UNIVERSITÀ DEGLI STUDI DI VERONA

Department of Biotechnology

XXXVI Cycle

2020-2023

PhD in Biotechnology

Overcoming challenges in *Cannabis sativa* breeding research with conventional and biotechnological tools.

With funding by:

Gruppo Padana

Coordinator: Prof. Matteo Ballottari

Tutor: Prof. Linda Avesani

Co-Tutors: Prof. Flavia Guzzo, Gianni Barcaccia

PhD Candidate: Marcello Borin

Table of Contents

List of abbreviations

CBD - Cannabinoid

CBC - Cannabichromenic Acid

CBG - Cannabigerolic Acid

THC - Tetrahydrocannabinol

THCa - Tetrahydrocannabinolic acid

SSR - Simple Sequence Repeats, also called microsatellites.

STS - Silver ThioSulfate

PCR - Polymerase Chain Reaction

cDNA - complementary DNA (retro transcribed RNA)

gDNA - Genomic DNA

FF - Female Flower

MF - Male Flower

IMF - Induced Male Flower

UPLC-HRMS - Ultra Performance Liquid Chromatography – High-Resolution Mass Spectrometry.

qRT-PCR - quantitative Reverse Transcription - Polymerase Chain Reaction

ERG - ethylene-related genes

Abstract

Cannabis sativa is a popular species known for its use in medical, recreational, and industrial purposes, but genetic and biotechnological research on this plant is still lacking due to historical and ongoing prohibition across the globe.

The general goal of the research conducted during my PhD is to overcome challenges in *C. sativa* breeding research integrating conventional and new biotechnological methods.

Firstly, we focused on deciphering the geographical origin of cannabis varieties, using SSR genomic markers and by the sequencing chloroplast markers.

This data helped us to statistically discriminate different cannabis varieties and identify maternal genetic backgrounds. These results will contribute to authentication, and improvement of breeding programs for *C. sativa*.

Furthermore, we investigated flower development and sex determination in *C. sativa* by transcriptomic and metabolomic analysis. Flower development experiments were conducted, and sex reversion protocols were tested by inducing the production of pollen in genetically female plants with chemical treatments. Together, these data will provide important tools not only to the scientific community but also to producers and consumers for the future of cannabis breeding.

Aim and outline of the PhD thesis

My PhD program is focused on the development of new tools for the characterization of *C. sativa*. Firstly, regarding genetic analysis we developed SSR markers, able to determine genetic distances and correlations between samples, the markers enable us to cluster individuals on the basis of geographical origin and were also used to access homozygosity, genetic uniformity, and genetic variation intra- and inter- variety. Additionally, we developed a method for the identification and sequencing of maternally transmitted markers (chloroplast markers) that proved to be useful in calculating genetic distances between varieties and assigning them to phylogenetic clusters.

Next, we developed quantitative RT-PCR analysis for specific transcriptomic analysis of genes involved in the cannabinoid pathway. The analysis was performed on flowering samples, together with a metabolomic analysis with a Liquid Chromatography coupled with mass spectrometry instrument. This analysis allowed us to study cannabinoid accumulation during flower development and to correlate it with a specific transcriptomic analysis.

Finally, we were also interested in understanding sex determination in cannabis as only female flowers are required for the industry and the presence of hermaphroditic behavior, which is possible in stress conditions, may ruin productivity and breeding experiments.

To understand sex determination, we performed chemical treatment experiments to induce the production of male flowers in female plants and then we perform a targeted transcriptome analysis in treated samples (that are genetically female plants with male flowers) and in normal female clones, to identify which gene transcriptions were altered, for shedding a light on genes responsible for sex switch, a topic highly debated in literature.

This research was performed in partial fulfillment of the Ph.D. Program of Marcello Borin by taking advantage of the Doctoral Research Fellowship funded by Gruppo Padana (Gruppo Padana Ortofloricoltura S.S., Via Olimpia 41, 31038 Paese, Treviso, Italy).

All genetic analysis were performed in collaboration with the University of Padova (DAFNAE: Dipartimento di Agronomia Animali Alimenti Risorse Naturali e Ambiente)

where the SSR analysis and cDNA sequencing experiments were conducted in the laboratory of Prof. Gianni Barcaccia.

Metabolomic and Transcriptomic analysis on flowering hemp samples was conducted at the University of Verona in the laboratory of Prof. Linda Avesani.

Finally, a large portion of activities during my second PhD year were conducted in Canada at University of Saskatchewan, Saskatoon, (SK) at the laboratory of Prof. Timothy Sharbel with the following project: Genetic, phenotypic, and biochemical analyses of worldwide accessions of Cannabis.

CHAPTER 1: Introduction

1.1 General Introduction to *Cannabis sativa*

Cannabis sativa L. is an agricultural plant species that today enjoys great interest because of its multiple uses in the recreational, medicinal, and industrial areas [1]. This plant can be cultivated to produce fibers (used to make different textiles), seeds (rich in unsaturated fatty acids for edible oils), and drugs from its female inflorescences that contain cannabinoids (compounds with psychotropic and psychopharmaceutical effects).

Among these latter, the principal psychoactive constituent of cannabis is THC (tetrahydrocannabinol), and the concentration of this metabolite is at the basis of the distinction between hemp and drug (marijuana) types, with hemp considered low in concentration, 0.3% in dry weight or less THC (non-psychoactive), and marijuana, on the other hand, containing up to 30% THC by dry weight.

The genus Cannabis belongs to the family of Cannabaceae (order Rosales). Its botanical classification had a very troubled genesis since the times of Linnaeus considering it was not clear whether the genus was mono- or polytypic [2–4]. In 1597, John Gerarde [5] first defined the plant species as dioecious, but the question remained open because monoecious plants can occur and hermaphroditism is also possible with plants that show both reproductive organs within the same flower [3,6,7]. All these biological variants are known to be very frequent in Cannabis fiber varieties [3]. Plants also manifest sexual dimorphism, with male individuals being often characterized by a shorter crop cycle and a taller stature than female ones. Lamarck originally recognized two interfertile species *C. sativa* (from Persia) and *C. indica* (from India) [8]. Based on this old taxonomy, many varieties available on the market are still classified as *C. sativa* × *C. indica* hybrids. As a matter of fact, the reproductive system of Cannabis plants is characterized by allogamy and anemophily, and therefore open pollination is necessarily responsible for a certain degree of hybridization between improved

and wild populations. This is why, according to Schultes, landraces of Cannabis should no longer exist since several decades [2]. Later on, Small and Cronquist [3] proposed a unique species system that is still widely accepted and that is based on two subspecies of *C. sativa*: *C. sativa* subsp. *sativa* and *C. sativa* subsp. *indica*. Although several authors, supporting the one species system for Cannabis, recommend classifying its varieties based on the cannabinoids and terpenoids profile [9,10], a molecular system based on DNA barcoding could represent a cost- and time-effective technique of great help in clarifying some of the taxonomic issues related to the genus Cannabis. DNA barcoding could also play a crucial role in the identification and characterization of those certified Cannabis strains, which are mainly derived from black market. According to Charlesworth et al. (2005) [11] the dioecious species evolved from a common monoecious ancestor shared by Cannabis and Humulus [1] both characterized by having sex chromosomes [12]. In particular, *C. sativa* possesses nine pairs of autosomes and a pair of X and Y sex chromosomes. The male sex is heterogametic (XY), while the female is homogametic (XX), and other authors reported distinct mechanisms involved in the determination of sex [13,14]. This uncertainty could derive from the fact that environmental conditions, and in particular abiotic stress factors, can influence the expression and the determination of sex [15]. Although the structure of sex chromosomes is poorly understood in Cannabis spp., since it is not detectable with standard microscopic techniques [13,16], the Y chromosome was shown to have larger dimensions than the X chromosome [13,17]. More recently, both male and female karyotypes of *C. sativa* L. were extensively characterized by DAPI banding procedures and FISH analyses using rDNA probes [18].

1.2 Cannabis legislation

1.2.1 History of Cannabis criminalization

After millennia of human engagement with the Cannabis plant control and regulation is a relatively recent notion, with the first instances around 200 years ago.

The first recorded one is from the short-lived banning of Cannabis in Egypt in 1800 by Napoleon, other localized prohibitions in the following years included bans in Rio de Janeiro in Brazil in 1830, in Egypt in 1868, in South Africa in 1870 and Greece in 1890; these early efforts were often in the context of the pursuit of colonial interests or for the social control of minority groups [20].

Cannabis was then included in the 1925 Geneva Opium Convention, this introduced relatively limited controls on transnational trade and did not require signatories to prohibit or restrict domestic production and consumption but marked a turning point as from that moment countries began to introduce national-level Cannabis controls [21].

For example, Britain introduced Cannabis in the Dangerous Drugs Act (1928) and federal prohibition in the USA eventually come in 1937 with the enactment of the Marihuana Tax Act. The United Nations Single Convention on Narcotic Drugs in 1961 finally consolidated and extended the global prohibition of Cannabis [22].

1.2.2 Modern Cannabis regulation

In recent years legalization and decriminalization initiatives are rapidly changing Cannabis legislation around the world, fully legal infrastructure for the cultivation, distribution and sale of Cannabis have occurred in Uruguay in 2013, Canada in 2018 and in various US states where legal Cannabis supply systems are present in contrast with federal prohibition.

Regulation statuses in the continents are presented in the following sections where the legislation situation in America, Europe, Oceania, Africa, and Asia will be briefly summarized [23].

- **America:** Within the Americas the most radical reforms to Cannabis laws have been taking place, with Uruguay as the first country in the world to legalize Cannabis (2013) and Canada the first in the G7 (2018) and numerous US states following suit. Notable other law reform initiatives in central/south America includes:

- Decriminalization of possession: Perù (1991), Colombia (1994), Argentina (2009), Chile (2005), Mexico (2009)
- Personal-use cultivation: Chile (2005),
- Licenses for cultivation: Colombia (2016), Perù (2019)
- Medical Cannabis legal on prescription: Chile (2015), Colombia (2016), Perù (2017) Brazil (2020)

❖ **Europe:** Significant changes to Cannabis law were first introduced in the Netherlands, where, since 1976, Cannabis use and possession was decriminalized; following this in the 1980s coffeeshops were established and tolerated with the contradiction that the Cannabis sold must be sourced from the illicit market, with some having been grown domestically and some imported. To resolve the issue, since 2019 a legal cultivation experiment is taking place, meanwhile Cannabis has been legal for medical use since 2003 in the Netherlands. Spain has also been at the forefront of legalization efforts with collective cultivation and possession accepted since 2003 [24]. In Belgium, while technically illegal, the cultivation of one Cannabis plant per person is tolerated since 2006.

