrante and V19), metabolite accumulation peaks temporally after the highest level of *CBCAS* transcription.

Therefore, it is evident from our data, that the major transcription of *CBCAS* is concentrated in the newly emerged seedling with one or two pairs of true leaves, and that the gene responsible for CBCA synthesis is already transcribed in the dry seed.

4. Discussion

In this study, focus was placed on the early stages of C. sativa seedling development, with the aim to investigate the onset of phytocannabinoids accumulation in different chemical phenotypes (I, III and V); besides, if and how much this accumulation trend is matched by transcription of the most relevant genes responsible for their biosynthesis was investigated. Six developmental stages were therefore chosen, starting from the dry seed (stage BCCH-00), known to be devoid of phytocannabinoids, to the plantlets with three pairs of leaves (BCCH-13), when the biosynthesis is supposed to be already activated.

The results shown offer the possibility of associating the content of different cannabinoids with the expression of genes involved in the relevant metabolism, with a particular focus on the specific *CBCAS* expression and CBC accumulation in juvenile tissues of different *C. sativa* chemotypes.

Only a few studies to date have focused on the synthesis of CBC; these studies clearly demonstrated an association between its accumulation with juvenile stages of the plant and showed that its content decreased with increasing plant age and size (Morimoto et al., 1997; 1998; de Meijer et al., 2009). This stage-dependent association is so strong in the case of CBC that de Meijer et al. (2009) were able to breed a line with very high CBC proportion by associating it to a "prolonged juvenile chemotype" (PJC) trait identified in a *C. sativa* accession. Normally, the accumulation of this compound in mature inflorescences, which are usually the primary site of maximal cannabinoid deposition, does not exceed 2 % w/dw (Gul et al., 2018; Stack et al., 2021).

Several studies suggested that the synthesis of CBCA in inflorescences results from the promiscuous activity of CBDAS, since the CBD:CBC ratios observed *in vivo*, are comparable with those obtained *in vitro* from the heterologous expression of CBDAS (Stack et al., 2021; Toth et al., 2020; Zirpel et al., 2018).

Based on the results reported in this work, the transcription of the *CBDAS* is strongly correlated to the CBDA production and is consecutive to CBC accumulation; while *CBCAS* showed a transcriptional profile which is independent and in agreement with the CBC accumulation, suggesting it is responsible for the synthesis of this phytocannabinoid. Localization of *CBCAS* in sessile trichomes (de Meijer et al., 2009), which are abundant in young leaves, could explain the preferential accumulation of this cannabinoid in juvenile tissues, and

consequently also its low accumulation levels in inflorescences, that are rich mostly in capitate-stalked trichomes.

It has been reported that CBCAS has a lower Vmax and kcat and therefore a lower catalytic capacity compared to CBDAS and THCAS, but a higher affinity for CBGA (Morimoto et al., 1998). These features may determine the higher accumulation of CBCA in the early stages of seedling life, when the transcription of *CBCAS* is higher than that of other cannabinoid synthases. As soon as *THCAS* or *CBDAS* begin to be transcribed and likely translated at increasingly high levels during development, their superior catalytic capacity causes the CBGA conversion to CBDA and THCA to become overwhelming compared to the comparatively modest catalytic activity of CBCAS, and as a result, CBCA concentrations are dramatically reduced in the adult plant tissues. Further, the already mentioned shift from the formation of sessile to stalked trichomes is also likely to contribute to the switch from prevalent CBCA to prevalent THCA or CBDA synthesis. Additionally, this could result in metabolic specialization, as also observed in trichomes of other species (Livingston et al., 2020).

A higher concentration of CBC during the juvenile stages has previously been associated with a final high THC amount (Gul et al., 2018); however, our results do not confirm this view, as the genotype with the highest accumulation of CBC is V19, with a high CBDA content at maturity.

In the present study, the THCA yield in extremely young tissues reaches the 0.56% w/dw at the BBCH-13 stage of V9 (chemotype I), but surprisingly at the same stage, the CBD content is higher than THC (0.69% w/dw). Chemotype I varieties usually lack the functional gene responsible for the synthesis of CBDA, as also verified by the genotyping of this accession, that, based on the B1080/B1192 molecular marker test ((Pacifico et al., 2006) data not shown) was positive exclusively for the presence of *THCAS*-associated fragment. On the other hand, it can be observed that in stage 11 of Ermo, a variety with very low or null cannabinoid content, THC is still present, despite lacking the gene responsible for its synthesis (Fulvio et al, 2021). These findings show that the chemotype is not conserved during plant growth, at least during the very early stages of plant development that, indeed, can be used to predict the chemotype of the plant only by DNA-based molecular makers and not by chemical analysis of cannabinoids.

Previous reports showed that both CBDAS and THCAS can produce to some limited extent non-specific cannabinoids as side products. Biochemical analyses (Thomas and Kayser, 2022; Zirpel et al., 2018) indicate

that a possible synthesis of side products by CBCAS activity is on the contrary less likely to be responsible for the presence of residual cannabinoids, and our data agree with this view. In *C. sativa* plants, this implies that genotypes with only CBDAS or only THCAS will also produce some THCA or CBDA as "extra" cannabinoids, respectively, even if no functional copies of the respective genes are present in their genome, since it appears as an intrinsic feature of both enzymes.

Our findings show that *THCAS* transcription and THCA synthesis are simultaneous in young tissues, in contrast to what was reported by Apicella et al. (2022), where a negative correlation between gene transcription and accumulation of the corresponding metabolite is reported (Apicella et al., 2022). The same is here observed for the *CBDAS* and partially for *CBCAS* genes, suggesting that the synthesis of the relative compounds is timely controlled by the transcription of synthases.

A positive correlation between cannabinoid gene expression and accumulation was also shown by Deguchi et al. (2022), who observed that overexpression of *CBDAS* is associated with a significant increase in CBDA in transformed plants (Deguchi et al.,2022). Also, Matchett-Oates et al. (2021), reported that *THCAS* silencing results in a reduction in THCA content, further evidence supporting the idea that *THCAS* and *CBDAS* are the limiting steps for the synthesis of CBDA and THCA, respectively (Matchett-Oates et al., 2021). In addition, a positive correlation between the *THCAS* gene transcription and the amount of THCA accumulated in female hemp inflorescences has recently been reported (Fulvio et al., 2023).

Besides cannabinoid synthase genes, a correlation is also observed between transcription of the *PT1* gene and accumulation of CBGA, the precursor cannabinoid. The results show that in V9, Fibrante and Ermo *PT1* is transcribed from stage 09, the same at which the first traces of cannabinoids are observed in the first two genotypes. In V19 the transcription of the gene appears later, starting at stage 11, in which also the presence of CBC and CBG is observed. The transcription of this gene could therefore represent the rate-limiting step for phytocannabinoid biosynthesis, as already suggested (Apicella et al.,2022). In fact, although the other genes involved in the pathway are already transcriptionally active, no phytocannabinoids are observed until the *PT* gene is transcribed.

Another interesting finding is that different mRNAs carrying information for the synthesis of cannabinoids are accumulated and conserved in the seed stored for long time. This occurs not only in the case of transcripts corresponding to genes upstream the synthesis of the first cannabinoid (*AAE1*, *PKS*, *OAC*), but also for the specific oxydocyclases like *CBCAS* and *CBDAS*. This result was particularly surprising as the seeds used for the experiments were derived from a germplasm collection stored at -4°C, for widely varying periods, from 3 (Fibrante, V19) to 8 years (V9).

Our data suggest that mRNAs bringing information for cannabinoid synthesis are stored during the final stage of seed development to be readily translated after imbibition, or during the first developmental stages of seedling growth. This truly remarks the apparent relevance of cannabinoid metabolism for the plant, and is not surprising, since about 3 days after seedling emergence, up to about 2% w/dw of CBC is accumulated in the young leaves.

Several studies highlighted how mRNAs stored in the dry seed are translated during germination (Bai et al., 2020, 2017) and the identification of transcribed genes involved in cannabinoid biosynthesis in long-time stored dry seeds offers further insights into the physiology of the *C. sativa* plant and the biological significance of phytocannabinoids.

5. Conclusions

A major challenge in hemp research lies in the still scarce and fragmented knowledge of the gene regulating cannabinoid metabolism. Our study strongly suggests that CBCA is formed by the action of the CBCAS, encoded by a gene which expression is concurrent to the accumulation of the metabolite in the juvenile tissues of *C. sativa* genotypes belonging to 3 different chemotypes. The transcription of *PT1* is proposed as the rate-limiting step of cannabinoid biosynthesis. In addition, it has been observed that transcription of cannabinoid synthase genes begins in the seed, before the germination process begins, attesting the relevance of this metabolism for the plant.

6. Funding

This research was funded by: UNIHEMP research project “Use of iNdustrIal Hemp biomass for Energy and new biocheMicals Production” (ARS01_00668) funded by European Regional Development Fund (within the PON R&I 2017-2020–Axis 2—Action II—OS 1.b). Grant decree UNIHEMP prot. no. 2016 of 27/07/2018; CUP B76C18000520005 and by CAMED - National medical cannabis: innovation and enhancement of the production of therapeutic cannabis plant material for national needs and research for the establishment of new varieties for pharmaceutical use. Collaboration agreement ex art. 15 of Law 241/90), MIPAAF-2021-0181330. (17/06/2021 - 16/06/2023).

7. References

Anderson, L.L., Ametovski, A., Lin Luo, J., Everett-Morgan, D., McGregor, I.S., Banister, S.D., Arnold, J.C., 2021. Cannabichromene, Related Phytocannabinoids, and 5-Fluoro-cannabichromene Have Anticonvulsant Properties in a Mouse Model of Dravet Syndrome. ACS Chem Neurosci 12, 330–339. https://doi.org/10.1021/acschemneuro.0c00677

Apicella, P. v., Sands, L.B., Ma, Y., Berkowitz, G.A., 2022. Delineating genetic regulation of cannabinoid biosynthesis during female flower development in Cannabis sativa. Plant Direct 6. https://doi.org/10.1002/pld3.412

Backer, R., Mandolino, G., Wilkins, O., ElSohly, M.A., Smith, D.L., 2020. Editorial: Cannabis Genomics, Breeding and Production. Front Plant Sci. https://doi.org/10.3389/fpls.2020.591445

Bai, B., Peviani, A., van der Horst, S., Gamm, M., Snel, B., Bentsink, L., Hanson, J., 2017. Extensive translational regulation during seed germination revealed by polysomal profiling. New Phytologist 214, 233–244. https://doi.org/10.1111/nph.14355

Bai, B., van der Horst, S., Cordewener, J.H.G., America, T.A.H.P., Hanson, J., Bentsink, L., 2020. Seed-stored mRNAs that are specifically associated to monosomes are translationally regulated during germination. Plant Physiol 182, 378–392. https://doi.org/10.1104/pp.19.00644

Citti, C., Ciccarella, G., Braghiroli, D., Parenti, C., Vandelli, M.A., Cannazza, G., 2016. Medicinal cannabis: Principal cannabinoids concentration and their stability evaluated by a high performance liquid chromatography coupled to diode array and quadrupole time of flight mass spectrometry method. J Pharm Biomed Anal 128, 201–209. https://doi.org/10.1016/j.jpba.2016.05.033

de Meijer, E.P.M., Hammond, K.M., Micheler, M., 2009. The inheritance of chemical phenotype in Cannabis sativa L. (III): Variation in cannabichromene proportion. Euphytica 165, 293–311. https://doi.org/10.1007/s10681-008-9787-1

Deguchi, M., Dhir, Seema, Potlakayala, S., Dhir, Sarwan, Curtis, W.R., Rudrabhatla, S., 2022. In planta Female Flower Agroinfiltration Alters the Cannabinoid Composition in Industrial Hemp (Cannabis sativa L.). Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.921970

Etchart, M.G., Anderson, L.L., Ametovski, A., Jones, P.M., George, A.M., Banister, S.D., Arnold, J.C., 2022. In vitro evaluation of the interaction of the cannabis constituents cannabichromene and cannabichromenic acid with ABCG2 and ABCB1 transporters. Eur J Pharmacol 922, 174836. https://doi.org/10.1016/J.EJPHAR.2022.174836

Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., Moschella, A., Alberti, I., Pecchioni, N., Cannazza, G., Mandolino, G., 2021. Analysis of sequence variability and transcriptional profile of cannabinoid synthase genes in cannabis sativa l. Chemotypes with a focus on cannabichromenic acid synthase. Plants 10. https://doi.org/10.3390/plants10091857

Fulvio, F., Righetti, L., Minervini, M., Moschella, A., Paris, R., 2023. The B1080/B1192 molecular marker identifies hemp plants with functional THCA synthase and total THC content above legal limit. Gene 858. https://doi.org/10.1016/j.gene.2023.147198

Galletta, M., Reekie, T.A., Nagalingam, G., Bottomley, A.L., Harry, E.J., Kassiou, M., Triccas, J.A., 2020. Rapid antibacterial activity of cannabichromenic acid against methicillin-resistant staphylococcus aureus. Antibiotics 9, 1–11. https://doi.org/10.3390/antibiotics9080523

Gul, W., Gul, S.W., Chandra, S., Lata, H., Ibrahim, E.A., Elsohly, M.A., 2018. Detection and Quantification of Cannabinoids in Extracts of Cannabis sativa Roots Using LC-MS/MS. Planta Med 84, 267–271. https://doi.org/10.1055/s-0044-100798

Hanuš, L.O., Meyer, S.M., Muñoz, E., Taglialatela-Scafati, O., Appendino, G., 2016. Phytocannabinoids: A unified critical inventory. Nat Prod Rep. https://doi.org/10.1039/c6np00074f

Izzo, A.A., Capasso, R., Aviello, G., Borrelli, F., Romano, B., Piscitelli, F., Gallo, L., Capasso, F., Orlando, P., di Marzo, V., 2012. Inhibitory effect of cannabichromene, a major non-psychotropic cannabinoid extracted

from Cannabis sativa, on inflammation-induced hypermotility in mice. Br J Pharmacol 166, 1444–1460. https://doi.org/10.1111/j.1476-5381.2012.01879.x

Kearsey, L.J., Prandi, N., Karuppiah, V., Yan, C., Leys, D., Toogood, H., Takano, E., Scrutton, N.S., 2020. Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases. FEBS Journal 287, 1511–1524. https://doi.org/10.1111/febs.15089

Kumar, P., Mahato, D.K., Kamle, M., Borah, R., Sharma, B., Pandhi, S., Tripathi, V., Yadav, H.S., Devi, S., Patil, U., Xiao, J., Mishra, A.K., 2021. Pharmacological properties, therapeutic potential, and legal status of Cannabis sativa L.: An overview. Phytotherapy Research. https://doi.org/10.1002/ptr.7213

Larionov, A., Krause, A., Miller, W.R., 2005. A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 6. https://doi.org/10.1186/1471-2105-6-62

Linciano, P., Citti, C., Luongo, L., Belardo, C., Maione, S., Vandelli, M.A., Forni, F., Gigli, G., Laganà, A., Montone, C.M., Cannazza, G., 2020. Isolation of a High-Affinity Cannabinoid for the Human CB1 Receptor from a Medicinal Cannabis sativa Variety: Δ9-Tetrahydrocannabutol, the Butyl Homologue of Δ9-Tetrahydrocannabinol. J Nat Prod 83, 88–98. https://doi.org/10.1021/acs.jnatprod.9b00876

Linciano P., Russo f., Citti C., Tolomeo f., Paris R., Fulvio F., Pecchioni N., Vandelli M.A., Laganà A., Capriotti A.L., Biagini G., Carbone L., Gigli G., Cannazza G. 2021. The novel heptyl phorolic acid cannabinoids content in different Cannabis sativa L. accessions,Talanta 235,122704. https://doi.org/10.1016/j.talanta.2021.122704.