Finally, in Portugal all drugs possession was decriminalized in 2001 and medical Cannabis use was legalized in 2018. Several other jurisdictions across Europe are taking steps for decriminalization and legalization, some examples are:

- Decriminalization of possession: Germany (1994);
- Personal-use cultivation: Italy (Supreme Court sentence, 2019), Germany (2024);
- Licenses for cultivation: Denmark (2018), Greece (2018);
- Medical Cannabis legal on prescription: Italy (2006, expansion in 2013), Austria (2008), Finland (2014), Germany (2017), Greece (2018), United Kingdom (2018), Ireland (2019).

❖ **Oceania:** In Oceania steps have been taken by both New Zealand and Australia; regarding New Zealand medical Cannabis has been legal since 2018 and, in 2020

launched its Medical Cannabis Scheme to fully regulate the medical Cannabis supply chain. In Australia medical Cannabis has been legal in numerous states and territories since April 2016 with great differences between jurisdictions.

- **Africa:** Regarding the African countries, access to medical Cannabis is now available in Zimbabwe (2018), South Africa (2018), Zambia (2019) Ghana (2020), and Malawi (2020). Although medical access has been approved, establishing a legalized and regulated supply has not yet been achieved.

- **Asia:** In Asia, medical Cannabis is now legal in Israel (1995), Lebanon (2020), Sri Lanka (2018) and Thailand (2019). Israel has been issuing licenses for medical Cannabis cultivation to individuals since the 1990s [25].

1.3 History of Cultivated Varieties

Traditionally, hemp-type and drug-type varieties have been bred mainly through mass selection. This method has been effectively used for the selection of Cannabis showing improved quality traits such as fiber, oil, and cannabinoid content [19]. Nevertheless, one of the main problems associated with the first attempts of cannabis genetic improvement was, on the one hand, the need to avoid hemp genotypes with high THC contents, on the other hand, the availability of uniform medical genotypes, which was often linked to clandestine growers. More recently, Cannabis cultivars were obtained from controlled mating using selected individuals from different landraces and cultivars. Usually, several selected individuals were used for open pollination so that each of the female plants could be fertilized by each of the male plants (i.e., intercrosses). Synthetic varieties were also obtained by open pollination using many female and male plants vegetatively propagated via cuttings (i.e., polycrosses). Heterosis (or hybrid vigor) has been a driving factor for breeding programs aimed at the development of both modern fiber- and drug-type cultivars. The heterotic effect is usually manifested by highly heterozygous plants produced by crossing two different

lineages and/or antagonist genotypes (i.e., using parental lines that show high homozygosity for antagonist gene forms across most of the loci). The first NLD/BLD (Narrow Leaflet Drug/Broad Leaflet Drug) hybrid was “Skunk No. 1” produced in the early 1970s.To obtain this variety, plants of the F2 progeny were chosen to carry out nine repeated inbreeding cycles aimed at increasing their homozygosity, then ten female and ten male plants were selected and vegetatively propagated to be used as parental lines in all possible pairwise cross combinations. Such a breeding strategy is very effective for the development of highly heterozygous synthetic varieties, especially if supported by progeny tests to assess the general combining ability (GCA) of parental lines. More frequently, selected F1 plants have been used to generate large segregating F2 populations from which favorable individual should be eventually cloned via cuttings or used in half- or full sibling mating. Cultivated varieties, or cultivars, were mainly produced by crossing a single male of one genetically distinct landrace with a single female of another landrace to produce a hybrid, heterozygous and vigorous offspring. In the subsequent F1generation, selected male or female progenies were bred by following one of these basic strategies: 1) Plants were inbred with one or more siblings to establish a relatively heterozygous or highly heterogeneous F2 population to be used in subsequent mass selection cycles to increase homozygosity and uniformity by intercrossing selected plants; 2) plants were backcrossed with a parental line (the seed parent or the pollen donor) to recover and fix specific traits before establishing mass selection; or 3) plants were outcrossed with an unrelated line (a plant from a third landrace) to integrate new traits and create new recombinants. Each of these breeding strategies was efficiently used to develop new cultivars using experimental hybrid materials that stemmed from crosses between distinct landraces. However, true F1 hybrid varieties were never bred in the past since agronomically super-pure inbred lines to be used yearly as parental lines were difficult to implement. Only recently some professional seed companies have produced and multiplied true F1 hybrid varieties by preserving vegetatively parental clones of the male and female lines. Nevertheless, if the parental clones are not fully homozygous and so genetically unstable, their hybrid progeny is frequently inconsistent phenotypically because of the genetic segregation of maternal and/or paternal traits. As a matter of fact, most seed

companies invest in breeding programs aimed at selecting superior female plants, while male plants are derived from the standard morphological analysis: an individual male is then used as a pollen donor in crosses performed with each of the female clones to produce commercial hybrid seed stocks. These seeds, which do not have the genetic constitution of F1 hybrids, are then widely distributed and grown to maturity so that female plants can be selected and multiplied by cuttings to achieve commercial sinsemilla production. In recent years, seeds of the so-called “all female” cultivars have been largely set by promoting artificially selfing: this is possible by applying hormones to some branches of female plants to let them produce also male flowers with viable genetically female pollen. As a consequence, the offspring of female plants fertilized with female pollen of masculinized branches include only genetically female progeny. This is a very efficient strategy for commercial *sinsemilla* production as all seeds generate useful female plants with no need to remove male plants, so it provides the benefits of asexual propagation (i.e., fixation of the female genotype), but with the advantages of sexual reproduction (i.e., reproduction via seeds in place of cuttings). However, female seeds can give rise to unstable populations characterized by some degree of genetic diversity, in contrast to clonal populations produced from female cuttings. In fact, under sexual reproduction, segregation and recombination mechanisms are all possible unless the parental lines are highly homozygous inbred lines suitable for breeding true F1 hybrids.

1.4 Sex determination and flower development in *Cannabis sativa*

Cannabis (2n=20) is mostly a dioecious plant with a XY sex-determination system; however, monoecious genotypes of *C. sativa* have been bred, especially for the production of seeds and oil used in industrial hemp [26]. Sex determination represents one of the main problems when breeding new Cannabis varieties since it can only be assessed at the beginning of flowering, when male and female flowers are visible and distinguishable.

Flowering and sex determination in Cannabis are important because only unfertilized female flowers are required for the industry (harvested for cannabinoid extraction) and the presence

of hermaphroditic behavior, which is possible in stress conditions, may ruin productivity and breeding experiments.

Cannabis is well known for containing hundreds of specialized metabolites with potential bioactivity, among these the main ones are cannabinoids, terpenes, and flavonoids; all are produced and accumulated in the glandular trichomes that are highly abundant mainly on female inflorescences. [27–29].

1.4.1 Sexual plasticity and the ethylene pathway

The genetic control of dioecy seems to be determined by two specific genes at linked loci acting as sex determinants [18,30,31]. Male plants would require a dominant suppressor of female organs and a dominant activator of maleness, while female plants would share homozygosity for their recessive alleles at both loci. For breeding purposes, male and female plants can then be identified in the early stages of development using Y-specific DNA markers [32,33]. Apart from that, the molecular mechanisms underlying dioecy are essentially unknown but, considering that this condition is fully reversible (e.g., through chemical products treatment), the hypothesis that those genic regions involved in both sexes development remain potentially functional throughout the entire life cycle cannot be excluded [34,35]. Given the role of homeotic genes in flower whorls identity (including anthers, pistils, and ovary), and the hypothesis for their involvement in sex determination[36–38], information on the ABCDE model in the Cannabis genus was characterized with transcriptomic data with the finding of homeotic genes in Barcaccia et al. 2020 [39].

Cannabis sativa L. belongs to the estimated 5-6% of dioecious angiosperm species [12], this provides an excellent opportunity to study the mechanisms involved in sex expression and the evolution of sex-determining regions and sex chromosomes.

Interestingly, dioecious Cannabis genotypes, especially in stress conditions, can display hermaphroditism [40], and photoperiod alteration can revert a plant from sexual to vegetative mode [41]. Finally, sex expression opposite to the XX/XY chromosomal complement can be induced with chemical reagents and environmental factors [42–44].

These facts imply that sex determination in Cannabis is very much plastic and not only controlled by the sexual chromosomes' presence.

Sexual plasticity is the ability of an organism with sex chromosomes to alter its phenotypic sex in response to environmental or physiological changes without resulting in genetic modifications [45].

The term sexual plasticity has largely been used to describe fluid sex expression in the animal kingdom such as in fish, frogs, aphids and bees [46–49]; the mechanisms responsible for these changes vary, but normally involves gene expression changes related to hormone synthesis and signaling, germ cell development as well as microRNAs. [50–53].

Recent studies highlight how sexual plasticity can also be used to describe the interaction between genotypic sex determination and environmental sex determination in dioecious plants containing sex chromosomes, such as *C. sativa* [41].

Research studies identified silver-containing compounds such as silver thiosulfate (STS), which is a potent inhibitor of ethylene signaling, to promote the production of male flowers on female plants; inversely aqueous ethephon ((2-Chloroethyl) phosphonic acid), which rapidly decomposes to produce ethylene gas, was shown to promote the production of female flowers on male plants, these results point to ethylene as being an effector of sexual plasticity [43,44,54,55].

The phenotypic outcomes of these flower reversal treatments in Cannabis have been observed and quantified but the underlying molecular mechanisms controlling ethylene-related sexual plasticity in cannabis have not been clearly explained yet.

Ethylene was discovered in 1901 [56] and has been extensively studied for its crucial role as a plant growth regulator in numerous plant processes such as seed germination, root and shoot growth, flowering, fruit ripening, leaf and fruit abscission, as well as responses to environmental stresses and senescence [57–60].

Thanks to the highly conserved nature of ethylene biosynthesis between species and the well-characterized pathways in model species like *Arabidopsis thaliana* it is possible to reconstruct the pathway in an under-researched plant such as *C. sativa*. The ethylene

biosynthesis and signaling is currently being explored with homology analysis and RNA-seq data.

1.4.2 Flower development and secondary metabolites in *Cannabis sativa*

As previously stated, female Cannabis flowers in Drug-Type varieties contain cannabinoids, and so, understanding flower development is essential to produce the metabolites of interest. Cannabis cultivators, after vegetative growth cultivation, usually conducted in a 18 to 20h of light per day, initiate flower development by decreasing the photoperiod to 12h, this transitions the plants to the reproductive phase, where shoot apical meristems (SAM) are transformed into an inflorescence meristems [61].

Photoperiod is known to have a wide-ranging effect on plant development, such as: controlling flowering time, meristem termination, bud dormancy, and branching. In Cannabis, as in other crops like wheat, onion, rice, photoperiod triggers the initial elongation of flower stalks and flower initiation [62].

In Cannabis the trichomes are where most of the secondary metabolites of interest are found, including phytocannabinoids and terpenes. Trichomes are small glandular projections, they are expressed on leaves and stems but especially on bracts: small leaf-like structures associated with buds of female flowers. As the number of daylight hours begins to decrease, buds undergo intensive branching, ultimately leading to higher trichome (and consequently cannabinoids) yield [62].

It is believed that Cannabis produces terpenes and phytocannabinoids for defense as they can protect the plant from microbial infestations and deter predation by insects [61].

Phytocannabinoid content can also be influenced by growing conditions like temperature, humidity, and soil nutrients [63].

THC (D9-tetrahydrocannabinol) is the most well-known phytocannabinoid, its chemical structure was reported by Gaoni and Mechoulam in 1964 [64], it is the main metabolite responsible for the euphoric effects induced by marijuana-type Cannabis.

All the most abundant phytocannabinoids have a singular common biosynthetic precursor which is Cannabigerolic acid (CBGA), from that three closely related enzymes catalyze the production of cannabichromenic acid (CBCA), cannabidiolic acid (CBDA), and D9-tetrahydrocannabinolic acid (D9-THCA) [65].

Phytocannabinoids represent a group of C21 or C22 (for the carboxylated forms) terpenophenolic compounds that are predominantly produced in Cannabis, more than 90 different cannabinoids have been reported in the literature, although many of these are breakdown products resulted from the chemical degradation of the three main ones (CBC, CBD and THC) [66].

Terpenes and terpenoids are considered to be the second-most important class of secondary metabolites produced by the Cannabis plant as they are largely responsible for the characteristic scent of Cannabis and are thought to contribute to its potential therapeutic benefits [61]. Terpenes identified in Cannabis consist of monoterpenes, 10- carbon hydrocarbons biosynthesized from geranyl pyrophosphate, and sesquiterpenes. Prevalent terpenoids in Cannabis include αL-pinene, β-pinene, β-myrcene, and β-caryophyllene.

1.5 Advances in Cannabis genomics

1.5.1 Conventional schemes and next generation methods for breeding Cannabis cultivars

Breeding methods conventionally used for the development of new varieties have been revolutionized since the advent of genomics applied to crop plant species. In fact, the examination of plant materials using molecular markers linked to single loci controlling specific traits of agronomic interest (i.e., marker-assisted selection, MAS) and the exploitation of multiple loci genotyping with molecular markers scattered throughout the genome (i.e., marker-assisted breeding, MAB) provide the opportunity to boost gain from selection [67].

Recent signs of progress in the development of multiplex SSR markers assays have been made in several crops [68–71], suggesting that these markers, especially when finely mapped and scattered throughout the genome, remain as relevant and cost-effective molecular tools at least for characterizing genetic resources and breeding new varieties.

For many years, the development of new varieties of medical Cannabis was not the exclusive preserve of breeders. Home growers who have acquired high-level skills and learned essential techniques of hybridization, selection, and cultivation have easily transitioned their activities from growing to breeding Cannabis lineages. In recent decades, home growers have created most of the Cannabis strains that have become popular in the market worldwide. Both medical (drug-type) and hemp (fiber-type) cultivars were traditionally developed for many years using mass selection. Cannabis varieties can then be easily preserved and multiplied via cuttings from individual plants that exhibit desirable traits matching a specific distinct phenotype. Propagation via cuttings is the main way to make prized varieties available as clones to maintain unaltered genotypes. When Cannabis varieties are multiplied and commercialized through seeds, open-pollinated OP synthetics and F1 hybrids represent the only populations that can be reproduced sexually, giving rise to offspring characterized by morphological distinctiveness and uniformity, and genetic stability across generations. Cannabis is a dioecious (and anemophilous) species, with male and female plants exhibiting stamens and pistils in separate flowers. As a consequence, outcrossing through wind-mediated cross-pollination is the only natural reproduction system of Cannabis spp. The genetic structure of both natural populations and experimental breeds obtained via mass selection can usually be composed of a combination of highly heterozygous genotypes that share a common gene pool. "*Selfing*" is also possible and can be accomplished by artificially generating monoecious plants with unisexual flowers (i.e., reversing the sex of flowers from female to male on some branches) to induce self-pollination. Attempts were made to transform the reproductive organs of Cannabis using irradiation [72] and streptovaricin [73] but the results were impractical. The successful use of other strategies, such as the feminization of male plants using ethephon [74] and the masculinization of female plants with silver thiosulfate [43,55,74] enabled to revolutionize breeding programs in Cannabis.

This latter treatment, in particular, is still largely used since thiosulfate inhibits the production of ethylene, a plant hormone that promotes the formation of female flowers. On the treated branches, the newly induced male flowers can develop anthers with viable pollen, while the other untreated branches of the plant will continue to grow female flowers. The female plants whose pistils are self-pollinated and their egg cells (X) fertilized by genetically female pollen (X) will give rise to a completely female progeny (XX). This method, exploitable for the multiplication of female plants by seeds, can be commercially more convenient than the female propagation by cuttings. Nevertheless, sexual reproduction can originate segregating populations, genetically unstable and characterized by phenotypic variability, negative features that are not shown by clones. The only way to successfully use seeds of Cannabis varieties is the one based on the development of true F1 hybrids by crossing genetically divergent but individually uniform parental inbreds. In addition to this strategy for selfing, the production of highly homozygous genotypes can be achieved from full-sibling crosses performed by hand between sister-brother individuals that belong to the same progeny and share the same two parental lines. Cannabis (sinsemilla) varieties were largely developed by crossing single male and female individuals belonging to genetically distinct landraces to create a pseudo-F1 hybrid. The genetic stability and uniformity of any new cultivar bred in this way can only be preserved as an individual clone through vegetative propagation through cuttings. To breed true F1 hybrid varieties, inbred lines stemmed via repeated selfing and/or full-siblings for some cycles can be used as parental stocks for the production of highly heterozygous hybrids through two-way crossing to exploit the effects of heterosis. Heterosis refers to the phenomenon in which F1 progeny obtained by mating two genetically divergent and antagonist inbred lines exhibit greater biomass, rate of development, and fertility than the two homozygous parents. This biological phenomenon has been extensively exploited for the development of crop varieties in several plant species and has been important for the development of modern fiber (hemp) cultivars but is still largely unexplored or undocumented in recreational (drug) cultivars. Since heterosis often results from the complementation in the hybrid of different deleterious (recessive) alleles that were present in one parental genotype by superior (dominant) alleles from the opposite parental genotype,

the development of F1 hybrids usually requires progeny tests for estimating the specific combining ability (SCA) of selected inbred lines in all possible pairwise cross-combinations (diallel design). This method not only requires the selection of individual breeding parents (single female and/or male plants) but also requires that some of the progeny plants are asexually propagated via cuttings to perform laboratory analyses and field trials. In particular, in each generation, the selection of the most appropriate plants from either selfing or full-siblings is based on agronomic, genomic, and metabolomic investigations to choose the best individuals in terms of agronomic performance, molecular genotypes, and biochemical profiles. Selected individuals should also be used to perform parallel progeny tests aimed at determining their SCA based on F1 hybrid evaluation. A key step for large-scale seed production is the use of an inbred female plant (XX) as the clonal seed parent line and another genetically divergent but complementary inbred female plant (XX) that has been masculinized as the clonal pollen parent. Thus, 100% of the F1 hybrid seeds will be female (XX): all-female seeds are produced by cross-pollination, but all-female plants are characterized by the same highly heterozygous and vigorous genotype. The same strategy can be exploited for breeding F1 varieties through two-way, three-way, or four-way hybrids using two, three, or four inbred lines derived from as many parental materials/landraces through intrasubspecific and intersubspecific hybridization. In fact, in addition to pure "indica" and "sativa" varieties, hybrid varieties with varying ratios of their genomes are common. For instance, among the most famous varieties worldwide, the "White Widow" exhibits approximately 60% "indica" and 40% "sativa" ancestry, and its plants exhibit traits from both parental biotypes. Nevertheless, the choice of the initial cross depends on the targeted Cannabis market (fiber vs. drug utilization genotype and tetrahydrocannabinol/cannabidiol ratio), as some varieties are bred mostly as medicinal Cannabis, and others are instead highly appreciated as recreational Cannabis. Breeding for fiber production includes both monoecious and dioecious cultivars showing a high percentage of primary fibers, fast-retting phenotypes, and distinctive morphological descriptors in low-THC plants. Breeding for the production of cannabinoids comprises THC-predominant or cannabidiol (CBD)-predominant cultivars. It is worth mentioning that a

limited number of cultivars have been specifically bred for seed production [29]. Considering the relevance of genomics and metabolomics in the development of next-generation Cannabis varieties, modern breeding methods must be based on the application of multidisciplinary skills and tools to assist professional agronomists in the evaluation or prediction, and early selection of plants with the highest potential in terms of molecular genotypes and biochemical profiles. Cannabinoids of breeding stocks can be assayed according to either quantity (i.e., percentage of cannabinoids in harvested material) or quality (i.e., THC/CBD ratio or chemotype). The quality of cannabinoids is strongly dependent on the genotype, whereas cannabinoid quantity is affected by agronomic practices, environmental conditions, and genotype x environment interactions.