Livingston, S.J., Quilichini, T.D., Booth, J.K., Wong, D.C.J., Rensing, K.H., Laflamme-Yonkman, J., Castellarin, S.D., Bohlmann, J., Page, J.E., Samuels, A.L., 2020. Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. Plant Journal 101, 37–56. https://doi.org/10.1111/tpj.14516

Maione, S., Piscitelli, F., Gatta, L., Vita, D., de Petrocellis, L., Palazzo, E., de Novellis, V., di Marzo, V., 2011. Non-psychoactive cannabinoids modulate the descending pathway of antinociception in anaesthetized rats

through several mechanisms of action. Br J Pharmacol 162, 584–596. https://doi.org/10.1111/j.1476-5381.2010.01063.x

Mangeot-Peter, L., Legay, S., Hausman, J.F., Esposito, S., Guerriero, G., 2016. Identification of reference genes for RT-qPCR data normalization in Cannabis sativa stem tissues. Int J Mol Sci 17. https://doi.org/10.3390/ijms17091556

Matchett-Oates, L., Spangenberg, G.C., Cogan, N.O.I., 2021. Manipulation of Cannabinoid Biosynthesis via Transient RNAi Expression. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.773474

Mishchenko, S., Mokher, J., Laiko, I., Burbulis, N., Kyrychenko, H., Dudukova, S., 2017. Phenological growth stages of hemp (Cannabis sativa L.): codification and description according to the BBCH scale. Žemės ūkio mokslai 24. https://doi.org/10.6001/zemesukiomokslai.v24i2.3496

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., 1997. Enzymological Evidence for Cannabichromenic Acid Biosynthesis.

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., n.d. PURIFICATION AND CHARACTERIZATION OF CANNABICHROMENIC ACID SYNTHASE FROM CANNABIS SATIVA.

Oláh, A., Markovics, A., Szabó-Papp, J., Szabó, P.T., Stott, C., Zouboulis, C.C., Bíró, T., 2016. Differential effectiveness of selected non-psychotropic phytocannabinoids on human sebocyte functions implicates their introduction in dry/seborrhoeic skin and acne treatment. Exp Dermatol 25, 701–707. https://doi.org/10.1111/exd.13042

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., Mandolino, G., 2006. Genetics and marker-assisted selection of the chemotype in Cannabis sativa L. Molecular Breeding 17, 257–268. https://doi.org/10.1007/s11032-005-5681-x

Piscitelli, F., Marzo, V. di, n.d. The ever-expanding world of the endocannabinoids: A concise introduction THE ENDOCANNABINOID SYSTEM: DISCOVERY AND EARLY DEFINITIONS.

Romano, B., Borrelli, F., Fasolino, I., Capasso, R., Piscitelli, F., Cascio, M.G., Pertwee, R.G., Coppola, D., Vassallo, L., Orlando, P., di Marzo, V., Izzo, A.A., 2013. The cannabinoid TRPA1 agonist cannabichromene inhibits nitric oxide production in macrophages and ameliorates murine colitis. Br J Pharmacol 169, 213–229. https://doi.org/10.1111/bph.12120

Shinjyo, N., di Marzo, V., 2013. The effect of cannabichromene on adult neural stem/progenitor cells. Neurochem Int 63, 432–437. https://doi.org/10.1016/j.neuint.2013.08.002

Small, E. 1975. Morphological variation of achenes of *Cannabis*. *Canadian Journal of Botany*, *53*(10), 978–987. https://doi.org/10.1139/b75-117

Stack, G.M., Toth, J.A., Carlson, C.H., Cala, A.R., Marrero-González, M.I., Wilk, R.L., Gentner, D.R., Crawford, J.L., Philippe, G., Rose, J.K.C., Viands, D.R., Smart, C.D., Smart, L.B., 2021. Season-long characterization of high-cannabinoid hemp (Cannabis sativa L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. GCB Bioenergy 13, 546–561. https://doi.org/10.1111/gcbb.12793

Tahir, M.N., Shahbazi, F., Rondeau-Gagné, S., Trant, J.F., 2021. The biosynthesis of the cannabinoids. J Cannabis Res 3. https://doi.org/10.1186/s42238-021-00062-4

Thomas, F., Kayser, O., 2022. Natural deep eutectic solvents enhance cannabinoid biotransformation. Biochem Eng J 180. https://doi.org/10.1016/j.bej.2022.108380

Tolomeo, F.; Russo, F.; Kaczorova, D.; Vandelli, M.A.; Biagini, G.; Laganà, A.; Capriotti, A.L.; Paris, R.; Fulvio, F.; Carbone, L.; et al., 2022. *Cis* Δ^9-tetrahydrocannabinolic acid occurrence in Cannabis sativa L. *J. Pharm. Biomed. Anal.*, *219*, 114958. https://doi.org/10.3390/molecules28031164

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K.C., Smart, L.B., 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. GCB Bioenergy 12, 213–222. https://doi.org/10.1111/gcbb.12667

Walsh, K.B., McKinney, A.E., Holmes, A.E., 2021. Minor Cannabinoids: Biosynthesis, Molecular Pharmacology and Potential Therapeutic Uses. Front Pharmacol. https://doi.org/10.3389/fphar.2021.777804

Zirpel, B., Kayser, O., Stehle, F., 2018. Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L. J Biotechnol 284, 17–26. https://doi.org/10.1016/j.jbiotec.2018.07.031

Chapter 3

The B1080/B1192 molecular marker identifies hemp plants with functional *THCA synthase* and total THC content above legal limit

This chapter has been extracted from "The B1080/B1192 molecular marker identifies hemp plants with functional *THCA synthase* and total THC content above legal limit" from Flavia Fulvio[1,2*], Laura Righetti[1*], Marco Minervini[1], Anna Moschella[1], and Roberta Paris[1]

1 CREA, Research Centre for Cereal and Industrial Crops, Bologna, Italy

2 Department of Sciences of Agriculture, Food Natural Resources and Engineering, University of Foggia, Foggia, Italy

*The first two authors equally contributed to the work

Published in Gene, https://doi.org/10.1016/j.gene.2023.147198

Abstract: In *Cannabis sativa* L. the presence of delta 9-tetrahydrocannabinolic acid (THCA) above legal limit is a challenging issue that still restricts the industrial exploitation of this promising crop. In recent years, the interest of entrepreneurs and growers who see hemp as a dynamic and profitable crop was joined by the growing knowledge on *C. sativa* genetics and genomics, accelerated by the application of high throughput tools. Despite the renewed interest in the species, much remains to be clarified, especially about the long-standing problem of THCA in hemp inflorescences, which could even result in the seizure of the whole harvest. Although several hypotheses have been formulated on the accumulation of this metabolite in industrial varieties, none is conclusive yet. In this work, individuals of a population of the hemp cultivar 'FINOLA' obtained from commercial seeds were investigated for total THC level and examined at molecular level. A marker linked to *THCA synthase* was found at a high incidence in both male and female plants, suggesting a considerable genetic variability within the seed batch. Full-length sequences encoding for putatively functional *THCA synthases* were isolated for the first time from the genome of both female and male plants of an industrial hemp variety and, using transcriptional analysis, the *THCA synthase* expression was quantified in mature inflorescences of individuals identified by the marker. Biochemical analyses finally demonstrated for these plants a 100% association between the predicted and actual chemotype.

1. Introduction

Cannabis sativa L. is a complex and polymorphic plant species, which produces a vast array of chemical compounds, making it attractive for many industrial applications. Cannabinoids are the major bioactive chemicals in *C. sativa*, with almost 150 different molecules described so far (Cerrato et al. 2021). Most common cannabinoids are produced by the condensation of olivetolic acid with geranylgeraniol-diphosphate, resulting in cannabigerolic acid (CBGA), from which cannabidiolic acid (CBDA), cannabichromenic acid (CBCA) and tetrahydrocannabinolic acid (THCA) are synthesized by three different enzymes named CBCA-, CBDA- and THCA- synthase (CBCAS, CBDAS and THCAS), respectively (Gülck and Møller 2020).

Based on both CBDA/THCA ratio and amount of the main cannabinoids, *C. sativa* plants can be classified in five chemical phenotypes (chemotypes): THCA is prevalent in chemotype I, chemotype II has both CBDA and THCA at similar concentrations, III is CBDA-dominant, IV is CBGA-dominant and V has only traces of cannabinoids (Mandolino and Carboni 2004). Chemotype I, II and III can be easily and rapidly distinguished by using molecular markers soon after seed germination. Among the markers capable to distinguish these chemotypes (Toth et al., 2020 and Wenger et al., 2020), the B1080/B1192 one developed by Pacifico et al. (2006) and recently detailed in Fulvio et al. (2021) is currently the easiest and least expensive to use for laboratories with basic equipment.

Despite years of breeding aimed at eliminating THCA, many industrial hemp varieties still accumulate traces of this cannabinoid, sometimes exceeding the current limit set by EU of 0.20% w/w (increased up to 0.30% from 2023) in the dry mature inflorescences. There are several possible origins for the accumulation of traces of THCA in industrial hemp. Different studies suggest that THCA is an unavoidable by-product of CBDAS activity (Zirpel et al. 2018). Likewise, CBCAS could be responsible for the low levels of THCA accumulation (Fulvio et al. 2021). The presence of *THCAS*-related molecular markers has been previously reported in certified hemp cultivars (Pacifico et al. 2008; Staginnus et al. 2014), and a more recent survey conducted by Borroto Fernandez et al. (2020) revealed the presence of up to 35% of individuals putatively carrying a *THCAS*-related marker. However, the identification of these fragments does not necessarily imply that the *THCAS* gene caught by such primers is complete. Van Bakel et al. (2011) reported the sequencing of reads corresponding to *THCAS* for 'FINOLA', describing them as pseudogenes, without the possibility of

assembling them into a gene encoding for a complete protein. Indeed, it has been hypothesized that these sequences are part of the complex genomic region at chromosome 7 that hosts the *cannabinoid synthases*, where functional genes, numerous pseudogenes, and additional full-length putatively functional sequences are interspersed with transposable elements (Grassa et al. 2021).

Among the hemp varieties known the be at risk to exceed the EU THC legal limit is 'FINOLA', a Finnish oilseed cultivar since 2003 in the list of crops eligible for subsidies from the European Union (https://finola.fi/news-info). It is a short, early flowering, dioecious variety and it is often used as the model *C. sativa* variety for genetic studies, since its genomic and transcriptomic data have long been available (van Bakel et al.,2011; Stout et al., 2012; Gagne et al., 2012).

In this work, the relationship among PCR markers for *THCAS*, the presence of complete coding sequences, their expression in female inflorescence and the accumulation of THC was investigated within individuals of the variety 'FINOLA' obtained from commercial certified seeds.

2. Material and methods

2.1. Plant materials

On the 20 May 2021, certified 'FINOLA' seeds were sown in paper pots with peat-based substrate and allowed to germinate in a greenhouse with natural light. Eight days after sowing, around 150 plantlets were transplanted to an experimental field in Bologna (GPS coordinates: 44.523791, 11.350306). After two weeks from the transplant, the first signs of sexual characterization became visible, and the youngest leaves of all plants were sampled for genotyping and stored at −20°C until DNA extraction.

Female plants only were allowed to reach flowering stage, then for a subset of 14 individuals, chosen based on the genotyping results, fully mature apical inflorescences (Figure 1A) were sampled, a part of them was immediately frozen in liquid nitrogen and stored at –80°C for RNA isolation and gene expression analysis, whilst the other part was oven dried at 40°C for a night for chemotype analyses.

In 2022, seeds of the same batch were germinated on wet filter papers, then 54 1-week old sprouts were moved to individual pots and grown in controlled environment (at 23°C and a photoperiod of 16/8 hours of light/dark) for about 6 weeks. Then, for each plant, leaves were collected for genotyping with the B1080/B1192 marker

associated with *CBDAS/THCA*S and male (Figure 1B) and female apical inflorescences (not fully mature) were sampled separately for chemotype analyses.

Moreover, seeds (a total of 72, taken random from the same certified batch) were also used for genotyping both for sex and for *CBDAS/THCAS*.

Figure 1. A) 'FINOLA' female mature inflorescence from the field experiment, collected for chemical and transcriptional analyses; B) 'FINOLA' male inflorescences with clusters of male flowers with closed sepals grown in controlled conditions, collected for chemical analyses.

2.2. Genotyping

2.2.1 DNA isolation

DNA was extracted from 100 mg of leaves sampled from male and female individuals grown in field and in controlled environment, and from single seeds, using the same protocol described below. Samples were homogenized in 400 µL of extraction buffer (200 mM Tris-HCl pH 8, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) with TissueLyser II (Qiagen) at 30 hz for 5 min. After centrifugation for 1 min at 13.000 rpm, nucleic

acids were precipitated adding one volume of isopropanol. After centrifugation for 5 min, the pellet was dried and resuspended in 50 µL ultrapure water.

2.2.2. Analyses of sex-associated markers

Analysis of DNA markers for sex was performed on DNA extracted from the 72 seeds using the primers developed by Mandolino et al. (1999) and the AccuStart II PCR ToughMix (Quantabio), with the specific 3-step cycling profile following manufacturer's instructions. The final reaction volume (25 µl) was composed of 12.5 µl of master mix (2X), 0.5 µM of each primer, 10 ng of template DNA and sterile water.

2.2.3. Analyses of CBDAS and THCAS-associated markers

To identify the presence of *CBDAS* and *THCAS*, the B1080/B1192 marker was used, according to Fulvio et al. (2021), using the AccuStart II PCR ToughMix (Quantabio). This marker is based on a three-primer multiplex system that can give as result either a 1080 bp fragment associated to the presence of *CBDAS*, or a 1192 bp fragment associated to *THCAS*, or both. Analyses of *CBDAS* and *THCAS*-associated markers was performed on all DNA samples extracted in this work.

2.3. Amplification, molecular cloning, and sequencing

Six female and two male plants positive for the 1192 bp amplification band were further analysed to isolate the full-length genes. The complete coding sequences were obtained by PCR using *THCAS* specific primers from Onofri et al. (2015). Amplification reactions and molecular cloning were performed after Fulvio et al. (2021). For each plant, one to five plasmids were isolated using the PureLink®Quick Plasmid Miniprep Kit (Thermo Fisher Scientific). The inserts were then sequenced with the Sanger method (Microsynth AG), using the pJET 1.2 sequencing primers. Sequences were BLAST-searched against the NCBI non-redundant database, filtered by *Cannabis sativa*.