1.5.2 Marker Assisted Breeding in *Cannabis sativa*

Since the advent of genomics applied to crop plant species, breeding methods conventionally used for the development of new varieties were rearranged and readapted, as for many traits selection can be assisted by molecular markers. In particular, both simple- and multiple-locus genotyping approaches proved their utility for improving the overall genetic stability and uniformity of cultivated populations as well as for pyramiding specific genes that control resistance or tolerance to both biotic and abiotic stresses. In addition to large panels of molecular markers useful for genotyping purposes, several next-generation platforms for genome sequencing and new biotechnological techniques for gene editing are nowadays available in many crop plant species. These molecular tools allow scientists to better characterize and estimate the breeding value of plant individuals and populations using lab analyses, materials which are then used by breeders for field trials to select the superior and ideal phenotypes showing distinctiveness, uniformity, and stability. The use of genomics in Cannabis has its roots around 25 years ago with the use of dominant markers such as RFLP, RAPD, and AFLP markers [75–79] to assess the genetic relatedness of species, varieties, and even individuals. Later on, microsatellite or SSR markers were shown to be more informative, reliable and reproducible than dominant markers for Cannabis genotyping [80–

82]. Specific marker alleles/variants were also identified as predictive and capable of discriminating hemp from marijuana [83]. Among the most relevant microsatellite-based studies conducted on Cannabis, two relatively recent researches deserve to be mentioned. In the first one, a panel of 13 SSR markers was used to test over 1,300 samples of fiber Cannabis and marijuana, together with accessions from local police seizure [84]. In the same year, Soler et al. [85] characterized the genetic structures of 154 individuals belonging to 20 cultivars of *C. sativa subsp. indica* and 2 cultivars of *C. sativa subsp. sativa* using a set of 6 SSR markers. However, despite the number of studies conducted using dominant markers and codominant microsatellites, only Soler et al. (2017) [85] opened to the concrete possibility of using these molecular tools for breeding goals, including the improvement and development of new varieties. Most of the studies were instead focused on germplasm management, genetic discrimination of varieties and forensic applications (e.g., drug vs. non-drug types identification). While any marker-assisted breeding strategy in Cannabis is still far to be explored, marker-assisted selection has already been successfully used. One of the main achievements that contributed the most to the shift from traditional to molecular breeding in Cannabis, is the release of the first two genomes of *C. sativa* in 2011 [17]. Since then, many studies focused on bioinformatic analyses of these genomes to mine molecular markers tightly linked to expressed genes [86] and hence useful for Cannabis marker-assisted characterization and selection studies. The availability of sequenced genomes also allowed the identification and exploitation of thousands of SNP variants, which together with Genotyping-by-Sequencing (GBS) approaches, enabled the analysis of the genetic diversity of several Cannabis accessions belonging to hemp and medical/recreational varieties. The use of GBS in Cannabis spp. has been recently described by Soorni et al. (2017) [87] which analyzed 98 samples from two Iran germplasm collections, obtaining over 24 thousand highly informative SNPs. Also, in this case, SNP markers proved to be useful not only to classify samples belonging to different Cannabis varieties but also to identify polymorphisms associated with genes belonging to the cannabinoid pathway, like THCAS and CBDAS (delta-9-tetrahydrocannabinolic acid synthase and cannabidiolic acid synthase, respectively) [17,88,89]. These markers could be extremely useful in breeding programs aimed at

developing new Cannabis varieties for fiber production (drug-free) or medical/recreational use. Using this approach, [90] developed a physical and genetic map of *C. sativa* focusing their attention on those genes involved in the cannabinoid synthase. In particular, authors coupled the genomes of Purple Kush and Finola varieties [17] to the Pacific Biosciences (PacBio) long-read single-molecule real-time (SMRT) sequencing and Hi-C technology to generate a combined genetic and physical maps of Cannabis. This provided new insights on the chromosome arrangement and the cannabinoid biosynthetic genes. Another milestone from the Laverty et al. (2019) [90] study is the identification of an important gene involved in the biosynthesis of cannabichromene, a cannabinoid with a weak activity on the CB1 and CB2 receptors (involved in the neural and psychoactive effect of THC and CBD) that could be possibly used in medical therapies against pain and gastro-inflammatory diseases [91]. More recently, based on the latest knowledge acquired on Cannabis genomics, [92] described the efficiency of a screening method based on KASP (Kompetitive Allele Specific PCR) technique for the identification of 22 highly informative SNPs involved in the biosynthetic pathway of cannabinoids and terpenes (important compounds for the recreational and medical Cannabis industries). It must be recognized that the increased knowledge on the most relevant Cannabis biosynthetic pathways has been possible thanks to the continuous refinement of available genomes together with the public delivery of new ones. Recently, McKernan et al. (2020) [93] sequenced and annotated 42 Cannabis genomes identifying SNPs useful for molecular breeding related not only to the cannabinoid synthesis but also to pathogen resistances. This could help in the production of medical/ recreational Cannabis without the risk of mildew contaminants that could be dangerous for consumers. In parallel, Gao et al. (2020) [94] assembled a new genome of *C. sativa* deriving from wild samples collected in Tibet using a combination of PacBio and Hi-C technologies. Despite all these efforts, an exhaustive meta-analysis of all the Cannabis genomics data published so far [1] demonstrated that the currently available Cannabis genome assemblies are: i) incomplete, with approximately 10% missing, 10–25% unmapped, and centromeres and satellite sequences unrepresented; ii) ordered at a low resolution and only partially annotated for what concerns genes, partial genes, and pseudogenes. Wrapping up if, on one hand, the enormous

interest raised by specific metabolic compounds (e.g., THC) has boosted the achievement of high levels of knowledge for specific biosynthetic pathways, on the other hand, the use of molecular markers for breeding new varieties is still in its embryonic phase and undoubtedly deserves further investigation to develop efficient tools transferable among laboratories. Considering the availability of a remarkable number of sequenced Cannabis genomes, the starting point could be the development and implementation of an informative and representative panel of polymorphic SSR marker loci scattered throughout the genome for standardized multilocus genotyping purposes.

CHAPTER 2: Cultivation for Cannabis sativa

2.1 Introduction

The use of the appropriate nutritional measures, cultivation methods as well as understanding the interactions between microbiome and plant growth are a vital aspects of medicinal plant production including medical Cannabis. Due to legal restrictions on Cannabis research, very little information is available concerning the effects of nutritional supplements and cultivation techniques on physiological and chemical properties of medical Cannabis, and their potential role in standardization of the active compounds in the plant material supplied to patients [95].

2.1.1 Light in Cannabis cultivation

Light is a key factor affecting plant growth, metabolism and function [96]. Light intensity and duration affect major plant developmental stages including the switch from the vegetative to the flowering stage (induced by photoperiodism in Cannabis), it is also involved in stomatal elongation and leaf expansion [97], as well as secondary metabolites production and accumulation [98].

Following years of illicit growing and prohibitions very little and incomplete information is available about responses of Cannabis to light in general, and, in particular, regarding light quality; its unique biologically active compounds (especially cannabinoids) are considered to be affected by light spectra but research on the topic is still its infancy.

A recent study by Danziger and Bernstein [96] showed that the accumulation of the primary cannabinoid CBGA, that was up to 400 % higher in the LED treatments compared with the HPS control and that the concentrations of some of the primary cannabinoids (THCA, CDBA or CBCA) decreased under LED cultivation, with a reduction up to 40 %.

Results also varied among Cannabis varieties suggesting that customizing the bioactive profile with the manipulation of the light spectrum is possible but must be characterized for the specific plants in use.

2.1.2 Effects of nutrients to cannabinoid and terpenoid yield

Studies show that nutritional supplements influence cannabinoid content in an organ- and spatial-dependent manner [95]. Here, experiments results regarding the supplementation of NPK (Nitrogen, Phosphorus, Potassium) on Cannabis growth and their effect on secondary metabolite biosynthesis and accumulation will be presented; NPK are the three primary mineral nutrients plants need to grow.

2.1.2.1 Nitrogen

Nitrogen (N) is a key plant nutrient, known to affect primary and secondary metabolism in plants [99]. Studies conducted with nitrogen concentrations from 30 to 320 mg/L showed that concentrations of most cannabinoids and terpenoids were highest under the deficient concentration (30 mg/L) but morpho-physiological state of the plants was optimal at supply rates of 160 to 320 mg/L [100].

Inflorescence yield quantity and quality ratio was at its highest at 160 mg/L with the highest secondary metabolite accumulation. Another study showed that excess uptake of nitrogen supply higher than 160 mg/L during vegetative growth promoted physiological and developmental restrictions; these were caused by ion-specific toxicity and by the induction of carbon fixation restrictions and lowering energy availability [101].

Regarding nitrogen source, plant species differ in the preference for NO_3 or NH_4 based nutrition [102]. Studies conducted in Cannabis with an increasing NO_3 to NH_4 ratio showed a 46% decrease in inflorescence yield with the increase in NH_4 supply from 0 to 50% [103]. Still, moderate levels of 10–30% NH_4 were deemed to be suitable for medical Cannabis

cultivation, as they did not significantly damage plant function and showed only very little adverse influence on yield and cannabinoid production [103].

2.1.2.2 Phosphorus

Phosphorus (P) is an essential macronutrient and a key element in nucleic acids and phospholipids, as well as in energy transfer processes in the cell, it therefore participates and affects central biosynthesis pathways [99].

In a research study where Cannabis growth was achieved under 5 to 90 mg/L phosphorus nutrient concentration [104] it was observed that:

- P concentrations lower than 15 mg/L were insufficient to support the optimal plant functions as reduced photosynthesis, transpiration, stomatal conductance, and growth were observed.
- Inflorescence biomass was at its highest between 30 to 90 mg/L but cannabinoid concentrations were highest under 30 mg/L.

It was noted that the lowest recommended P supply for optimal yield quantity is 30 mg/L but that different varieties differed in their response with one with an optimal yield at 90 mg/L.

During experiments conducted on Cannabis during vegetative growth (long photoperiod) results showed that 30 mg/L supply is optimal for plant function [105].

2.1.2.3 Potassium

Potassium (K) is an essential plant nutrient and takes part in many key physiological processes such as: stomatal regulation, protein synthesis, photosynthesis, enzyme activation, osmoregulation, and the uptake and accumulation of other essential cations such as Ca and Mg [106,107]. In experiments where medical Cannabis was grown in controlled environment conditions, under 15 to 240 mg/L of K supply [108] showed that:

- 15 mg/L impaired development and functions provoking visual chlorosis and decreasing inflorescence yield.

- 175 mg/L supply was demonstrated to be optimal for plant function and high yield but increasing it to 240 mg/L showed over-supply symptoms.

It was concluded that, as secondary metabolite concentrations were inversely correlated to K supply, 60 mg/L of K supply was suggested to maintain both high yields combined with high secondary metabolites productivity.

2.1.3 Plant density and shape

A major challenge for utilizing Cannabis for modern medicine is the spatial variability of cannabinoids concentration in the plant, which results in differences in medical potency depending on flower positioning on the canopy.

Environmental conditions, such as micro-climates altered by plant density can affect secondary metabolism and so, canopy architecture and plant density play essential roles in the standardization of the cannabinoid profile [109].

Recent studies showed that cannabinoid concentrations were reduced in the lower part of the plant with the increase in plant density thus highly reducing the cannabinoid uniformity across the plant [110]; it was noted that this decreased inflorescence yield per plant but increased yield per area and so, depending on the final product goals (yield quantity per area or chemical stability), different cultivation practices should be applied.

Parallel studies showed that low cannabinoid concentrations in inflorescences at the bottom of the plants correlate with low light penetration, and so, increasing light penetration by defoliation or removal of bottom branches and leaves, increases cannabinoid concentrations locally thus reaching higher levels of plant spatial uniformity [111].

2.1.4 Microbiome interactions with plant growth

Recent studies are trying to investigate the interactions between the microbiome and important agronomic and production factors such as seed germination, phosphorus solubilization, mycelial growth inhibition [112], vegetative growth and bud yield [113–115] in *C. sativa*.

A research study showed that high germination rates were observed with Bacilli strains as well as significant inhibition of mycelial growth [112], which is important as fungal growth has been reported to cause diseases damaging crop production [116].

At the same time, in numerous research experiments, microbial biostimulation was shown to have several other benefits such as:

- increasing phosphorus solubilization by improved bioavailability of non-useable tricalcium phosphate into a useable form of monophosphate [112];
- improving yields of cannabinoids and enhance the balance of cannabidiol (CBD) and tetrahydrocannabinol (THC) proportions [114];
- promoting *C. sativa* growth by synergistic Pseudomonas–Bacillus interaction with up to 30% increase at the total dry plant weight [115];
- producing larger and more robust plants that yielded 16.5% more product with the use of commercial microbial biostimulant (Mammoth P^{TM}) [113].

Research on this subject will help future advancements in sustainable cultivation practices, disease management, and nutrient availability enhancement resulting in improved plant growth and higher yields.

2.2 Materials and methods

2.2.1 Hemp cultivation

Eleven hemp-type Cannabis varieties were grown at Gruppo Padana (Gruppo Padana Ortofloricoltura S.S., Via Olimpia 41, 31038 Paese, Treviso, Italy).

The industrial hemp varieties used had European origins: Italy, France, Poland, Hungary, Netherlands, and Finland and were all grown starting from authenticated seeds.

2.2.1.1 Agronomic conditions

All varieties were cultivated in greenhouses with artificial lighting additions if necessary, depending on the season and the desired plants state if required, lighting hours per day were increased to 18 for vegetative growth or 12 for flower initiation and development.

Seeding was conducted in 96- cells seedling trays with 4 cm holes where germination rates were calculated for all varieties, results varied significantly depending on seeds age and among varieties.

Afterwards plants were transferred to increasingly bigger vases, from 14 up to 32 cm in diameter for further tests and experiments. The growing medium used was Jongking fertilizer (Jongkind Substrates, Oosteinderweg 35, 1432 AX Aalsmeer, Netherlands).
(https://www.jongkind-substrates.com/en/jongkidn-substrates).

2.2.1.2 Genotyping experiments on hemp samples

104 plants from all 11 hemp varieties available were selected for testing genotyping markers, samples for DNA extractions were collected from young leaves (usually at 3rd or 4th internode) typically from 4 to 5 weeks' old plants.

Varieties studied had different sexual behaviors with 5 monoecious and 6 dioecious varieties present, details and characteristics of the samples are shown in the following table (Table 2.1).

Variety	Origin	Leaf	Number of Samples	Sex behavior of the samples		
				Dioecious (Male)	Dioecious (Female)	Monoecious
ITA1	Italy		10	7	3	
ITA2	Italy		9	4	5	
ITA3	Italy		10	2	8	
HUN1	Hungary		11	6	5	
HUN2	Hungary		10	5	5	
FIN	Finland		11	3	8	
NED	Netherlands		13			13
POL	Poland		8			8
FRA1	France		10			10
FRA2	France		9			9
FRA3	France		3			3

Table 2.1: Hemp samples collected for testing genotyping markers.

Samples from these plants have been used to develop and test both SSR marker based, and chloroplast sequencing based genotyping protocols, in addition, as will be described afterwards plants from this list have also been used for flowering experiments.

2.2.2 Drug-type Cannabis cultivation

During my second PhD year, at my stay at the University of Saskatchewan, Saskatoon, (SK, Canada) in the laboratory of Prof. Timothy Sharbel, different cultivation conditions were tested, involving seeding, cuttings from mother plants, vegetative plant growth, flowering and sex reversion experiments and inflorescences exsiccation and analysis.

Plants were grown at the Controlled Environment Facility (College of Agriculture and Bioresources), 8 drug-type THC variety and one undefined Canadian hemp variety (Tab 2.2) varieties were cultivated and experimented upon.

Variety	Additional Information
Namibian	Drug-type TCH variety, supposed African origin
East Coast Sour Diesel	Drug-type TCH variety, supposed USA bred
Congo LMN	Drug-type TCH variety, supposed African origin
The King	Drug-type TCH variety, supposed progenitor to Pink Kush
Special K	Drug-type TCH variety, supposed Canadian bred
GMO cross	Drug-type TCH variety, unknown origin
Timewarp	Drug-type TCH variety, Canadian bred on Texada Island
Hashmiri	Drug-type TCH variety, unknown origin
Canadian Hemp	Hemp-type Cannabis, Canadian origin, undefined variety

Table 2.2: List of Cannabis varieties cultivated at the University of Saskatchewan.

2.2.2.1 Growing media and fertilizer

The growing media used for all growth stages “Sunshine Mix #4” was composed of roughly 60-70% sphagnum peat moss with 30-40% perlite with added dolomite lime as a pH buffer and wetting agent (https://www.sungro.com/retail-product/sunshine-mix-4/).

As a fertilizer “mega crop” was used diluted at 1g/L, its stated composition was the following:

- 9.7% total Nitrogen (N) with 9.1% NO_3 and 0.6% NH_4
- 7.8% Available Phosphate (P_2O_5)
- 15.1% Soluble Potash (K_2O)
- 8.1% Calcium (Ca)
- 3% Magnesium (Mg)

- 3% Sulfur (S)
- 0.02% Boron (B)
- 0.04% Copper (Cu) as Amino Chelate Copper
- 0.15% Iron (Fe) as Amino Chelate Iron
- 0.7% Zinc (Zn) as Amino Chelate Zinc
- 0.1% Molybdenum (Mo)
- 0.15% Silicon Dioxide (SiO_2)
- 1% Kelp Extract

2.2.2.2 Phytotrons used for Cannabis cultivation.

Three phytotrons were used for our experiments, in all of them temperature was controlled at 24 °C and humidity was maintained between 50 and 60%.

- **Phytotron #1** (4.5 m^2 area, 2 m height) was used for seed germination, stem cuttings clonal propagation from mother plants, and vegetative growth prior to flowering induction.

  Lighting was maintained for 18h a day at 600 µmol/m^2s intensity on leaf canopy with Fluorescent/Incandescent 30,000 lux height adjustable lamps.

  Seeding was conducted in a 50 cells seedling tray with 5 cm cubic holes filled with Sunshine Mix #4 media, clonal cuttings were grown in high humidity containers to induce root formation, and finally, vegetative growth was achieved in 15 cm diameter and 15 cm high vases.
- **Phytotron #2** (1.5 m^2 area, 1.5 m height) was used for flower development experiments. Plants in vegetative state were transferred to larger vases (28 cm diameter and 25 cm height) briefly prior to flower initiation, flowering was achieved by changing the photoperiod from 18 to 12h per day.

  Lighting was maintained at 600 µmol/m^2s intensity on leaf canopy with Fluorescent/Incandescent 84,000 lux height adjustable lamps.

- **Phytotron #3** (3.3 m^2 area, 2 m height) was used for flower growth, collection and exsiccation, flowering was maintained for 70 days after induction with all cultivated THC-rich varieties.

  To collect flower material for further analysis plants were cut and leaves were removed, they were then left in the dark at 15 °C until exsiccation which occurred in around 2 weeks.

2.2.2.3 Flowering, sex reversion and sex determination experiments

Numerous flowering experiments were conducted to model flowering behaviors and sex reversion protocol.

The first experiments were conducted at Gruppo Padana where, as stated, both monoecious and dioecious hemp varieties were cultivated; flowering timing was also distinct among varieties with one, from Finland, exhibiting autoflowering behaviors which resulted in change from vegetative to flowering state without the need for photoperiod reduction.

Main flowering experiments conducted included:

- Basic flowering timing experiment where all 11 cultivated hemp varieties (Table 2.1) were changed from vegetative to flowering state by photoperiod reduction, this test gave us information about a timetable for flower initiation and development for each variety, essential for further experiments. Flowers from the HUN1 varieties were also collected in liquid nitrogen for testing metabolomic and transcriptomic protocols (results presented in **Chapter 5***: Flower development studies*).
- Sex reversion tests where Silver Thiosulfate was used with different protocols in dioecious varieties [55,74], an effective procedure was found to achieve a complete Induced Male Flower (IMF) production on treated branches; this was accomplished firstly in hemp varieties cultivated at Gruppo Padana (Table 2.1) and, later, also on drug-type Cannabis varieties cultivated at the University of Saskatchewan (Table 2.2).

- With the accumulated knowledge a final flowering experiment where samples from Female Flowers (FF), Male Flowers (MF) and Induced Male Flower (IMF) were collected from three varieties (HUN1, HUN2 and ITA3, from Table 2.1) for metabolomic and transcriptomic experiments.