2.4. Gene expression analysis

RNA was isolated from 100 mg of frozen samples, using the Spectrum™ Plant Total RNA Kit with on-column DNAse I treatment (Sigma Aldrich). Total RNA was eluted in 50 µL of DEPC water. Five hundred ng of total RNA were reverse transcribed with the High-Capacity RNA to cDNA kit (Thermo Fisher Scientific).

The expression of *CBDAS* and *THCAS* was evaluated in qPCR, using the specific primer pairs and the protocol described in Fulvio et al. (2021). Data were normalised against the expression of *Glyceraldehyde-3-Phosphate*

Dehydrogenase (*GAPDH*) and *tubulin* reference genes (Mangeot-Peter et al. 2016). Amplification efficiency for all assays was estimated from a regression line obtained with a five-point serial dilution of a pool of cDNA. Relative quantification was calculated using the delta delta Ct method, corrected for the PCR amplification efficiencies (Pfaffl 2001), using as calibrator the sample with the lowest expression level. Three technical replicates were performed on each sample.

2.5. Chemotype analysis

CBD and THC quantification was performed with a gas chromatograph TRACE 1300 (Thermo Scientific) equipped with autosampler and flame ionisation detector, according to the protocol detailed in Pacifico et al. (2006). Cannabinoid concentrations are given in percentage of weight over inflorescence dry weight (% w/dw). Multiple t-test with Bonferroni correction was performed using Microsoft Excel to compare the concentration of CBD, THC, and their sum in two groups of plants characterised by the presence or absence of THCAS-associated marker.

3. Results

3.1. CBDAS and THCAS associated markers

Three sets of experiments were carried out on (i) plants in field, (ii) plants in controlled environment and (iii) seeds. Looking into the details, it appears clear that the genotype with *THCAS*-associated band only is rare and it was therefore detected in different proportions (1.9 to 6.8%) in the different experiments, suggesting a bias possibly due to the sizes of the sample sets. The presence of *THCAS*- in combination with *CBDAS*-associated bands was however much more common, with a range between 23.6 and 38.6%. Finally, individuals positive only to *CBDAS*-associated marker were the most prevalent in all cases. Overall, a total of 258 individuals and seeds of 'FINOLA' from the same batch were assessed for the *CBDAS* and *THCAS* associated markers and out of these 4.7% showed only the PCR fragment of 1192 bp associated to *THCAS*, whilst 33.3% showed both the fragments associated to *CBDAS* and *THCAS* and 62.0% only the PCR fragment associated to *CBDAS*. Details about the results are shown in Table 1.

Table 1. Number of individuals, divided by sex and experiment, positive for the PCR markers for *CBDAS* (B1080), *THCAS* (B1192) or both (B1080/B1192)

Experiment	Sex	B1192	B1080/B1192	B1080	TOTAL plants
Plants in field	Male	9	38	22	69
	Female	0	13	50	63
Plants in controlled environment	Male	0	18	11	29
	Female	1	0	24	25
Seeds	Male	1	8	26	35
	Female	1	9	27	37
TOTAL markers	n.	12	86	160	258
	%	4.7	33.3	62.0	

3.2. *THCAS* sequencing

At least one full length *THCAS* sequence, identical to the functional sequence annotated with the accession number AB212829.1 (Kojoma et al. 2006) was found in each tested plant. Moreover, three new, unique, and complete sequences were obtained from female plants and submitted to GenBank with ID numbers: ON007340-42.

3.3. *CBDAS* and *THCAS* gene expression

Results of relative quantification from a subset of mature female inflorescence from the field, given in Figure 2, are clearly in agreement with the genotype revealed by the molecular marker. No expressed *THCAS* was detectable in individuals with only the *CBDAS*-associated PCR fragment, whilst both *CBDAS* and *THCAS* were expressed in plants with both fragments, suggesting that the marker B1080/1192 is suitable to detect the functional forms of these synthases, avoiding pseudogenes.

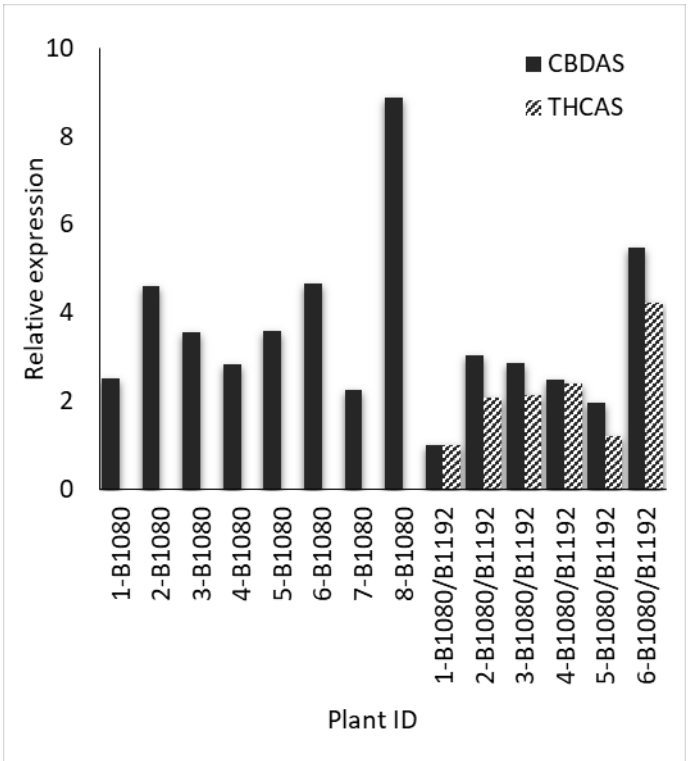


Figure 2. Relative expression of *CBDAS* and *THCAS* in selected plants. Data are normalised over the reference genes *GAPDH* and *Tubulin.* IDs of plants positive for *CBDAS*-associated marker are numbered from 1 to 8, followed by B1080; those of plants with both *CBDAS and THCAS* -associated markers are followed by B1080/B1192. Sample 1-B1080/B1192 is used as calibrator and its expression levels are set as 1.

3.4. Chemotype analysis

Individuals from the field experiment with only functional *CBDAS* in general had a higher CBD content, only traces of THC and CBD to THC ratio in the range of 13.7 to 15.9, falling within the chemotype III classification. Plants with both functional *CBDAS* and *THCAS* had on average a lower CBD content and a higher THC content, with a CBD to THC ratio between 0.4 and 0.7, as expected for chemotype II plants. The difference in both CBD and THC content between plants with and without *THCAS*-associated marker is significant, whilst the difference in the sum of the two cannabinoids is not (*adjusted p-value* equal to 0.007, 0.0005 and 0.9 respectively). The THC levels exceeded the 0.2% w/dw in all the samples of this experiment with *THCAS*-associated marker.

Chemotype analyses showed a complete agreement with the genotyping also for individuals from the experiment conducted in controlled environment. Chemotype III plants had a CBD/THC ratio, whenever THC

was above the limit of detection, in the range of 9.8 to 38.2, while chemotype II had a ratio between 0.3 and 2.4. In this experiment however, the flowers from the only female plant carrying the *THCAS* gene did not exceed the legal limit for THC, probably because they were collected before they could reach full maturity; while seven out of 18 analysed male inflorescences from plants positive for the *THCAS* gene, collected when sepals were still closed, did, reaching up to the 0.42 % w/dw. Results are shown in Table 2 and Figure 3.

Table 2. Total CBD and THC content in selected female plants from the experiment in field, expressed as % of dry inflorescence, their sum and ratio. IDs of plants positive for *CBDAS*-associated marker are numbered from 1 to 8, followed by B1080; those of plants with both *CBDAS*- and *THCAS*-associated markers are followed by B1080/B1192**.**

Plant ID	CBD % dw	THC % dw	CBD+THC % dw	CBD/THC ratio
1-B1080	1.00	0.07	1.07	14.9
2-B1080	1.23	0.08	1.31	15.3
3-B1080	0.95	0.06	1.01	14.7
4-B1080	0.62	0.04	0.66	15.9
5-B1080	0.83	0.06	0.89	15.1
6-B1080	0.82	0.06	0.88	13.7
7-B1080	0.80	0.05	0.85	15.3
8- B1080	0.65	0.04	0.69	14.8
1-B1080/B1192	0.26	0.46	0.72	0.6
2-B1080/B1192	0.57	0.84	1.41	0.7
3-B1080/B1192	0.20	0.48	0.68	0.4
4-B1080/B1192	0.52	0.98	1.50	0.5
5-B1080/B1192	0.24	0.36	0.60	0.7
6-B1080/B1192	0.76	1.28	2.04	0.6

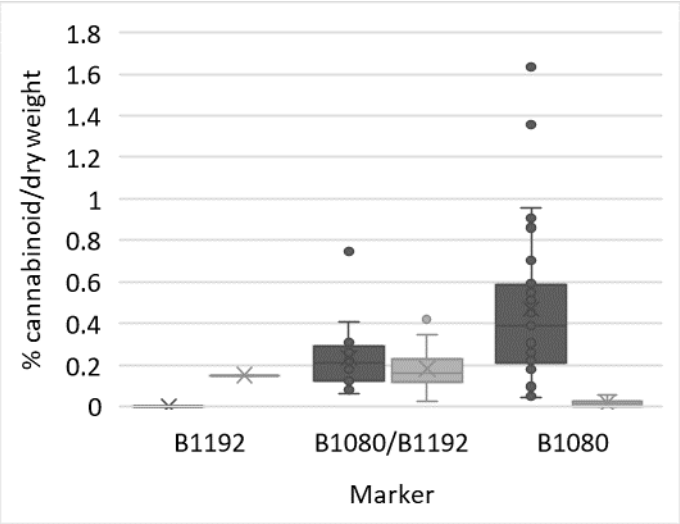


Figure 3. Box and Whiskers plot of percentage of CBD (dark grey) and THC (light grey) over dry weight detected in flowers from plants grown in controlled environment, divided according to the results of the B1080/B1192 marker. B1192 (n=1), B1080/B1192 (n=18), and B1080 (n=35).

4. Discussion

Hemp is a sustainable crop, needing low maintenance, minimal processing, and being its fiber, shiv, and seeds valuable sources of different bioproducts. Both foods and non-foods products can be produced by a single hemp plant, and this overcomes food security concerns. A recent survey of U.S. stakeholders about priorities for hemp research (Ellison, 2020) revealed the urgent need to develop genetic tools to help improve its cultivation and increase trait stability and uniformity. The dioecism and high heterozygosity of the *C. sativa* species result in a high genetic and phenotypic diversity, which for breeders is a goldmine of traits that can be selected and used for genetic improvement. However, for growers, unstable genetics lead to poor harvests, low production and yields, resulting in significant economic losses, and sometimes also in more serious consequences, such as the seizure of entire crops due to high THC levels (Olejar and Park, 2022).

Among the most popular cultivars there is 'FINOLA', grown mostly for its seeds and oil. Overall, in the 'FINOLA' batch here investigated, 38% of all tested individuals carried the *THCAS* gene, alone or in combination with *CBDAS*. These numbers are much higher than the 15% observed in 2020 by Borroto Fernandez et al., suggesting that different origin of certified seeds can heavily affect the rate of *THCAS*

occurrence within a hemp variety and potentially pose risks to exceed THC legal limit. Indeed, from all male and female plants randomly chosen among the ones positive for the *THCAS*-associated marker, it was possible to isolate full-length *THCAS* coding sequences. Each of the analysed DNA carries at least one full-length gene identical to the sequence with accession number AB212829, and three new sequences were also identified from female plants. To the best of our knowledge, this is the first report concerning the presence of one or more full-length functional *THCAS* in the genome of an industrial hemp variety obtained from certified seeds. In mature inflorescences from the field experiment positive for the B1192 marker, it was also detected an active transcription of *THCAS* genes. No expression was instead detectable in plants positive only to the *CBDAS*-associated marker. Moreover, the expression of *THCAS* correlates to the amount of THC accumulated in female flowers (Pearson correlation coefficient = 0.9), which for these plants was in all cases much higher than the legal limit and in one case even exceeding 1%. Residual THC in plants in which *THCAS* expression was not detected was on the contrary in all cases lower, and probably caused by unspecific enzymatic activity (Zirpel et al., 2018; Toth et al., 2020; Fulvio et al., 2021).

In the experimental subset analysed from the field experiment, positivity for the *THCAS* marker not only perfectly predicts the presence and the expression of this gene but was also able to identify individuals in which THC accumulation was higher than the legal limit, as they possess a *THCAS* sequence encoding for a fully functional enzyme. These results enrich the information offered by the only few studies that have analysed this issue in hemp so far, suggesting that the presence of *THCAS*-related marker indicates the presence of THCA content above 0.3% in mature inflorescences (Borroto Fernandez et al., 2020; Toth et al., 2020). Despite the need to better understand these mechanisms, it seems unlikely that in plants lacking such functional THCAS the residual THC would ever reach levels at risk of triggering legal issues. Our data strongly suggest that the problem of "hot hemp", containing THC above legal limits, is genetically based, not due to environmental stresses. To avoid THC in hemp crops it is necessary to avoid the presence of *THCAS* and in practical terms the key to achieve this result is the use of marker assisted selection during seed multiplication for improving varietal selection and guaranteeing certified seed batches with THC levels below the legal limit.

Starting from these observations of presence, expression and products of functional THCA synthase in the field experiment, we then repeated a similar investigation on plants grown in a controlled environment, where

male flowers were also tested for cannabinoids. Concordance between genotyping and chemical results was again complete and, interestingly, in many instances these male inflorescences exceeded the legal limit for THC. For breeding purposes, we then tried to apply the genotyping protocol also to seeds, and we showed that it could be easily used to characterize seed batches, as well as to guide breeding choices, in the attempt to remove functional *THCAS* from hemp varieties.

The presence of functional *THCAS* contaminated seeds could result from cross-pollination or could originate from unstable genetics. *C. sativa* male plant produces a huge amount of pollen grains, evolved for maximum dispersal by wind, which makes them capable of traveling long distances (Cabezudo et al., 1997; Small and Antle, 2003). However, as pointed out by Olejar and Park (2022), a hemp field exceeding legal THC as a consequence of cross-pollination should not occur where only certified seed of varieties with low THC level is admissible for hemp cultivation, as it is the case in EU. To avoid the potential risk of having crops with uncertain genetics, efforts would be required to ensure that the new seed lots are produced in safe conditions, within large buffer zones or even in indoor facilities to preserve the varietal purity.

On the other hand, modern hemp varieties were obtained through a counter-selection for THC level and it cannot be excluded that without a strict selection during seed multiplication the *THCAS* occurrence in hemp varieties will increase. This work promotes the discussion of the still little-debated issue of the hemp-varieties genetic stability, which is still a long way from being achieved. The use by hemp breeders and seed companies of genetic tools, such as the marker used in this study, and others developed more recently, must be encouraged as they are a reliable support for the whole hemp supply chain to ensure greater uniformity and varietal correspondence, securing crops and growers from possible seizure.