Additionally, we also focused on testing a reliable molecular marker for sex determination in Cannabis plants. Separating male and female plants at early developmental stages is useful due to the influence of gender on agriculturally significant traits [117]. Sequence-characterized amplified region (SCAR) markers, such as SCAR119, have been developed to identify male plants, which are presented with a single band of 119 bp [33].
To test the reliability of this molecular assay, we applied this analysis to the entire germplasm (all dioecious hemp and drug-type samples) by using the protocol described by the authors and the following oligos PCR primers:
SCAR119_F: 5'-TCAAACAACAACAAACCG-3';
SCAR119_R: 5'-GAGGCCGATAATTGACTG-3'.
Electrophoresis was performed using a 1.5% agarose gel on a Mupid®-one apparatus (Advance®) with Tris-acetate-EDTA buffer (TAE) and SYBR Safe for staining. The presence and size of PCR products were visualized using an UV transilluminator (Uvitec Cambridge®).

2.3 Results and discussion

2.3.1 Cannabis Cultivation

Hemp plants samples cultivated at Gruppo Padana were used for multiple genotyping experiments such as SSR marker testing (**Chapter 3:** *Developing and Testing SSR Molecular Markers in Cannabis sativa*) and chloroplast sequencing (**Chapter 4:** *Chloroplast sequencing markers in Cannabis sativa*).

Samples collected from drug-type plants cultivated at the University of Saskatchewan were used for developing and testing Chloroplast sequencing markers (**Chapter 4:** *Chloroplast sequencing markers in Cannabis sativa*).

2.3.2 Flowering experiments

Plants from all cultivated varieties (hemp and drug-type) were used for flowering and sex reversion experiments.

Flower development was induced by reducing the photoperiod from 18 to 12h of light per day and was observed and documented for all cultivated varieties.

Distinguishing flower identity between male or female flowers was visually possible around 7 days after induction, pollen release form male anthers was observed from around 14 to 21 days after induction and seed formation at around 25 to 31 days after induction.

Visual confirmation of flower identity can be seen in Figure 2.1 for dioecious female and male flowers as well as for monoecious ones.

Fig 2.1: Flower development at 14 days after induction for dioecious female (Left image), male (Middle image) and monoecious (Right image) varieties.

Regarding sex determination experiment we confirmed that the SCAR119 molecular marker was 100% effective in predicting the sex of the samples as a 119 bp amplicon was detected

in all the male individuals but not in female and monoecious plants. Testing was conducted on DNA samples extracted from young leaves in the plants vegetative state, after flower development predicted sex was checked visually and was successful in all cases.

2.3.2.1 Sex-reversion experiments

After numerous sex-reversion testing, a protocol was established which could induce the production of male flowers on genetically female treated plants.

The protocol consisted in preparing the following Silver ThiuSulfate (STS) solution (in 100 mL H_2O): 0.051g Silver Nitrate (3 mM); 0.190g Sodium Thiusulfate (12 mM).

The resulting solution was applied by spraying the leaves of the target plants, around 20 mL of solution were found to be necessary for the assurance of the successful treatment of a plant, volume of solution needed depended on size; about 2 mL of solution were deemed necessary for the treatment of each branch. The treatment was conducted twice, firstly as soon as the photoperiod plant was changed from long (vegetative growth) to short (flowering stage) and subsequently 5 days after.

The treatment was successful in sex-reverting all flowers on treated branches with 100% male flowers observed there, the only female flowers were rarely detected on new branches grown directly from the low section on the main stalk; flowers from lower branches were therefore never collected in sex-reversion experiments to avoid the possibility of gathering non-reverted samples.

The successful treatment and the resulting male-induced flower development compared to control female flower development can be observed in Fig 2.2 where images taken from 10 to 33 days after flower induction in five different drug-type varieties are presented.

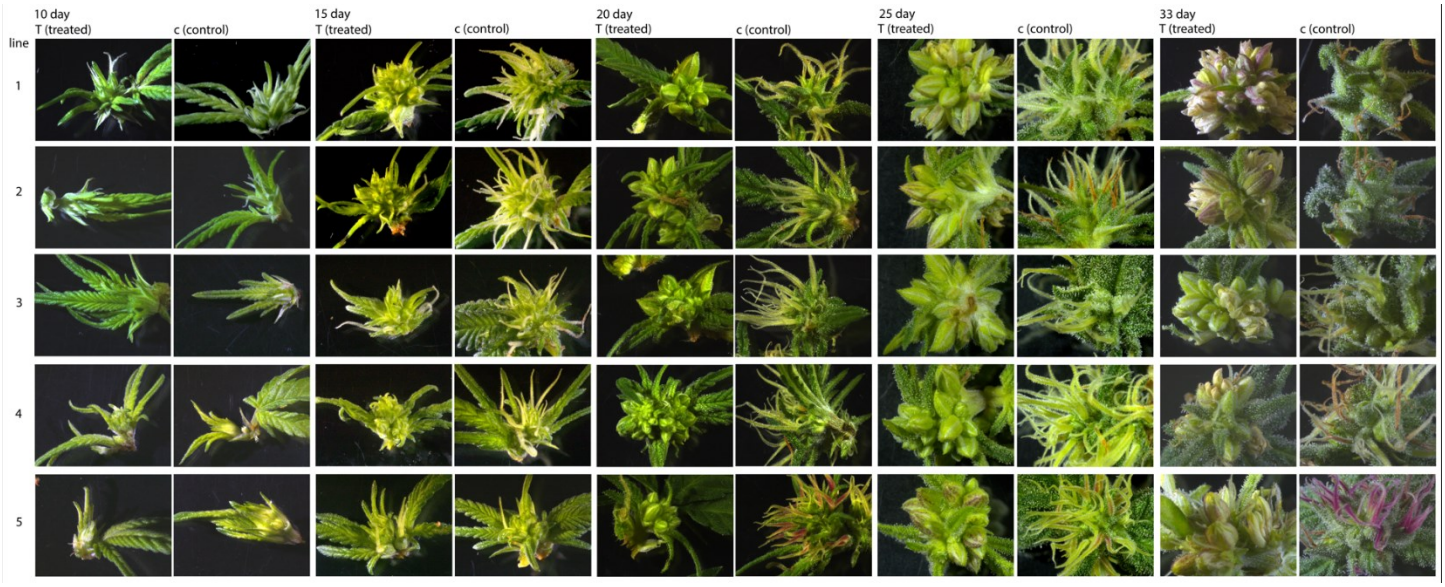

Figure 2.2: flower development for 5 drug type cannabis varieties (line 1: Timewarp; line 2: East Coast Sour Diesel; line 3: Special K; line 4: Congolese LMN; line 5: The King). For each line T (treated = male induced flower) and c (control = normal female flower) images are confronted.

2.3 Conclusions

Successful cultivation procedures and flowering protocols allowed us to collect many samples for molecular experiments.

Leaf samples collected during vegetative growth were used for developing and testing genotyping methods such as with SSR markers (**Chapter 3:** *Developing and Testing SSR Molecular Markers in Cannabis sativa*) and chloroplast sequencing (**Chapter 4:** *Chloroplast sequencing markers in Cannabis sativa*) as well as sex-determination protocols with SCAR119.

Samples collected during flowering and sex-reversion experiments were instead used as starting materials in **Chapter 5:** *Flower development studies* for metabolomic and transcriptomic analysis investigating flower development, sexual identity and cannabinoid pathway and content.

CHAPTER 3: Developing and Testing SSR Molecular Markers in Cannabis sativa

3.1 Introduction

C. sativa (2n=2x=20) is a popular species belonging to the Cannabaceae family. Despite its use for medical, recreational, and industrial purposes as well as its long history, the genetic and biotechnological research on this plant species is in its infancy due to the legal implications and the prohibition campaigns linked to its utilization and cultivation. For many years, the development of new varieties has not exclusively been pursued by breeders because home growers have easily transitioned their activities from growing to breeding Cannabis lineages. Recently, the use of genomics has led to a boost in approaches aimed at marker-assisted selection, germplasm management, genetic discrimination, authentication of cultivated varieties or cultivars and forensic applications, such as protection of intellectual property rights.

In recent times, molecular markers in Cannabis have been mostly employed to investigate genetic diversity, geographical origin, relationships between cultivated germplasm, varietal characterization and sex determination as well as to generate genetic linkage maps [118]. Among molecular markers, SSRs possess numerous desirable features, such as a high level of reproducibility and codominant heredity. Moreover, in contrast to other markers such as SNPs, microsatellites do not require high-throughput technologies and computational resources for their development and analysis [70,119,120].

Microsatellites have been frequently used in Cannabis cultivars and genotypes to evaluate genetic diversity. While early studies have identified and characterized a small number of polymorphic SSR sites specific for Cannabis sativa, [80–82], these markers have been recently used to discriminate between drug- and fiber-type (hemp) varieties [15,81,85–87,121–124] demonstrating their effectiveness as an alternative method to biochemical analyses [1].

Regarding geographical origin, a several studies have reported that the highest genetic diversity is found within hemp varieties compared to drug varieties, and it is possible to differentiate distinct gene pools in Europe and central Asia [86,125,126].

A previous study had developed a panel of SSR markers, organized in multiplex assays, to predict in silico from polymorphic and widely distributed microsatellite regions of Cannabis spp. [39]. Nonetheless, the exploitation of molecular markers for the development of commercial varieties through marker-assisted breeding (MAB) schemes is rare.

During my PhD program two molecular markers methods for genotyping were developed: a simple sequence repeat (SSR) informative panel test and a chloroplast sequencing test (described later in ***Chapter 4***) were established and proven to work on hemp and "medical" Cannabis samples.

Regarding the SSR method we started from 41 nuclear SSR regions designated by in silico analyses, we selected 20 highly polymorphic and discriminant marker loci that were first tested in over 104 individual plants belonging to 11 distinct hemp varieties. The selected markers were demonstrated to be successful in accessing heterozygosity/homozygosity, genetic uniformity/stability and genetic variation within and among varieties. Population structure analysis identified 8 genetic groups, clustering individuals based on sexual behaviors (dioecious and monoecious) and geographical origins.