5. Conclusions

To date in the EU the THC content limit for hemp cultivars is at 0.2% of inflorescence dry weight (set to increase to 0.3% from 2023). Although selection has been applied for years to industrial varieties in the attempt to eliminate the accumulation of this cannabinoid, still large genetic and phenotypic instability is found. The present work describes the identification of *THCA synthase* sequences via molecular markers and sequencing in a population obtained from commercial seed of ‘FINOLA’ and the transcription of this gene in

inflorescences, in association with a consistent final concentration of the metabolite. Such a wide distribution of the functional *THCAS* gene in a population from certified seeds may lead to an excessive accumulation of the psychotropic cannabinoid, putting the cultivation at high risk of seizure and therefore endangering both crop and grower. The origin of this wide diffusion is still unclear, and further research should be carried out with the aim of obtaining industrial hemp varieties that are stable in terms of residual THC content beyond the environmental conditions. The genotyping by molecular markers, allowing an early prediction of the chemotype, therefore becomes an effective tool to be used during breeding and multiplication.

6. Funding

This research was funded by UNIHEMP research project "Use of iNdustrIal Hemp biomass for Energy and new biocheMicals Production" (ARS01_00668) funded by European Regional Development Fund (within the PON R&I 2017-2020–Axis 2—Action II—OS 1.b). Grant decree UNIHEMP prot. no. 2016 of 27/07/2018; CUP B76C18000520005.

7. References

Borroto Fernandez, E., Peterseil, V., Hackl, G., Menges, S., de Meijer, E., Staginnus, C., 2020. Distribution of Chemical Phenotypes (Chemotypes) in European Agricultural Hemp (Cannabis sativa L.) Cultivars. J. Forensic Sci. 65, 715–721. https://doi.org/10.1111/1556-4029.14242

Braich, S., Baillie, R.C., Jewell, L.S., Spangenberg, G.C., Cogan, N.O.I., 2019. Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis. Sci. Rep. 9, 1–12. https://doi.org/10.1038/s41598-019-53023-6

Cabezudo, B., Recio, M., Sanchez-laulhi, M., del Mar Trigo, M., Javier Toro, F., Polvorinos, F., 1997. Atmospheric transportation of marihuana pollen from North Africa to the southwest of Europe, Atmos. Environ.

Cerrato, A., Citti, C., Cannazza, G., Capriotti, A.L., Cavaliere, C., Grassi, G., Marini, F., Montone, C.M., Paris, R., Piovesana, S., Laganà, A., 2021. Phytocannabinomics: Untargeted metabolomics as a tool for cannabis chemovar differentiation. Talanta 230, 122313. https://doi.org/10.1016/J.TALANTA.2021.122313

Ellison, S., 2021. Hemp (Cannabis sativa L.) research priorities: Opinions from United States hemp stakeholders. GCB Bioenergy 13, 562–569. https://doi.org/10.1111/gcbb.12794

Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., Moschella, A., Alberti, I., Pecchioni, N., Cannazza, G., Mandolino, G., 2021. Analysis of sequence variability and transcriptional profile of cannabinoid synthase genes in cannabis sativa l. Chemotypes with a focus on cannabichromenic acid synthase. Plants 10. https://doi.org/10.3390/plants10091857

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. Proc Natl Acad Sci U S A 109, 12811–12816. https://doi.org/10.1073/pnas.1200330109

Gao, S., Wang, B., Xie, S., Xu, X., Zhang, J., Pei, L., Yu, Y., Yang, W., Zhang, Y., 2020. A high-quality reference genome of wild Cannabis sativa. Hortic. Res. https://doi.org/10.1038/s41438-020-0295-3

Grassa, C.J., Weiblen, G.D., Wenger, J.P., Dabney, C., Poplawski, S.G., Motley, S.T., Michael, T.P., Schwartz, C.J., 2021. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana, New Phytol. https://doi.org/10.1111/nph.17243

Gülck, T., Møller, B.L., 2020. Phytocannabinoids: Origins and Biosynthesis. Trends Plant Sci. 25, 985–1004. https://doi.org/10.1016/j.tplants.2020.05.005

Kojoma, M., Seki, H., Yoshida, S., Muranaka, T., 2006. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" Cannabis sativa L. Forensic Sci. Int. 159, 132–140. https://doi.org/10.1016/j.forsciint.2005.07.005

Mandolino, G., Carboni, A., 2004. Potential of marker-assisted selection in hemp genetic improvement. Euphytica 140, 107–120. https://doi.org/10.1007/s10681-004-4759-6

Mangeot-Peter, L., Legay, S., Hausman, J.F., Esposito, S., Guerriero, G., 2016. Identification of reference genes for RT-qPCR data normalization in Cannabis sativa stem tissues. Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17091556

Olejar, K.J., Park, S.H., 2022. Industry-Based Misconceptions Regarding Cross-Pollination of Cannabis spp. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.793264

Onofri, C., de Meijer, E.P.M., Mandolino, G., 2015. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. Phytochemistry 116, 57–68. https://doi.org/10.1016/j.phytochem.2015.03.006

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., Mandolino, G., 2006. Genetics and marker-assisted selection of the chemotype in Cannabis sativa L. Mol. Breed. 17, 257–268. https://doi.org/10.1007/s11032-005-5681-x

Pacifico, D., Miselli, F., Carboni, A., Moschella, A., Mandolino, G. 2008. Time course of cannabinoid accumulation and chemotype development during the growth of Cannabis sativa L. Euphytica, 160(2), 231–240. https://doi.org/10.1007/s10681-007-9543

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR, Nucleic Acids Res.

Small, E., Antle, T., 2003. A preliminary study of pollen dispersal in cannabis sativa in relation to wind direction. J. Ind. Hemp 8, 37–50. https://doi.org/10.1300/J237v08n02_03

Staginnus, C., Zörntlein, S., de Meijer, E., 2014. A PCR marker Linked to a THCA synthase polymorphism is a reliable tool to discriminate potentially THC-rich plants of cannabis sativa L. J. Forensic Sci. 59, 919–926. https://doi.org/10.1111/1556-4029.12448

Stout, J.M., Boubakir, Z., Ambrose, S.J., Purves, R.W., Page, J.E., 2012. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. Plant J. 71, 353–365. https://doi.org/10.1111/j.1365-313X.2012.04949.x

Toth, J.A., Stack, G.M., Cala, A.R., Carlson, C.H., Wilk, R.L., Crawford, J.L., Viands, D.R., Philippe, G., Smart, C.D., Rose, J.K.C., Smart, L.B., 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. Glob. Change Biol. Bioenergy 12, 213–222. https://doi.org/10.1111/gcbb.12667

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., Page, J.E., 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biol. 12. https://doi.org/10.1186/gb-2011-12-10-r102

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. Am. J. Bot. 107, 1423–1432. https://doi.org/10.1002/ajb2.1550

Zirpel, B., Degenhardt, F., Zammarelli, C., Wibberg, D., Kalinowski, J., Stehle, F., Kayser, O., 2018. Optimization of Δ9-tetrahydrocannabinolic acid synthase production in Komagataella phaffii via post-translational bottleneck identification. J. Biotechnol. 272–273, 40–47. https://doi.org/10.1016/j.jbiotec.2018.03.008

Chapter 4

Optimization of a protocol for micropropagation of Italian *Cannabis sativa* L. varieties for pharmaceutical uses

This chapter has been extracted from "Ottimizzazione di un protocollo per la micropropagazione di varietà italiane di *Cannabis sativa* L. per uso medico" from Roberta Paris[1], Flavia Fulvio[1,2], Anna Moschella[1], Cristina Baldin[3], Ilaria Alberti[3] submitted to ACTA ITALUS HORTUS on 13/01/2023

[1] CREA—Research Centre for Cereal and Industrial Crops, Via di Corticella 133, 40128 Bologna, Italy;

[2] Department of Sciences of Agriculture, Food Natural Resources and Engineering, University of Foggia, Via Napoli 25, 71122 Foggia, Italy

[3] CREA—Research Centre for Cereal and Industrial Crops, Via G. Amendola 82, 45100 Rovigo, Italy;

Abstract: *Cannabis sativa* (L.) is rich in active metabolites of medical interest, like phytocannabinoids, among which the majors are Δ^9-tetrahydrocannabinol and cannabidiol. The two varieties currently used for the national production of *Cannabis*-based drugs have been developed at CREA-CI and mother plants are maintained exclusively by agamic route in indoor facilities. In this work, the optimization of an *in vitro* micropropagation protocol is reported, used to support the *in vivo* maintenance and multiplication activities of the mother plants. To optimize explants multiplication and elongation, three treatments with different hormone combinations were tested. In addition, an auxin-supplemented medium was compared to the control medium to evaluate its effect on promoting root generation. Following each trial, the chemotype of explants was verified via molecular markers and gas chromatography using the leaves as plant material for the analyses. Subsequently, the protocols were extended also to materials of an èlite genotype collection, to maintain the most interesting accessions also through *in vitro* culture. Together our findings represent an attempt to overcome the lack of research on *C. sativa* that limited the progress of this crop and to implement *in vitro* micropropagation protocols for germplasm maintenance and as a tool to support breeding and apply the New Breeding Technques.

1. Introduction

Cannabis sativa L., is a plant of great interest both for fiber and seed production (industrial hemp), and as medicinal *Cannabis*, whose female inflorescences, rich in bioactive compounds, are used for the production of drugs. Among these, phytocannabinoids, are the most significant compounds. Of the approximately 150 identified to date (Hanus et al., 2016), the main and most abundant are cannabigerolic acid (CBGA), cannabidiolic acid (CBDA), tetrahydrocannabinolic acid (THCA) and cannabichromenic acid (CBCA), with their respective neutral forms (CBG, CBD, THC and CBC). Their different quali-quantitative composition and the presence of other classes of biomolecules (including terpenoids and phenylpropanoids), acting in synergy with the former (resulting in the so-called "entourage effect"), determine the different therapeutic effects of *Cannabis* preparations. The discovery of the endocannabinoid system and the receptors on which phytocannabinoids act (Mechoulam and Parker, 2015) has certainly contributed to the renewal of interest in the use and research on hemp for medical use, through the production of different types of *Cannabis*-based drugs. Some are produced synthetically (such as Dronabinol and Nabilon); others are made from *C. sativa* extracts, such as Sativex and Epidiolex (Legare et al., 2022). Finally, unlike other herbal medicines, mature and dried inflorescences of *C. sativa* can be used as such. In this case, only standardized cultivation protocols allow the production of inflorescences with the desired amount of phytocannabinoids and under the most suitable phytosanitary conditions.

Cannabis-based drug production began in Italy in 2016 at the Military Chemical-Pharmaceutical Factory (SCFM) in Florence following an agreement between the Ministry of Defence and the Ministry of Health and includes the standardized cultivation of plant clones derived from mother plants and the pharmaceutical processing of the female inflorescences. The plant production cycle begins at CREA-CI's Rovigo facility with the preparation of rooted herbaceous cuttings obtained from mother plants, bred indoors. Two magistral preparations, FM1 and FM2, based on inflorescences of two varieties of *C. sativa*, CINBOL and CINRO developed by CREA-CI in Rovigo, are available to date. CINBOL is characterized by a predominant THCA chemical phenotype (chemotype I), while CINRO has a CBDA:THCA ratio of 1.5:1 (chemotype II). In this work, conducted within the framework of the national projects Pronacanamed (DM 0089914 of 02/12/2016) and CAMED (collaborative agreement under Art. 15 of Law 241/90, MIPAAF-2021-0181330), the possibility

of flanking the maintenance and renewal of mother plants, and propagation activity by cuttings with micropropagation was evaluated. Indeed, medical *Cannabis* is generally propagated using clonal methods, to achieve mass-production of high-quality plants genetically and phenotypically uniform. Micropropagation represents an alternative to clonal propagation, allowing the production of a high number of plants without the need for large spaces, and obtaining disease-free plants as they are cultivated in aseptic environments without the intervention of abiotic stresses. Although several protocols have been published for this notoriously recalcitrant species (Adhikary et al., 2021; Monthony et al., 2021), considerable differences and limitations are still reported in the response of culture according to the genotype. The objective of the present study was the development and optimization of a micropropagation protocol suitable for the medical Italian varieties and other élite clones, belonging to plant germplasm and representing pre-breeding material to breed new genotypes for Pharmaceutical and Medical Industry.

2. Material and methods

2.1 Plant material

As a source of explant, *C. sativa* mother plants of the CINBOL and CINRO variety were grown indoor in 11-liter pots containing rockwool granules, under controlled temperature and humidity conditions (T 24°C; RH 60%) and exposed to long photoperiod (18 h of light) with LED lamps to maintain vegetative growth for 6-9 months (Fig. 1A). Young apical shoots of 4-5 cm were taken from the mother plants and used for *in vitro* propagation.

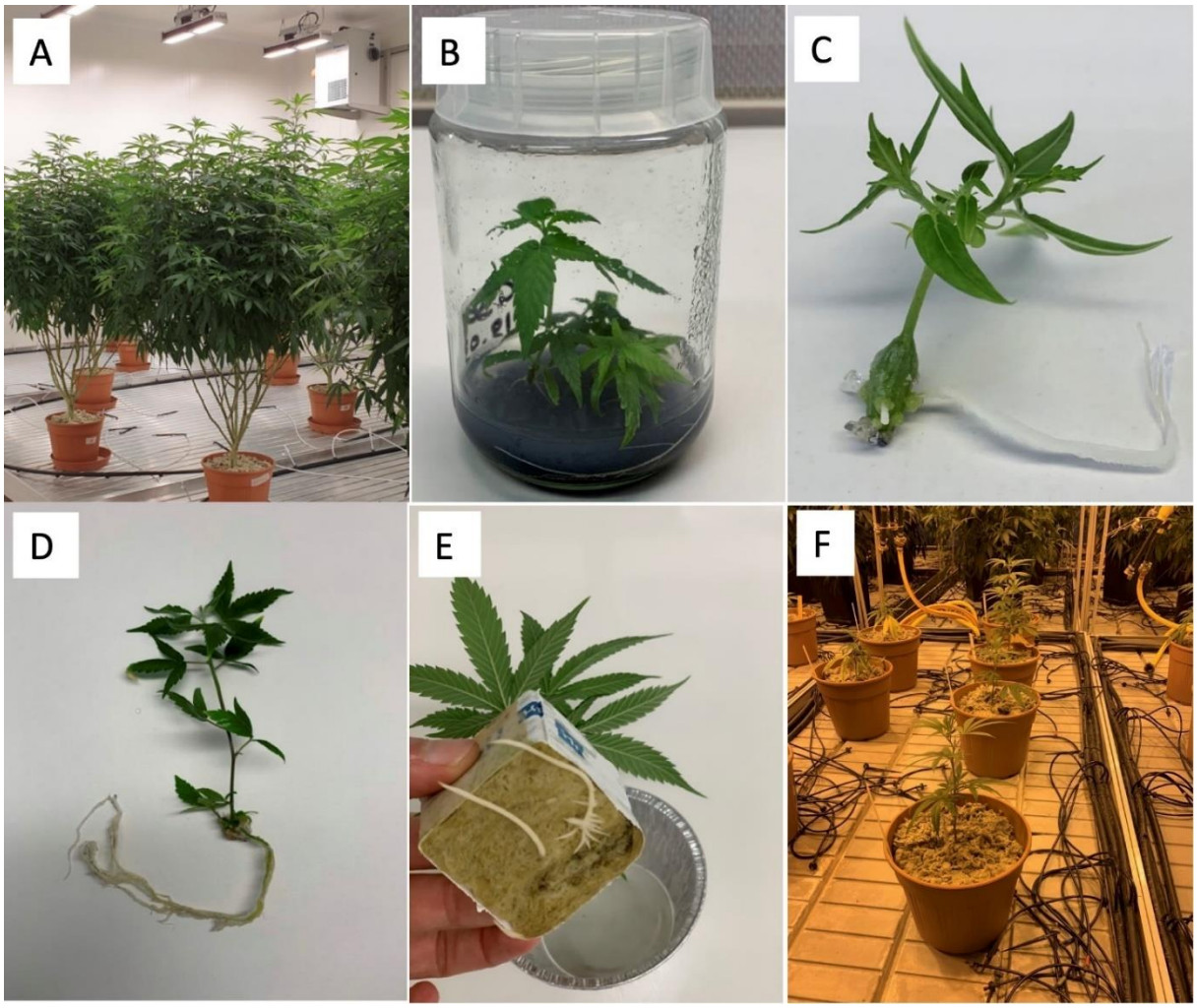


Figure 1. A. Mother plants in the growth chamber; B. *C. sativa* micropropagated shoots on MS30ca medium; C: Shoots proliferation on MS culture medium supplemented with TDZ and GA3; D: Micro-sprout with root system; E: *Ex vitro* sprouts in rock wool blocks; F: Sprouts from acclimatized *in vitro* culture.