Results obtained during my genotyping experiments over the course of my PhD program are presented in this chapter, the results of genotyping hemp with SSR markers were used to publish the article: *Developing and testing molecular markers in Cannabis sativa (Hemp) for their use in variety and dioecy assessments* [127].

3.2 Materials and methods

The research work was started by developing the microsatellite panel in silico and then testing its efficacy on eleven hemp varieties; what follows is a detailed description of the materials and methods used to do so.

3.2.1 Characterization of microsatellites in the Cannabis genome

Cannabis genome is diploid (2n = 2x = 20) and its haploid nuclear genome size is estimated to be 818 Mbp for females (karyotype XX) and 843 Mbp for males (karyotype XY) [13]. The *C. sativa* plastid and mitochondrial genomes are 153,871 bp [128] and 415,545 bp [129], respectively. Among the 12 Cannabis genomes available in GenBank, 5 were assembled at the chromosome level, while the remaining ones are considered drafts at the contig (6) or scaffold (1) assembly level. The *C. sativa* cs10 genome (BioProject ID: PRJNA560384), which was the most recent, the best-assembled and, thus, considered the representative genome of this species, was chosen for microsatellites or simple sequence repeat (SSR) searches using MISA (MIcro SAtellites Identification Tool) [130]. The parameters were set as follows: minimum of 15 repetitions for mononucleotide motifs, 8 for dinucleotides, 5 for trinucleotides, and 4 for tetra-, penta-, and hexanucleotides. A total of 126,593 perfect and 12,017 compound SSR regions were identified, with a density equal to 148 SSRs/Mbp (0.34% of the total length of the genome). This value is slightly higher but still comparable with those found for 15 other plant genomes, including *Solanum melongena*, *Capsicum annuum*, *Nicotiana tabacum*, *Petunia axillaris*, and *Coffea canephora* by Portis et al. [131], which ranged from 60 to 140 SSRs/Mbp according to the same search parameters for SSRs [131]. Most of the SSR sequences detected in *C. sativa* exhibited a length between 15 and 19 nucleotides (60.1%), 26.5% of the sequences were 20–29 nucleotides long, 5.4% presented a length of 30–39 nucleotides and the remaining 8% were more than 40 nucleotides in length. The motif category responsible for the longest microsatellites was the dinucleotides, for which 16.7% of the sequences showed >20 repetitions and, hence, were more than 40 nucleotides long. A second and more stringent SSR analysis was performed to identify sites suitable for genotyping analysis; longer and putatively, more polymorphic sites were searched, increasing the stringency of the parameters to a minimum of 20 repetitions for mononucleotides, 15 for dinucleotides, 10 for trinucleotides, and 7 for tetra-, penta-, and hexanucleotides. The resulting 23,900 sequences were scored with a density of 28.2 SSRs/Mbp, with a total length equal to 0.13% of the genomic sequence. The most abundant

motifs identified were the dinucleotide and the trinucleotide motifs, accounting for 55.3% and 23.9% of the total length of the SSR sequences, respectively (Figure 3.1), followed by mononucleotides motifs (18.4%), while the remaining tetra-, penta-, and hexanucleotide motifs accounted for only 2.2% of the total length (with 0.6, 0.3, and 1.3% richness, respectively). The most abundant type of SSR repeat was A/T for mononucleotides (the only type of this motif), AG/CT for dinucleotides (88.7% of the total length of this motif category), and AAT/ATT for trinucleotides (84.4%).

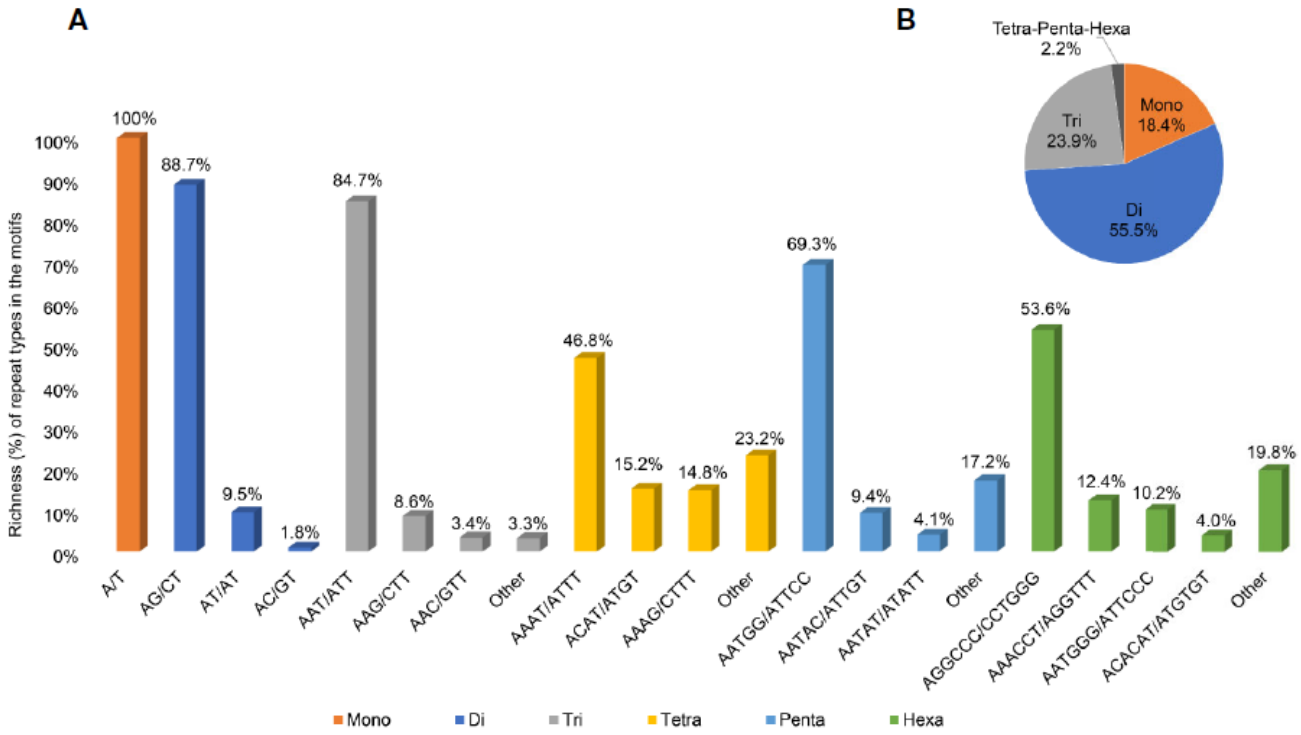

Fig 3.1: Information on SSR regions: (A) Abundance of the main repeat types of SSR sites in the Cannabis cs10 genome (Y axis represent the % base pairs among the total base pairs of each motif; information was collected on the Mono-, Di-, Tri-, Tetra-, Penta-, and Hexa- motif types). (B) Abundance of each motif among the total SSR sites.

To develop a panel of SSR loci that are exploitable for marker assisted breeding (MAB) purposes, several microsatellites were selected within each linkage group, with an average of four per chromosome, to cover the entire genome at a density equal to or greater than one SSR every 5Mb. The selection was performed taking into consideration chromosomal position, nucleotide length, and repetitive motifs. SSR-specific primer pairs were designed using the Geneious plug-in Primer3 [132] following the same criteria described by Palumbo et al. [69] and using the same parameters for all genomic loci to make multiplex PCR assays possible. The panel of markers was also developed considering their presence in a single copy to avoid nonspecific PCR products and their polymorphic nature through an in-silico comparison of cs10 with two additional genomes (Finola SAMN02981385 and Purple Kush SAMN09375800). A total of 41 SSR primer pairs were designed, with an average of four per chromosome.

3.2.2 Plant materials of Cannabis

To test these markers developed in silico Cannabis samples were needed. Fresh samples from 104 individuals representing 11 different Cannabis sativa industrial hemp varieties of high breeding value kindly provided by Gruppo Padana (Paese, TV, Italy) were collected in December 2020 and stored at -20 °C. Varieties were selected based on their geographical origin (Italy, Finland, Hungary, France, Poland and Netherlands) and their sex behavior (dioecious and monoecious), the detailed information on these varieties is presented on **Chapter 2.2.1.2:** *Genotyping experiments on hemp samples* Table 2.1.

Genomic DNA was extracted from 70–100 mg of each sample using the "DNeasy® 96 Plant Kit" (Qiagen, Hilden, Germany) following the manufacturer's instructions. The quality of the genomic DNA samples was assessed by electrophoresis on a 1% (w/v) agarose gel stained with 1X SYBR® Safe™ DNA Gel Stain (Life Technologies, Carlsbad, CA, USA) in Tris–acetate-EDTA (TAE) running buffer. The yield and purity were evaluated using a NanoDrop 2000c UV–Vis Spectrophotometer (Thermo Scientific, Pittsburgh, PA, USA). Following

DNA quantification, all the DNA samples were diluted to a final concentration of 20 ng/µl to be used as templates for PCR amplification.

3.2.3 Analysis of the SSR marker loci

The amplification reactions were performed using the three-primer strategy reported by Schuelke [133] with some modifications. Briefly, for each primer pair, universal sequences (namely, M13 and PAN1–3) were used to tag the 5' end of the forward primer and were utilized in PCR assays in combination with M13, PAN1, PAN2 and PAN3 fluorophore-labeled oligonucleotides. The fluorophores used in all amplification reactions were 6-FAM, VIC, NED and PET. As a preliminary step, 41 SSR primer pairs in-silico designed [39] were evaluated to amplify a subset of DNA samples (at least one for each population). Each PCR contained Platinum Mix 2X (Applied Biosystems, Carlsbad, CA, USA), GC Enhancer 10X, 0.25 µM tailed forward primer, 0.75 µM reverse primer, 0.5 µM fluorophore-labeled oligonucleotides, 40 ng of DNA template and H_2O to a final volume of 20 µL. Reactions were performed in a Veriti® 96 Well Thermal Cycler (Applied Biosystems®) as follows: 95 °C for 5 min; 35 cycles at 95 °C for 30 s, 55 °C (or 57 °C) for 45 s and 72 °C for 45 s; and a final extension of 30 min at 60 °C. The quality of the PCR products was assessed by electrophoresis on a 2% (w/v) agarose gel stained with 1X SYBR® Safe™ DNA Gel Stain (Life Technologies) using Tris–acetate-EDTA (TAE) running buffer. DNA fragment capillary electrophoresis was then performed with an ABI PRISM 3130xl Genetic Analyzer (Thermo Fisher) using LIZ500 (Applied Biosystems) as the molecular weight standard. Fragment analysis was conducted using Peak-Scanner V1.0 (Applied Biosystems). A subset of 20 primer pairs was selected based on their amplification capability, polymorphism and repeatability. These primer pairs were organized into 4 multiplexes (Table 3.1) and then used to amplify all 104 genomic DNA samples using the same conditions previously reported.