2.2 Sterilization of explants and *in vitro* propagation of explants.

To initiate shoots *in vitro*, explants free of disease symptoms and nutritional deficiencies were harvested from stock plants. Leaves were removed from explants, which were placed in a solution of 0.5% (w/v) sodium hypochlorite and 0.1% Tween 20 for 15 min, intermittently shaking, followed by two washes in sterile water under a laminar flow sterile hood (Lubell et al., 2021). After drying on filter paper, the explants were cut, separating apex and 2-cm stem portions comprising one or two nodes, and placed in baby food culture jars (4-cm diameter x 9. 5-cm high) sealed with Magenta-cap, into which was dispensed 25 mL of sterile initiation culture medium. The initiation medium contained macro elements, micro elements, and vitamins according to

Murashige and Skoog (1962) fortified with 30 gr/L sucrose and 8 gr/L agar (SIGMA) adjusted at pH 5.7. The cultures were maintained for about 4 weeks in Binder growth chamber illuminated with L 18W 865 Cold Light Fluorescent Lamps (43 W/m2; 140000 Lx) at 25°C, and long photoperiod (18 h). Then the cultures were transferred and maintained for 3-4 weeks on MS30 medium supplemented with 500 mg/L of activated charcoal (MS30ca).

2.3 In vitro multiplication

To stimulate the elongation and development of axillary shoots, three hormone combinations were used in addition to MS30ca medium:

- T05: MS30ca supplemented with Thidiazuron (TDZ) 0.5 µM (Lata et al., 2009).
- T057: MS30ca fortified with TDZ 0.5 µM in combination with gibberellic acid (GA3) 0.7 µM (Lata et al., 2009).
- mT2: MS30ca with metatopoline (mT) 2 µM (Lata et al., 2016).

The initial test was conducted on 10 shoots per genotype on the 4 different media. Subsequently, 20 explants were evaluated on the two media that gave the best result.

2.4 Rooting of explants and ex vitro acclimation

For rooting of explants, MS30ca culture medium was compared with MS medium with a halved concentration of salts and vitamins (MS/2), supplemented with 2.5 M indolbutyric acid (IBA). Rooted seedlings were gently taken out of the medium, and the roots were washed with sterile water from the residual medium adhering to the roots. Then they were transferred *ex vitro* into rockwool cubes, previously acidified with sterile water at pH 4.5 overnight, then kept for 4 hours in a nutritive solution made from Canna Coco A + B, placed in plastic trays covered with film, in order to maintain high humidity, and kept in the same growth chamber. Humidity was gradually reduced by making holes in the film until fully acclimatized.

2.5 Chemotype assessment

The chemotype of plants was verified by DNA analysis with molecular markers B1080/B1192 according to the protocol detailed in Fulvio et al. (2021) and by quantification of total CBD and THC in the leaves using a TRACE 1300 gas chromatograph (Thermo Scientific, Milan, Italy) equipped with FID detector, as reported in

Pacifico et al. (2006). Cannabinoid concentrations are given as percent by weight over the dry weight of the starting material (% w/w).

3. Results and discussions

The main objective of this work was to test some existing protocols for the propagation of *C. sativa* to verify their adaptability for two genotypes (CINBOL and CINRO) currently used for the national production of medicinal *Cannabis* in Italy.

The sterilization protocol employed resulted in effective disinfection of the source material, close to 100% of the explants. This result is influenced by the fact that the starting material was raised indoors to reduce the risk of the occurrence of phytosanitary problems.

The newly developed shoots were transferred after 4 weeks to the new growth medium supplemented with 500 mg/L activated charcoal (MS30ca), in which they can remain for up to 8 weeks, before renewal (Fig. 1B). Plants of both varieties benefited from the presence in the culture medium of activated charcoal to improve cell growth and development and indicated as useful in removing plant toxic exudates, such as polyphenols (Thomas, 2008). However, even after two months of culture the shoots do not multiply and showed little elongation (average length around 2 cm).

To promote shoot proliferation, after 4 weeks of culture in the MS30ca medium, three hormone combinations were tested using two different cytokinins (TDZ and mT), in the presence or absence of gibberellin.

Several reports regarding TDZ use in *Cannabis* micropropagation, suggest that TDZ induces shoot regeneration better than other cytokinins (Lata et al., 2009; Lata et al.,2016; Piunno et al., 2019). Here we tested TDZ, alone and in combination with gibberellin (GA_3) to evaluate its potential synergistic effect. Indeed, it was reported that cytokinins, by stimulating the proliferation of shoots inhibit stem elongation and consequently the plant develops very short internodes.

The optimal response in terms of proliferation was recorded on MS medium supplemented with TDZ and GA_3, for both varieties (Fig. 1C), with an average number of shoots developed from a single explant of 8.0±4.9 and 5.7±1.0 for CINBOL and CINRO, respectively. Shoots had an average height of 1.1±0.2 for CINBOL and 0.9±0.4 cm for CINRO.

However, the use of TDZ (0.5 μM) alone did not show comparable results in terms of shoot formation. For CINRO the average number of shoots developed from a single explant was 3.5±2.1 and 1.2±1.2 for CINBOL, with an average height of 0.5±0.1 and 0.9±0.2 cm, respectively.

The use of mT at the indicated dose proved to be unsuitable for multiplication of both varieties, and even lethal for most plants of CINBOL (70%), without positive effects compared with the MS30ca, but rather showing worse results in terms of the number and quality of explants.

After 8 weeks of culture, explants grown on the standard MS30ca medium succeeded in rooting (70.8% for CINBOL and 75% for CINRO), showing a well-developed root system, as observed in Fig. 1D. No rooting was obtained on cytokinin-added medium (T057 e *m*T2) nor IBA-added MS/2 after 8 weeks.

Plants with a developed root system were acclimated with a success rate close to 100% (Fig. 1E). Plants were gradually transferred to larger pots containing inert substrate and transferred to indoor facilities (Fig. 1F) to form new mother plants for cuttings production after verification of chemotype maintenance.

Chromatographic analysis (Fig. 2A, B), quantification of CBD and THC (Fig. 2C) and molecular markers (Fig. 2D) confirmed the chemotype of the plants *in vitro*.

The protocol developed on the two medical *Cannabis* varieties is currently in use for the maintenance of 12 élite genotypes from the CREA germplasm collection.

The illustrated procedure was also applied to a varietal selection experiment for obtaining a high CBDA line from feminized seed. From about 40 seedlings in the vegetative stage (40 days post-sowing), at least 4 explants were taken and subcultured until the plants reached maturity, when they were analysed for cannabinoid content. Based on the results obtained from the inflorescence analysis, 4 plants with the desired trait (high CBD) were finally selected to undergo further cultivation cycles and select the excellent clone that will be used for the production of mother plants with high CBD content for pharmaceutical use. In fact, the level of THC was detected above the 0.3% legal limit set by the Common Agricultural Policy (Regulation (EU) 2021/2115) and for this reason, this accession can't be classified as industrial hemp, even if it has a chemotype III as most of the certified varieties registered in the European Catalogue.

4.Conclusions

The use of micropropagation of medical *Cannabis* genotypes is extremely useful and highly desirable to ensure the availability of material with a well-defined chemical profile and phytosanitary safe. Maintenance of hormone-free synthetic medium does not require special costs. Multiplication can be achieved through a 2-3week subculture on the medium fortified with cytokinin TDZ and gibberellic acid. Rooting occurs on the standard medium after about 8 weeks of culture with percentages above 70%. The procedure using propagation techniques, molecular markers and cannabinoid content analysis is an approach that can be employed for the selection and maintenance of elite germplasm.

5. Funding

This research was funded by PRONACANAMED: Pilot project for the national production of *Cannabis*-based substances and preparations of plant origin' financed by the MIPAAF (12/2016 - 06/2019) and by CAMED - National medical cannabis: innovation and enhancement of the production of therapeutic cannabis plant material for national needs and research for the establishment of new varieties for pharmaceutical use. Collaboration agreement ex art. 15 of Law 241/90), MIPAAF-2021-0181330. (17/06/2021 - 16/06/2023).

6. References

Adhikary D., Kulkarni M., El-Mezawy A., Mobini S., Elhiti M., Gjuric R., Ray A., Polowick P., Slaski J. J., Jones M.P., Bhowmik P. 2021. Medical Cannabis and Industrial Hemp Tissue Culture: Present Status and Future Potential. Front. Plant Sci. 12:627240. doi: 10.3389/fpls.2021.627240

Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., Moschella, A., Alberti, I., Pecchioni, N., Cannazza, G., Mandolino, G., 2021. Analysis of sequence variability and transcriptional profile of cannabinoid synthase genes in *Cannabis sativa* L. chemotypes with a focus on cannabichromenic acid synthase. Plants 10: 1857. https://doi.org/10.3390/plants10091857

Lata H, Chandra S, Khan I, ElSohly MA. 2009. Thidiazuron-induced high-frequency direct shoot organogenesis of Cannabis sativa L. In Vitro Cell Dev Biol Plant. 45:12–19. https://doi.org/10.1007/s11627-008-9167-5

Lata H., Chandra S., Techen N., Khan I.A., Elsohly A., 2016. In vitro mass propagation of Cannabis sativa L.: A protocol refinement using novel aromatic cytokinin meta-topolin and the assessment of eco-physiological, biochemical and genetic fidelity of micropropagated plants. Journal of Applied Research on Medicinal and Aromatic Plants 3:18–26.

Legare CA, Raup-Konsavage WM, Vrana KE. Therapeutic Potential of Cannabis, Cannabidiol, and Cannabinoid-Based Pharmaceuticals. Pharmacology. 2022;107(3-4):131-149. doi: 10.1159/000521683. Epub 2022 Jan 28. PMID: 35093949.

Lubell-Brand J.D., Kurtz L.E., Brand M.K., 2021. An In Vitro-Ex Vitro Micropropagation System for Hemp. HortTechnology 31(2): 199–207.

Mechoulam R., Parker L., 2015. The Endocannabinoid System and the Brain. Annual Review of Psychology 64:21-47.

Monthony A.S., Page S.R., Hesami M., Jones A.M. P., 2021. The Past, Present and Future of Cannabis sativa Tissue Culture. Plants 10(1): 185.

Murashige, T. and Skoog, F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Plant Physiology, 15, 473-497. https://doi.org/10.1111/j.1399-3054.1962.tb08052.x

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., Mandolino, G., 2006. Genetics and marker-assisted selection of the chemotype in Cannabis sativa L. Molecular Breeding 17, 257–268. https://doi.org/10.1007/s11032-005-5681-x

Concluding remarks

The use of *C. sativa* for therapeutic purposes is well documented since ancient times, but due to its narcotic effect and the development of alternative medicines, its potential remained underappreciated, until, in the 21st century, its benefits came powerfully back into the limelight through a large number of studies that have demonstrated its usefulness against a large set of diseases.

The breeding of genetic materials for industrial applications and the development of appropriate protocols for cultivation are prerequisites to guarantee the production of standardized products based on *Cannabis*.

Due to the state of the illegality of the plant for historical legal restrictions and stigma surrounding cannabis, research has been limited and there is still much to learn about the optimal growing conditions, cultivation techniques, and potential uses of cannabis for industrial purposes.

Overall, the lack of research on industrial and medical cannabis is a significant barrier to fully realizing the potential benefits of this plant. It is crucial that policymakers, researchers, and industry stakeholders prioritize funding and resources to support further research in this area, so that we can develop a more comprehensive understanding of the benefits and risks associated with cannabis as a crop.

In recent years, we have witnessed a considerable increase in literature focusing both on the medicinal properties of the plant and the agronomic and physiological aspects of its cultivation. However, there is also a rich segment of insiders who, beyond the most accredited scientific sources, have developed their own knowledge of the plant, which, being mostly privately funded research, is not disseminated and therefore remains the prerogative of a fortunate few. A common effort to cooperation and organisation all the scattered findings could help in the process of improving *Cannabis*.

The results obtained in this thesis cover a wide range of aspects that are important for future research both on industrial and on medicinal cannabis. The research benefited from access to a unique *Cannabis* genetic resource collection conserved at CREA-CI located in Rovigo.

The results of the first chapter examined the sequence variability of *CBCAS* genes in different accessions, and although these genes are almost completely identical, 28 polymorphisms leading to amino acid mutations were identified. It was hypothesised that given the high sequence similarity between the *CBCAS* and *THCAS* genes, the former might be involved in the synthesis of the little, but ubiquitous THCA in industrial varieties.

However, no studies have so far reported the formation THCA as a by-product of the CBCAS enzyme, suggesting that this enzyme is not the source of THCA in strains lacking a functional THCA enzyme. To further investigate the role of the *CBCAS* gene, the second chapter investigated its transcription in a set of samples from the early developmental stages of different chemical phenotypes. A correlation was observed between CBCA accumulation and *CBCAS* transcription, suggesting that the gene is responsible for metabolite biosynthesis, which occurs predominantly in the very early stages of seedling development. Furthermore, in this work, the transcription of the *PT1* gene was identified as the rate-limiting step for cannabinoid synthesis. The finding of the stage when phytocannabinoids started to be synthesized and accumulate in the plants is also of great impact on the researchers' community, suggesting the most suitable stage which study the gene expression of candidate genes for the regulation of this pathway, which is still scarsely known.

These first chapters offer the opportunity to observe the relationship between chemotype and genotype through a forward genetics approach, and at the same time provide the basis for a more accurate study of the *CBCAS* gene, the understanding of whose regulation may provide valuable information for defining the biological significance of CBCA accumulation at such a crucial stage of plant life.

Overall, the third chapter provides important insights into the genetic basis of THC production in hemp plants and highlights the potential for using molecular markers to improve the efficiency and legality of hemp farming, potentially avoiding crop losses due to legal noncompliance. The presence of excessive amounts of THC in industrial *C. sativa* inflorescences is a major concern among growers, so much so that the limit for cultivation was recently raised to 0.3% w/w. Although selection has been applied to industrial varieties for years in an attempt to eliminate the accumulation of this cannabinoid, there is still great genetic instability.

Another interesting result of this work is the finding that male inflorescences also accumulate phytocannabinoids, especially THC, to a significant extent, not previously reported in any scientific paper. In future studies, it will be worth investigating this aspect in order to understand whether the male inflorescence may also represent a source of cannabinoids or other valuable bioactive compounds.

In the last chapter, chemotype detection becomes a useful tool for verifying the correct *in vitro* preservation of the two genotypes currently used for the production of medical *Cannabis* in Italy. This work consisted of the

development of a protocol involving the use of an efficient and reliable hormone medium for the mass production and conservation of germplasm under aseptic conditions and long periods.

All the findings presented in this thesis represent an important starting point for functional genomics approaches in *Cannabis*, to exploit the different potential of this plant aimed at the development of improved genotypes for cannabinoid-based therapies or a wide range of industrial utilization involving the exploitation of the chemically and genetic intrinsic diversity of the plant.

List of publications produced during the PhD program

Bassolino, L., Buti, M., **Fulvio, F.**, Pennesi, A., Mandolino, G., Milc, J., Francia, E., Paris, R. In silico identification of myb and bhlh families reveals candidate transcription factors for secondary metabolic pathways in cannabis sativa L. (2020). DOI: 10.3390/plants9111540

Martinelli, T., **Fulvio, F.**, Pietrella, M., Focacci, M., Lauria, M., Paris, R. In Silybum marianum Italian wild populations the variability of silymarin profiles results from the combination of only two stable chemotypes (2021). DOI: 10.1016/j.fitote.2020.104797

Fulvio, F., Martinelli, T., Paris, R. Selection and validation of reference genes for RT-qPCR normalization in different tissues of milk thistle (Silybum marianum, Gaert.) (2021). DOI: 10.1016/j.gene.2020.145272

Pieracci, Y., Ascrizzi, R., Terreni, V., Pistelli, L., Flamini, G., Bassolino, L., **Fulvio, F.**, Montanari, M., Paris, R. Essential oil of cannabis sativa l: Comparison of yield and chemical composition of 11 hemp genotypes (2021). DOI: 10.3390/molecules26134080

Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., Moschella, A., Alberti, I., Pecchioni, N., Cannazza, G., Mandolino, G. Analysis of sequence variability and transcriptional profile of cannabinoid synthase genes in cannabis sativa l. Chemotypes with a focus on cannabichromenic acid synthase (2021). DOI: 10.3390/plants10091857

Linciano, P., Russo, F., Citti, C., Tolomeo, F., Paris, R., **Fulvio, F.**, Pecchioni, N., Vandelli, M.A., Laganà, A., Capriotti, A.L., Biagini, G., Carbone, L., Gigli, G., Cannazza, G. The novel heptyl phorolic acid cannabinoids content in different Cannabis sativa L. accessions (2021). DOI: 10.1016/j.talanta.2021.122704

Russo, F., Tolomeo, F., Vandelli, M.A., Biagini, G., Paris, R., **Fulvio, F.**, Laganà, A., Capriotti, A.L., Carbone, L., Gigli, G., Cannazza, G., Citti, C. Kynurenine and kynurenic acid: Two human neuromodulators found in Cannabis sativa L. (2022). DOI: 10.1016/j.jpba.2022.114636

Fulvio, F., Martinelli, T., Bassolino, L., Pietrella, M., Paris, R. A single point mutation in a member of FAD2 multigene family resulted in the generation of a high oleic line of Silybum marianum (L.) Gaertn. (2022). DOI: 10.1016/j.indcrop.2022.114930

Tolomeo, F., Russo, F., Kaczorova, D., Vandelli, M.A., Biagini, G., Laganà, A., Capriotti, A.L., Paris, R., **Fulvio, F.**, Carbone, L., Perrone, E., Gigli, G., Cannazza, G., Citti, C. Cis-Δ9-tetrahydrocannabinolic acid occurrence in Cannabis sativa L. (2022). DOI: 10.1016/j.jpba.2022.114958

Martinelli, T., **Fulvio, F.**, Pietrella, M., Bassolino, L., Paris, R. Silybum marianum chemotype differentiation is genetically determined by factors involved in silydianin biosynthesis (2023). DOI: 10.1016/j.jarmap.2022.100442

Fulvio, F., Righetti, L., Minervini, M., Moschella, A., Paris, R. The B1080/B1192 molecular marker identifies hemp plants with functional THCA synthase and total THC content above legal limit (2023). DOI: 10.1016/j.gene.2023.147198

Appendices (Chapter 1)

Appendix A

Chemotype assessment

Table A1. Primer sequences used for chemotype assessment.

Primer Name	Sequence (5′->3′)
THCAS/CBDAS fw	AAGAAAGTTGGCTTGCAG
CBDAS rev	ATCCAGTTTAGATGCTTTTCGT
THCAS rev	TTAGGACTCGCATGATTAGTTTTTC

Table A2. PCR reaction set up for three-primers marker system.

Component	Final Concentration
Water	Up to 25 µl
10× PCR Buffer, -Mg	1×
50 mM MgCl2	1.5 mM
10 mM dNTP Mix	0.2 mM each
10 µM forward primer	0.2 µM
10 µM reverse primer 1	0.2 µM
10 µM reverse primer 2	0.2 µM
Template DNA	10 ng
Taq DNA Polymerase (5U/µL)	0.04 U/µL

Table A3**.** PCR program for three-primers marker system.

Step	Temperature	Time	Cycles
Initial denaturation	94 °C	3 min	1
Denaturation	94 °C	45 s	

Annealing	62 °C	1 min	32
Extension	72 °C	2 min	
Final extension	72 °C	10 min	1

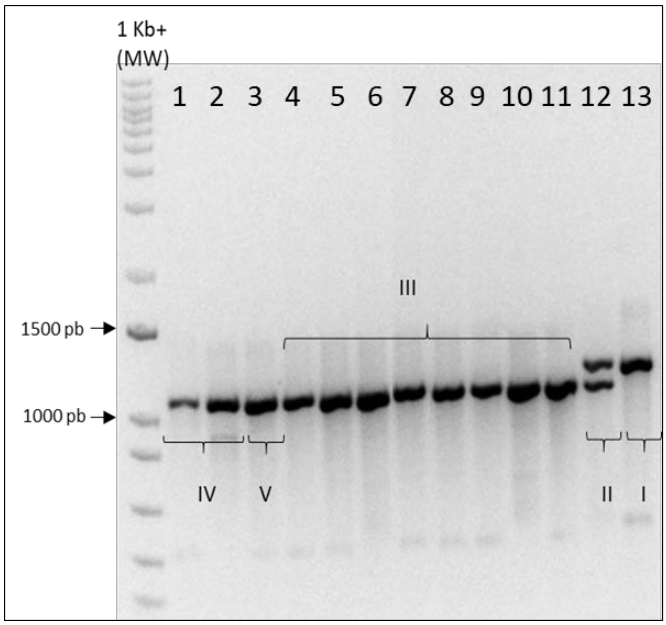


Figure A1. Chemotype determination. The respective chemotypes are indicated for each genotype (I–V). From left to right the genotypes are: 1, Santhica 27; 2, Bernabeo; 3, Ermo; 4, Fibranova; 5, Eletta Campana; 6, Futura 75; 7, Carmagnola; 8, Carmaleonte; 9, Codimono; 10, C.S.; 11, Fibrante; 12, CINRO; 13, CINBOL.

The chemotype assessment was performed by molecular markers described by Pacifico et al. (2006) in order to confirm the chemotypes of the 13 genotypes.

The amplification reaction, consisting of 10 ng of DNA and the recombinant Taq DNA Polymerase (Invitrogen), was performed in a T-Gradient Thermocycler (Biometra) in a final volume of 25 µL.

5 µL of PCR products were checked after adding 1 µL of 6× DNA loading dye (Thermo Fisher Scientific) by electrophoresis on 1% agarose gel (TAE buffer 1×), stained with Gel Red (Biotum). 1Kb plus DNA Ladder (Invitrogen) was used as marker for amplicon size. DNA was visualized under UV light on a transilluminator and digitally photographed with AlphaImager HP (Alpha Innotech).

Appendix B

Optimization of RT-qPCR primers for selective and quantitative amplification of target *cannabinoid synthase* genes.

Primer design and amplification reactions were optimized to distinguish the different cannabinoid synthases sequences, developing highly specific assays for gene detection and transcriptional analyses.

Gene-specific primer pairs were designed using the software Primer3. Since the different genes encoding for Cannabinoid Synthases share high similarity at the nucleotide level, primers were carefully designed in presence of SNPs differentiating genes, identified by sequences alignment, by keeping the SNPs at the 3' end of the primers. The primer pairs designed and employed in this work are shown in Table 3 of the main text.

Primer specificity was verified in different ways as described elsewhere (Pagliarani et al., 2013). First, primer pairs were searched by BlastN against *Cannabis sativa* non-redundant database at NCBI. Second, the presence of a single PCR product after end-point PCR and the absence of non-target amplification were assessed. The PCR reactions were performed in a 10 μL total volume, with 200 μM dNTPs, 0.2 U Phusion DNA Polymerase (Thermo Scientific) and different concentrations of the specific primer pairs. As a template, 0.5 pg of plasmid DNA, 1.5 μL of cDNA (diluted 1:10, corresponding to 2.5 ng starting RNA) and 10 ng DNA were used.

Different pJET1.2 plasmid vectors (CloneJET PCR Cloning Kit, Thermo Scientific) obtained as described in paragraph 2.5 of the main text, were used in this second step: one containing the specific gene targeted by the primer pair (positive control), and one or more plasmids containing other cannabinoids synthase gene sequences (negative controls).

PCR was performed following this condition: 98 °C for 30 s, 30 cycles of 10 s at 98 °C, 15 s at variable temperatures, and 15 s at 72 °C, followed by a final extension of 5 min at 72 °C. In order to optimize the reaction to get a specific product, different primers concentrations and annealing temperatures were tested until the best condition was found.

The reaction was considered optimized only when the amplification band was present in the positive control and absent in the negative ones (Figures A2–A4).

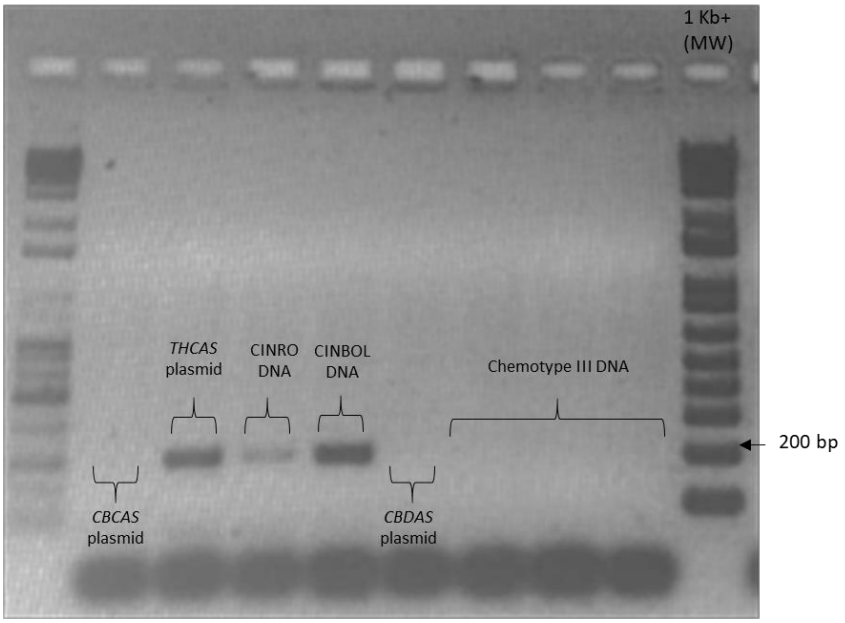


Figure A2. Specificity of primers for *THCAS*. The expected amplification band of 198 bp was obtained from pJET1.2 containing the *THCAS* and DNA of CINRO and CINBOL. Amplification was not obtained from pJET1.2 plasmid containing the *CBCAS*, and the *CBDAS*, and neither using DNA from hemp varieties as a template.

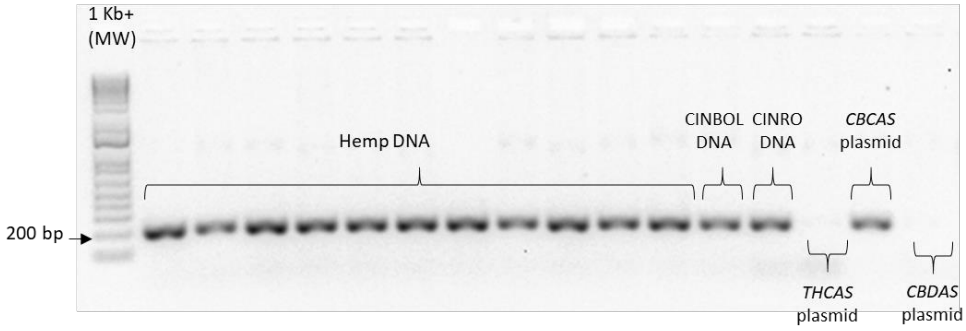


Figure A3. Specificity of primers for *CBCAS* genes. The expected amplification band of 198 bp was obtained from pJET1.2 containing the CBCAS and DNA of all tested genotypes. Amplification was not obtained from pJET1.2 plasmid containing the *THCAS* and the *CBDAS*, indicating a great specificity of these primer pairs.

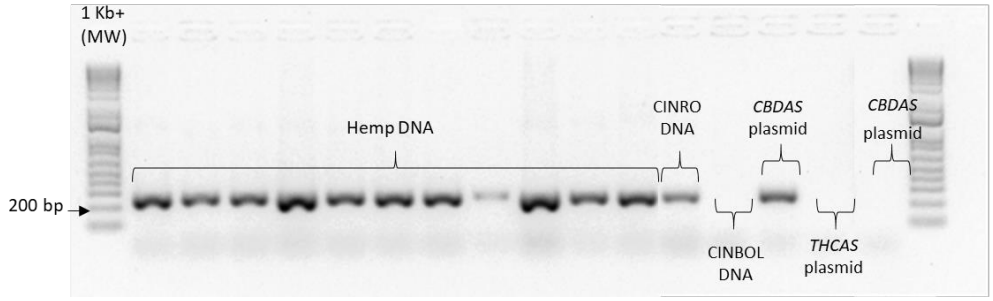


Figure A4. Specificity of primers for *CBDAS*. The expected amplification band of 241 bp was obtained from pJET1.2 containing the CBDAS and DNA from hemp and CINRO. Amplification was not obtained from pJET1.2 plasmid containing the *THCAS* and *CBCAS*. Primers work only on functional CBDAS sequences.

Primer pairs can be used as markers on DNA or to verify if the specific target gene is expressed using cDNA as template, as indicated in Figure A5.

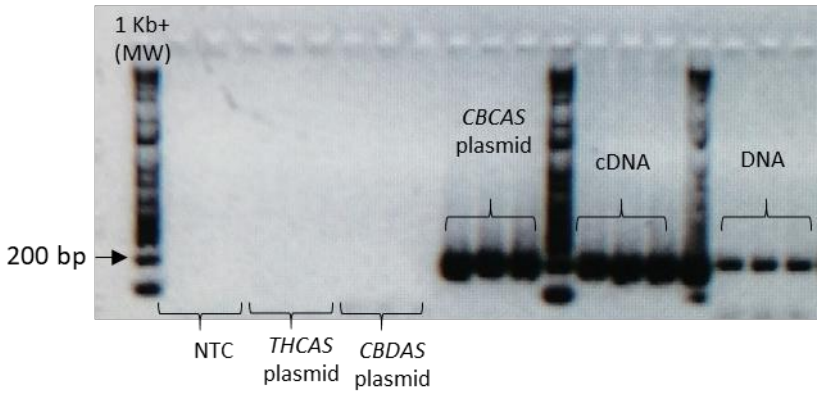


Figure A5. Agarose gel electrophoresis results of *CBCAS* primers amplification on plasmids containing each cannabinoid synthases, genomic DNA and cDNA of a hemp plant in triplicate. The PCR assays contained also 2 negative controls, where water was added instead of template (lane 1 and 2).

Finally, primers specificity was tested by RT-qPCR with a RotorGene 6000 (Corbett). In this specificity assay, two parameters were changed to optimize the amplification conditions: the annealing temperature (from 58 to 62 °C) and the primers concentration (from 75 nM to 1750 nM). Amplification reactions were conducted in a final volume of 10 µL containing the Power Up® SYBR master mix (Thermo Fisher Scientific), primer pair (different concentrations) and RNase-free water. As template, cDNA, DNA and plasmid DNA were used as indicated above.

Different amplification conditions were used depending on the annealing temperature of the primers used: when this was below 60 °C, a two-step method was used, consisting of 15 s at variable temperature for annealing and 1 min at 72 °C for extension; when this was above 60 °C, a one-step method was used, consisting of 60 s at the desired temperature in which both annealing and extension are performed. Finally, to ensure the absence of not-specific PCR products and primer dimers, a heat dissociation protocol (from 60 °C to 99° C) was performed and a dissociation curve for each sample was generated. Primers were considered as specific when they amplify a positive control and give no signal for negative controls (in both cases the controls are represented by plasmids with a known sequence). The production of a single melting curve peak in positive controls and samples suggested that the amplicon was specific (Figures A6-A8).

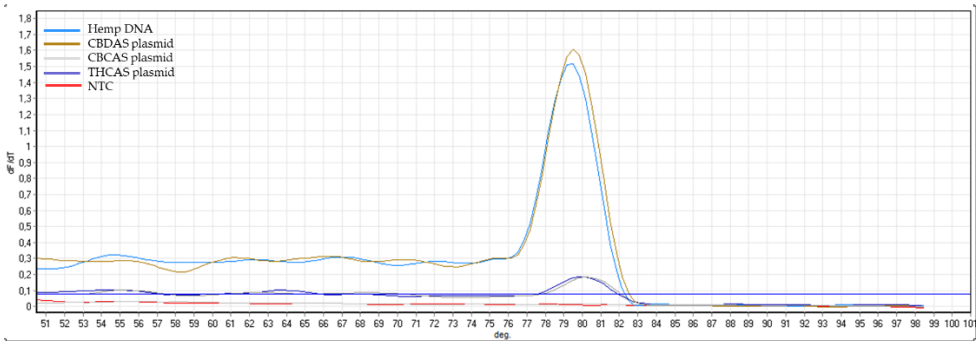


Figure A6. Melting curve plots, which display data collected during a melting curve stage with primers CBDAS_F and CBDAS_R. Peaks in the melting curve may indicate the melting temperature (Tm) of a target or identify nonspecific PCR amplification. No peaks were seen for the no template control (H_2O) and for negative controls (*THCAS* plasmid and *CBCAS* plasmid).

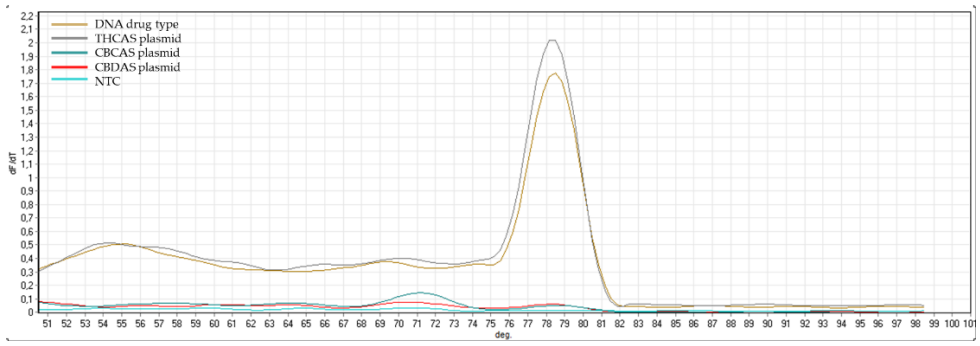


Figure A7. Melting curve plots, which display data collected during a melting curve Scheme 2. O) and for negative controls (*CBDAS* plasmid and *CBCAS* plasmid).

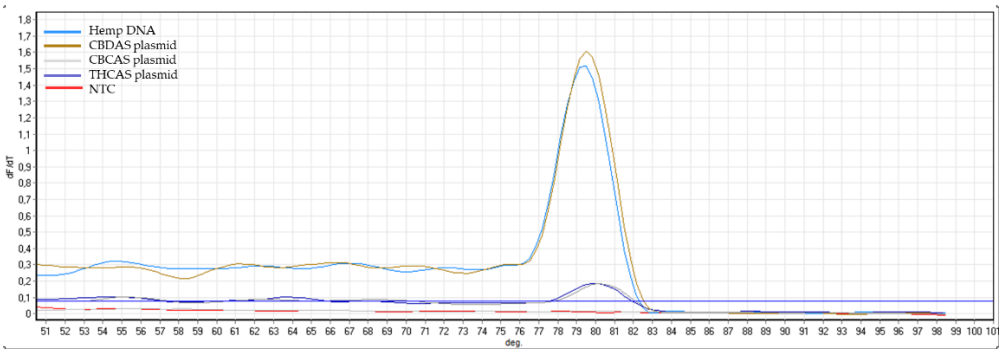


Figure A8. Melting curve plots, which display data collected during a melting curve stage with primers CBCAS_F and CBCAS_R. Peaks in the melting curve may indicate the melting temperature (Tm) of a target or identify nonspecific PCR amplification. No peaks were seen for the no template control (H_2O) and for negative controls (*THCAS* plasmid, *CBDAS* plasmid).

Supplementary Files (Chapter 1)

Supplementary File S1: nucleotide sequence alignment of *CBCAS* genes downloaded from sequenced genomes, AB212830 and a *CBCAS* sequenced from one of the genotypes under study.

CLUSTAL multiple sequence alignments by MUSCLE (3.8)

```
AB212830.1                          ATGAATTGCTCAGCATTTTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
NC_044378.1:25821957-25823591       ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
QKVJ02004887.1:13943-15577          ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
MNPR01002882.1:12890-14524          ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
MXBD01006695.1:30197-31831          ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
AGQN03005496.1:2986-4620            ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
QVPT02000161.1:1531540-1533174      ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
MW429551                            ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
                                    ************ **** ******************************************

AB212830.1                          TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
NC_044378.1:25821957-25823591       TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
QKVJ02004887.1:13943-15577          TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
MNPR01002882.1:12890-14524          TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
MXBD01006695.1:30197-31831          TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
AGQN03005496.1:2986-4620            TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
QVPT02000161.1:1531540-1533174      TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
MW429551                            TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
                                    ************************************************************

AB212830.1                          TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
NC_044378.1:25821957-25823591       TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
QKVJ02004887.1:13943-15577          TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
MNPR01002882.1:12890-14524          TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
MXBD01006695.1:30197-31831          TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
AGQN03005496.1:2986-4620            TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
QVPT02000161.1:1531540-1533174      TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
MW429551                            TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
                                    ************************************************************

AB212830.1                          ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
NC_044378.1:25821957-25823591       ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
QKVJ02004887.1:13943-15577          ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
MNPR01002882.1:12890-14524          ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
MXBD01006695.1:30197-31831          ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
AGQN03005496.1:2986-4620            ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
QVPT02000161.1:1531540-1533174      ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
MW429551                            ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
                                    ************************************************************

AB212830.1                          CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
NC_044378.1:25821957-25823591       CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
QKVJ02004887.1:13943-15577          CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
MNPR01002882.1:12890-14524          CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
MXBD01006695.1:30197-31831          CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AGQN03005496.1:2986-4620            CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
QVPT02000161.1:1531540-1533174      CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
MW429551                            CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
                                    ************************************************************

AB212830.1                          AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
NC_044378.1:25821957-25823591       AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
QKVJ02004887.1:13943-15577          AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
MNPR01002882.1:12890-14524          AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
MXBD01006695.1:30197-31831          AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
AGQN03005496.1:2986-4620            AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
QVPT02000161.1:1531540-1533174      AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
MW429551                            AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
                                    ************************************************************

AB212830.1                          TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
NC_044378.1:25821957-25823591       TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
QKVJ02004887.1:13943-15577          TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
MNPR01002882.1:12890-14524          TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
```

```
MXBD01006695.1:30197-31831          TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
AGQN03005496.1:2986-4620            TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
QVPT02000161.1:1531540-1533174      TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
MW429551                            TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
                                    ************************************************************

AB212830.1                          GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
NC_044378.1:25821957-25823591       GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
QKVJ02004887.1:13943-15577          GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
MNPR01002882.1:12890-14524          GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
MXBD01006695.1:30197-31831          GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
AGQN03005496.1:2986-4620            GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
QVPT02000161.1:1531540-1533174      GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
MW429551                            GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
                                    ************************************************************

AB212830.1                          TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
NC_044378.1:25821957-25823591       TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
QKVJ02004887.1:13943-15577          TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
MNPR01002882.1:12890-14524          TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
MXBD01006695.1:30197-31831          TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
AGQN03005496.1:2986-4620            TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
QVPT02000161.1:1531540-1533174      TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
MW429551                            TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
                                    ************************************************************

AB212830.1                          GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
NC_044378.1:25821957-25823591       GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
QKVJ02004887.1:13943-15577          GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
MNPR01002882.1:12890-14524          GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
MXBD01006695.1:30197-31831          GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
AGQN03005496.1:2986-4620            GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
QVPT02000161.1:1531540-1533174      GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
MW429551                            GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
                                    ************************************************************

AB212830.1                          GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
NC_044378.1:25821957-25823591       GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
QKVJ02004887.1:13943-15577          GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
MNPR01002882.1:12890-14524          GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
MXBD01006695.1:30197-31831          GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
AGQN03005496.1:2986-4620            GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
QVPT02000161.1:1531540-1533174      GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
MW429551                            GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
                                    ************************************************************

AB212830.1                          TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
NC_044378.1:25821957-25823591       TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
QKVJ02004887.1:13943-15577          TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
MNPR01002882.1:12890-14524          TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
MXBD01006695.1:30197-31831          TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
AGQN03005496.1:2986-4620            TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
QVPT02000161.1:1531540-1533174      TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
MW429551                            TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
                                    ************************************************************

AB212830.1                          ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
NC_044378.1:25821957-25823591       ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
QKVJ02004887.1:13943-15577          ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
MNPR01002882.1:12890-14524          ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
MXBD01006695.1:30197-31831          ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AGQN03005496.1:2986-4620            ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
QVPT02000161.1:1531540-1533174      ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
MW429551                            ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
                                    ************************************************************

AB212830.1                          AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
NC_044378.1:25821957-25823591       AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
QKVJ02004887.1:13943-15577          AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
MNPR01002882.1:12890-14524          AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
MXBD01006695.1:30197-31831          AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
AGQN03005496.1:2986-4620            AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
QVPT02000161.1:1531540-1533174      AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
MW429551                            AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
                                    ************************************************************
```

```
AB212830.1                          TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
NC_044378.1:25821957-25823591       TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
QKVJ02004887.1:13943-15577          TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
MNPR01002882.1:12890-14524          TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
MXBD01006695.1:30197-31831          TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AGQN03005496.1:2986-4620            TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
QVPT02000161.1:1531540-1533174      TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
MW429551                            TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
                                    ************************************************************

AB212830.1                          AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
NC_044378.1:25821957-25823591       AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
QKVJ02004887.1:13943-15577          AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
MNPR01002882.1:12890-14524          AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
MXBD01006695.1:30197-31831          AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
AGQN03005496.1:2986-4620            AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
QVPT02000161.1:1531540-1533174      AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
MW429551                            AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
                                    ************************************************************

AB212830.1                          GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
NC_044378.1:25821957-25823591       GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
QKVJ02004887.1:13943-15577          GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
MNPR01002882.1:12890-14524          GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
MXBD01006695.1:30197-31831          GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
AGQN03005496.1:2986-4620            GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
QVPT02000161.1:1531540-1533174      GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
MW429551                            GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
                                    ************************************************************

AB212830.1                          GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
NC_044378.1:25821957-25823591       GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
QKVJ02004887.1:13943-15577          GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
MNPR01002882.1:12890-14524          GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
MXBD01006695.1:30197-31831          GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AGQN03005496.1:2986-4620            GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
QVPT02000161.1:1531540-1533174      GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
MW429551                            GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
                                    ************************************************************

AB212830.1                          AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
NC_044378.1:25821957-25823591       AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
QKVJ02004887.1:13943-15577          AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
MNPR01002882.1:12890-14524          AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
MXBD01006695.1:30197-31831          AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
AGQN03005496.1:2986-4620            AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
QVPT02000161.1:1531540-1533174      AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
MW429551                            AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
                                    ************************************************************

AB212830.1                          TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
NC_044378.1:25821957-25823591       TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
QKVJ02004887.1:13943-15577          TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
MNPR01002882.1:12890-14524          TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
MXBD01006695.1:30197-31831          TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
AGQN03005496.1:2986-4620            TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
QVPT02000161.1:1531540-1533174      TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
MW429551                            TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
                                    ************************************************************

AB212830.1                          TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
NC_044378.1:25821957-25823591       TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGAT
QKVJ02004887.1:13943-15577          TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGAT
MNPR01002882.1:12890-14524          TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
MXBD01006695.1:30197-31831          TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
AGQN03005496.1:2986-4620            TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
QVPT02000161.1:1531540-1533174      TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
MW429551                            TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
                                    ************************************************************ *

AB212830.1                          GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
NC 044378.1:25821957-25823591       GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
QKVJ02004887.1:13943-15577          GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
MNPR01002882.1:12890-14524          GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
MXBD01006695.1:30197-31831          GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
AGQN03005496.1:2986-4620            GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
```

```
QVPT02000161.1:1531540-1533174      GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
MW429551                            GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
                                    ************************************************************

AB212830.1                          TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
NC_044378.1:25821957-25823591       TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
QKVJ02004887.1:13943-15577          TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
MNPR01002882.1:12890-14524          TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
MXBD01006695.1:30197-31831          TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
AGQN03005496.1:2986-4620            TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
QVPT02000161.1:1531540-1533174      TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
MW429551                            TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
                                    ************************************************************

AB212830.1                          TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
NC 044378.1:25821957-25823591       TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTGTGTGTCCCAAAATCCAAGATTGGCG
QKVJ02004887.1:13943-15577          TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTGTGTGTCCCAAAATCCAAGATTGGCG
MNPR01002882.1:12890-14524          TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
MXBD01006695.1:30197-31831          TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
AGQN03005496.1:2986-4620            TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
QVPT02000161.1:1531540-1533174      TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
MW429551                            TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
                                    ********************************* ************************

AB212830.1                          TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
NC_044378.1:25821957-25823591       TATCTCAATTATAGGGACCTTGATTTAGGAAAATCTAATCCTGAGAGTCCTAATAATTAC
QKVJ02004887.1:13943-15577          TATCTCAATTATAGGGACCTTGATTTAGGAAAATCTAATCCTGAGAGTCCTAATAATTAC
MNPR01002882.1:12890-14524          TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
MXBD01006695.1:30197-31831          TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
AGQN03005496.1:2986-4620            TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
QVPT02000161.1:1531540-1533174      TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
MW429551                            TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
                                    ********************************* ************************

AB212830.1                          ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
NC_044378.1:25821957-25823591       ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
QKVJ02004887.1:13943-15577          ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
MNPR01002882.1:12890-14524          ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
MXBD01006695.1:30197-31831          ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
AGQN03005496.1:2986-4620            ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
QVPT02000161.1:1531540-1533174      ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
MW429551                            ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
                                    ************************************************************

AB212830.1                          GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
NC_044378.1:25821957-25823591       GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
QKVJ02004887.1:13943-15577          GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
MNPR01002882.1:12890-14524          GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
MXBD01006695.1:30197-31831          GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
AGQN03005496.1:2986-4620            GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
QVPT02000161.1:1531540-1533174      GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
MW429551                            GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
                                    ************************************************************

AB212830.1                          CCACCGCATCATCAT
NC_044378.1:25821957-25823591       CCACCGCGTCATCAT
QKVJ02004887.1:13943-15577          CCACCGCGTCATCAT
MNPR01002882.1:12890-14524          CCACCGCGTCATCAT
MXBD01006695.1:30197-31831          CCACCGCGTCATCAT
AGQN03005496.1:2986-4620            CCACCGCGTCATCAT
QVPT02000161.1:1531540-1533174      CCACCGCGTCATCAT
MW429551                            CCACCGCGTCATCAT
                                    ******* *******
```

Input sequences:
((5'3') indicates that the downloaded sequences have been converted into their respective inverse complement to perform the alignment)

```
>MNPR01002882.1:12890-14524 Cannabis sativa cultivar Cannatonic
Cannabis.v1_scf2882_q, whole genome shotgun sequence
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
```

ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>MXBD01006695.1:30197-31831 Cannabis sativa cultivar Pineapple Banana Bubba Kush
006705Cannabis_1597M_SHL, whole genome shotgun sequence (5'3')
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>AGQN03005496.1:2986-4620 Cannabis sativa cultivar Purple Kush 005500F, whole
genome shotgun sequence (5'3')
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT

ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>QVPT02000161.1:1531540-1533174 Cannabis sativa cultivar Jamaican Lion DASH
000163F_arrow, whole genome shotgun sequence (5'3')
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>NC_044378.1:25821957-25823591 Cannabis sativa chromosome 7, cs10
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA

CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGAT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTGTGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAATCTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>QKVJ02004887.1:13943-15577 Cannabis sativa cultivar Finola 004887F, whole genome shotgun sequence (5'3')
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGAT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTGTGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAATCTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT
>AB212830.1 Cannabis sativa gene for tetrahydrocannabinolic acid synthase, partial cds
ATGAATTGCTCAGCATTTTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA

CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCATCATCAT
>MW429551
ATGAATTGCTCAACATTCTCCTTTTGGTTTGTTTGCAAAATAATATTTTTCTTTCTCTCA
TTCAATATCCAAATTTCAATAGCTAATCCTCAAGAAAACTTCCTTAAATGCTTCTCGGAA
TATATTCCTAACAATCCAGCAAATCCAAAATTCATATACACTCAACACGACCAATTGTAT
ATGTCTGTCCTGAATTCGACAATACAAAATCTTAGATTCACCTCTGATACAACCCCAAAA
CCACTCGTTATTGTCACTCCTTCAAATGTCTCCCATATCCAGGCCAGTATTCTCTGCTCC
AAGAAAGTTGGTTTGCAGATTCGAACTCGAAGCGGTGGCCATGATGCTGAGGGTTTGTCC
TACATATCTCAAGTCCCATTTGCTATAGTAGACTTGAGAAACATGCATACGGTCAAAGTA
GATATTCATAGCCAAACTGCGTGGGTTGAAGCCGGAGCTACCCTTGGAGAAGTTTATTAT
TGGATCAATGAGATGAATGAGAATTTTAGTTTTCCTGGTGGGTATTGCCCTACTGTTGGC
GTAGGTGGACACTTTAGTGGAGGAGGCTATGGAGCATTGATGCGAAATTATGGCCTTGCG
GCTGATAATATCATTGATGCACACTTAGTCAATGTTGATGGAAAAGTTCTAGATCGAAAA
TCCATGGGAGAAGATCTATTTTGGGCTATACGTGGTGGAGGAGGAGAAAACTTTGGAATC
ATTGCAGCATGGAAAATCAAACTTGTTGTTGTCCCATCAAAGGCTACTATATTCAGTGTT
AAAAAGAACATGGAGATACATGGGCTTGTCAAGTTATTTAACAAATGGCAAAATATTGCT
TACAAGTATGACAAAGATTTAATGCTCACGACTCACTTCAGAACTAGGAATATTACAGAT
AATCATGGGAAGAATAAGACTACAGTACATGGTTACTTCTCTTCCATTTTTCTTGGTGGA
GTGGATAGTCTAGTTGACTTGATGAACAAGAGCTTTCCTGAGTTGGGTATTAAAAAAACT
GATTGCAAAGAATTGAGCTGGATTGATACAACCATCTTCTACAGTGGTGTTGTAAATTAC
AACACTGCTAATTTTAAAAAGGAAATTTTGCTTGATAGATCAGCTGGGAAGAAGACGGCT
TTCTCAATTAAGTTAGACTATGTTAAGAAACTAATACCTGAAACTGCAATGGTCAAAATT
TTGGAAAAATTATATGAAGAAGAGGTAGGAGTTGGGATGTATGTGTTGTACCCTTACGGT
GGTATAATGGATGAGATTTCAGAATCAGCAATTCCATTCCCTCATCGAGCTGGAATAATG
TATGAACTTTGGTACACTGCTACCTGGGAGAAGCAAGAAGATAACGAAAAGCATATAAAC
TGGGTTCGAAGTGTTTATAATTTCACAACTCCTTATGTGTCCCAAAATCCAAGATTGGCG
TATCTCAATTATAGGGACCTTGATTTAGGAAAAACTAATCCTGAGAGTCCTAATAATTAC
ACACAAGCACGTATTTGGGGTGAAAAGTATTTTGGTAAAAATTTTAACAGGTTAGTTAAG
GTGAAAACCAAAGCTGATCCCAATAATTTTTTTAGAAACGAACAAAGTATCCCACCTCTT
CCACCGCGTCATCAT

Supplementary Table S. Summary of features of the *Cannabis sativa* genomes accessed.

Accession number	Cultivar	Genome representation	Assembly level
GCA_003417725.2	Finola	Full	Chromosome
GCA_000230575.5	Purple Kush	Full	Chromosome
GCA_003660325.2	Jamaican Lion DASH	Full	Contig
GCA_002090435.1	Pineapple Banana Bubba Kush	Full	Contig
GCA_001865755.1	Cannatonic	Full	Contig
GCA_001510005.1	LA Confidential	Full	Contig
GCA_001509995.1	Chemdog91	Full	Scaffold
GCA_900626175.2	cs10/CBDRx	Full	Chromosome

Supplementary Table S2. Sequence variation within *Cannabis* germplasm.

Genotype	**No. of *CBCAS cds***	**Accession code**	**No. of *THCAS cds***	**Accession code**	**No. of *CBDAS cds***	**Accession code**
C.S.	6	MW429517, MW429528, MW429536, MW429537, MW429538, MW429539	0	.	2	MW429549, MW429550,
Carmagnola	5	MW561075, MW429518, MW429519, MW429520, MW429521	0	.	1	MW429550
Carmaleonte	5	MW429515, MW429516, MW429517, MW429551, MW561074	0	.	4	MW429541, MW429542, MW429549, MW429550
Fibrante	5	MW429528, MW429529, MW429530, MW429551, MW561073	0	.	2	MW429548, MW429550
Fibranova	3	MW429518, MW429531, MW429532	0	.	3	MW429550, MW429546, MW429547
Futura 75	4	MW429517, MW429525, MW429526, MW429527	0	.	1	MW429550
Santhica 27	3	MW429551, MW429523, MW429524	0	.	1	MW429549
Eletta Campana	2	MW429534, MW429535	0	.	2	MW429549, MW429550
Codimono	1	MW429540	0	.	1	MW429550
Bernabeo	2	MW429551, MW429522	0	.	1	MW429549
Ermo	2	MW429551, MW429533	0	.	5	MW429549, MW429550, MW429543, MW429544, MW429545
CINBOL	1	MW429551	3	E33090, MW429552, MW429553	0	.
CINRO	1	MW429551	1	E33090	1	MW429550

Supplementary Table S3. SNPs found in the *CBCAS* sequences found in *Cannabis* germplasm

Accession	13	18	41	45	46	63	71	73	79	111	138	168	198	254	257	287	300	410	437	440	441	600	856	970	1074	1110	1124	1127	1138	1162	1170	1258	1372	1410	1422	1424	1438	1518	1560	1581	1582	1628
AB212830	G	T	T	A	T	C	A	A	A	C	A	C	G	T	C	G	C	C	C	C	G	G	G	C	A	G	C	T	C	A	G	G	G	C	A	A
MW429515	A	C						T																						G							G
MW429516	A	C			A															T																	G
MW429517	A	C																					C														G
MW429518	A	C																			A						T										G
MW429519	A	C											A									C															G
MW429520	A	C								T						A														T	T						G
MW429521	A	C		G													T																				G
MW429522	A	C						T																														G
MW429523	A	C							T															T		T								T				G
MW429524	A	C							T															T		T								G				G
MW429525	A	C								T																				T	T						G
MW429526	A	C									T																	A										G
MW429527	A	C		G										C			T																					G
MW429528	A	C																T																				G
MW429529	A	C																																			C	G
MW429530	A	C							T																				A					G				G
MW429531	A	C																	T																		T		G
MW429532	A	C																							C													A		G
MW429533	A	C						T															C																G
MW429534	A	C														A																							G
MW429535	A	C						T										G															A						G
MW429536	A	C				G																										T		G					G
MW429537	A	C										T			T																								G
MW429538	A	C		G													T																				T			G
MW429539	A	C					G																C																G
MW429540	A	C	C																															G					G
MW561075	A	C											A																										G
MW561074	A	C							T																									G					G
MW561073	A	C			A																																		G
MW429551	A	C																																					G
AA position	5		14	15	16	21	24	25	27					85	86	96		137	146	147			286			370	375	376	380	388	390	420	458			475	480	506			528	543
	A		I	I	F	F	Q	I	I					V	T	S		T	T	A			D			L	A	G	A	V	K	G	H			Q	A	W			N	H
Substitution	↓		↓	↓	↓	↓	↓	↓	↓					↓	↓	↓		↓	↓	↓			↓			↓	↓	↓	↓	↓	↓	↓	↓			↓	↓	↓			↓	↓
	T		T	M	I	L	R	F	L					A	I	N		M/R	I	V			H			F	V	E	T	F	N	S	Y			R	S	C			H	R

Supplementary Table S4. SNPs found in the *CBDAS* sequences in *Cannabis* germplasm

Accession code	105	348	420	681	770	953	958	1036	1426	1559	1621
E55107	G	T	A	T	T	G	G	T	C	T	C
MW429541	T	G									
MW429542	T								T	C	
MW429543	T			C	C				T		
MW429544	T							C	T		
MW429545	T					T			T		
MW429546	T		T								
MW429547	T										G
MW429548	T						A				
MW429549	T								T		
MW429550	T										
AA position					257	318	320	346	476	520	541
Substitution					F ↓ S	G ↓ V	V ↓ M	W ↓ R	P ↓ S	V ↓ A	R ↓ G

Supplementary Files (Chapter 2)

Supplementary Table 1. Primer sequences and primer features

Primer name	Primer sequence	T ann	Eff	R^2	Amplicon size (bp)	Concentration
AAE3	Fw:CCTCTACACCTCCCCAACAA; Rv:AGAAAGCCATGGGAGACCTT	60	1.31	1	164	0.3 μM
OLS	Fw:CACGAGATGCAAACTCTGGA; Rv:ACACGCTTCACTGAGGGACT	60	0.91	0.91	215	0.3 μM
OAC	Fw:CACAGAAGCCCAAAAGGAAG; Rv:CAACATGGGCAGGATGAATA	60	1.11	1	194	0.3 μM
PT	Fw:AACAAATTACGACCCGGAAG; Rv:GGTAAGATCCACCACTAAATTATACGA	60	1.4	0.99	167	0.3 μM
AAE1	Fw:TGTCATTGTTTGGCCCACTA; Rv:AGGGACCACACCTAGCATTG	60	1.09	0.99	166	0.3 μM
THCAS	Fw:AAACTTCCTTAAATGCTTCTCAA; Rv:TAAAATAGTTGCTTGGATATGGGAGTT	58	1.01	0.98	198	0.3 μM
CBCAS	Fw:ATTTTGGGCTATACGTGGTGGA; Rv:CCTTTGATGGGACAACAACA	61	1	1	86	0.3 μM
CBDAS	Fw:GCAATACACACTTACTTCTCTTCAGTTTTC; Rv:ACGTAGTCTAACTTAATCTTGAAAGCAC	61.5	1.15	0.99	241	0.3 μM