Locus name	Chloroplast Location	Start (position on Chloroplast)	End (position on Chloroplast)	Expected Size	Multiplex	Fluo Dye	Ta (°C)	Motif	Forward Primer	Reverse Primer
SSR_6–3	6	35,062,092	35,062,261	170–200	1	M13	55	(AAT)10	ATCTCATTTTCCGTACCTGTT	CTAATTCTCAACTTAACCGCG
SSR_2–2	2	27,019,093	27,019,345	240–300	1	M13	55	(TGA)12	TAGTAGTAGTAGTGCCTGAGG	ACCTTAACAACACCACAACTA
SSR_X-1	X	12,090,959	12,091,352	390–450	1	M13	55	(TC)40	TTGTCAAGGGAGCTTAGTTAG	ATGTGTATTTCTCGCCTGTTA
SSR_4–2	4	38,738,240	38,738,472	230–260	1	PAN1	55	(AT)17	CAGAGTTTGGTCCTTTTCAAA	CACGGATTTTAAGCATTGGAT
SSR_2–3	2	49,240,375	49,240,744	350–410	1	PAN1	55	(GA)22	CTCCCTGCCATTAGACAAATA	CCAGGAGGTAATTTTCTGCTA
SSR_7–3	7	51,776,452	51,776,692	230–280	1	PAN2	55	(CT)22	ACTGTGAACTGTCCTTTTACA	AACAACCTGAAATCCGAAAAG
SSR_3–3	3	59,258,629	59,258,880	250–300	1	PAN3	55	(AG)21	CAAAGAAAGCAGGCATTAGTT	CTCTCTGTGAATGTGATCTGT
SSR_2–1	2	15,695,145	15,695,388	240–260	2	M13	55	(AAT)11	GGCAGGAAAAATCTCAAACAT	ACATTGGAATTAGACAGAGCA
SSR_4–1	4	3,414,697	3,414,947	230–270	2	PAN1	55	(ATA)21	GTTGGTTATGTGTTAGGGTCT	GTTATGGACAAACAATGCATG
SSR_8–2	8	13,924,026	13,924,199	180–220	2	PAN2	55	(CT)21	CATCACACCAGGTACCAATAT	CATGAAACAACGTTGGGTTAT
SSR_5–2	5	34,558,385	34,558,643	250–300	2	PAN2	55	(CT)32	TGGCTGAAAGTAAGAAAAGAC	TTATCGCTCAAAACACTCAAC
SSR_6–1	6	3,764,859	3,765,058	200–270	2	PAN3	55	(AT)17	ACTTCACATGAGATTGAGAACA	TCCTTTGGATTCATTAAGTTGT
SSR_X-3	X	71,305,129	71,305,410	280–350	2	PAN3	55	(TC)41	ACAGTAGTTTTCAGGGTTGAA	TCACACCAATATCTATCAGCC
SSR_1–4	1	86,039,144	86,039,328	180–220	3	M13	55	(TTA)17	TCAAGTTACGTAATCCCCAAA	CCTAAGCACAAGGTTAAATCAT
SSR_3–1	3	12,247,530	12,247,829	300–340	3	M13	55	(TC)32	TGATTTTGCGACCCTTTTATG	CTTTTGCAGGTACATCCAAAA
SSR_8–4	8	50,925,135	50,925,396	280–330	3	PAN1	55	(TC)22	TATGCATCCATTGTACCTGTT	TAATGTTTGTGTGTGTGCAAA
SSR_9–4	9	58,895,568	58,895,670	110–150	4	PAN1	57	(CT)16	TTTCCTGCTCACCTTAAACC	AACCTATATTGAGACGAACCG
SSR_1–1	1	12,756,851	12,757,030	180–220	4	PAN1	57	(TC)33	AAACTGACAGCTTAAGCATTC	TGGGCATGTACTCTATCACTA
SSR_5–5	5	82,565,436	82,565,719	270–290	4	PAN2	57	(GA)18	AGAGGAAGGAAAGAGAGCTAT	CACGAGGGAGCCTTATTAATA
SSR_6–4	6	63,517,285	63,517,456	170–180	4	PAN3	57	(CT)30	ACGAGACTTTACAGAGAACAA	AGATAGGGAAGAACACAACAC

Table 3.1: Information on the 20 microsatellite markers validated using a core collection of high breeding value plant materials of Cannabis, positions of the loci was assessed on the *C. sativa* cs10 genome (BioProject ID: PRJNA560384).

3.2.4 Molecular data analysis

Statistical analyses for all SSR marker loci were performed using the PopGene software package v. 1.32 [134]. The observed number of alleles per locus (ne), Levene's observed heterozygosity (Ho; [135]) Nei's expected heterozygosity (He; [136]) and the average heterozygosity (Ha; [137]) were computed for each SSR locus and over all SSR markers. PopGene software was also used to estimate F-statistics [138], including the heterozygosity within (Fis) and between (Fit) subpopulations as well as the fixation index (Fst), according to Wright [139]. The phenotypic diversity of the allele profiles was estimated using Shannon's information index (I) as reported by Lewontin [140]. Gene flow (Nm) estimates among subpopulations were derived from the fixation index as described by McDermott and McDonald [141].

The genetic similarity (GS) in all the pairwise comparisons and the Principal Component Analysis (PCoA) for the two main dimensions were conducted under default settings using NTSYS software package v. 2.21c [142]. Moreover, the cluster dendrogram was built based on the unweighted pair-group with arithmetic mean (UPGMA) method by applying Dice's coefficient [143], and a bootstrap analysis was conducted with 1,000 resampling replicates [144] using Past3 software [145].

The population structure of the *C. sativa* collection was investigated using the model-based (Bayesian) clustering algorithm implemented in STRUCTURE software v.2.2 [146] which groups individuals according to marker allele combination and distribution. The number of founding groups ranged from 1 to 15, and the method described by Evanno et al. (2005) [147] was used to evaluate the most likely estimation of K. A burn-in of 2·105 and a final run of 10^6 Markov chain Monte Carlo (MCMC) steps were set. GENALEX v 6.503 [148] was used to investigate the total genetic variation by means of analysis of molecular variance (AMOVA) and to evaluate the presence of private alleles among populations. Alleles were considered private when present with at least 5% frequency in one group and absent in all others.

3.2.5 SSR Marker characteristics

Starting from 41 primer pairs amplifying an equivalent number of SSR loci, with an average of 4 markers per chromosome, all were tested on a subset of DNA samples obtained comprehending at least one sample from each variety. Of these loci, 4 did not produce any amplicon, 6 produced nonspecific products and 11 were discarded for the high number of missing data among samples. The remaining 20 loci were used to screen on the entire collection, comprised of 104 individuals belonging to 11 different hemp varieties of European origin, both monoecious and dioecious.

Table 3.2 summarizes the number of detected alleles per locus (N), the frequency of the most abundant allele (pi), the observed heterozygosity (Ho), expected heterozygosity (He),

average heterozygosity (Ha) and the polymorphism information content (PIC) of the 20 selected loci.

Additional information on the primers such as: expected size, primer annealing temperature, location on the genome and primer sequence sequences is found in *Supplementary Table S1: Multiplex SSR Primer List*.

Locus	General Statistics			H-statistics			F-statistics			Nm
	Na	pi	PIC	Ho	He	Ha	Fis	Fit	Fst	
SSR_6–3	7	0.54	0.57	0.50	0.62	0.25	0.00	0.22	0.22	0.86
SSR_2–2	7	0.55	0.58	0.42	0.63	0.21	0.16	0.37	0.25	0.60
SSR_X-1	28	0.13	0.91	0.53	0.92	0.25	0.42	0.46	0.07	1.40
SSR_4–2	11	0.43	0.70	0.22	0.74	0.10	0.73	0.76	0.10	0.69
SSR_2–3	24	0.25	0.88	0.70	0.89	0.34	0.18	0.24	0.08	1.69
SSR_7–3	17	0.21	0.88	0.83	0.89	0.41	0.03	0.10	0.07	1.67
SSR_3–3	23	0.20	0.90	0.70	0.91	0.34	0.19	0.26	0.09	1.47
SSR_2–1	7	0.77	0.35	0.07	0.38	0.03	0.74	0.84	0.37	0.23
SSR_4–1	11	0.70	0.47	0.20	0.50	0.09	0.68	0.72	0.13	0.42
SSR_8–2	20	0.17	0.90	0.84	0.91	0.39	0.07	0.16	0.10	1.33
SSR_5–2	21	0.14	0.91	0.32	0.92	0.13	0.67	0.73	0.17	0.37
SSR_6–1	14	0.18	0.86	0.35	0.88	0.14	0.64	0.69	0.14	0.78
SSR_X-3	25	0.13	0.92	0.70	0.93	0.29	0.31	0.37	0.10	0.87
SSR_1–4	8	0.62	0.56	0.57	0.59	0.26	0.16	0.21	0.06	2.78
SSR_3–1	15	0.16	0.88	0.52	0.90	0.22	0.45	0.53	0.14	1.05
SSR_8–4	17	0.16	0.88	0.41	0.90	0.18	0.56	0.61	0.12	0.78
SSR_9–4	17	0.23	0.88	0.63	0.89	0.27	0.35	0.41	0.09	1.24
SSR_1–1	18	0.23	0.87	0.36	0.88	0.14	0.63	0.69	0.15	0.63
SSR_5–5	8	0.70	0.46	0.08	0.49	0.03	0.86	0.89	0.21	0.51
SSR_6–4	3	0.67	0.40	0.38	0.47	0.17	0.29	0.42	0.17	1.05
Mean	15.05	0.36	0.74	0.47	0.76	0.21	0.41	0.48	0.14	1.02
St. Dev.	7.12	0.23	0.20	0.23	0.19	0.11	0.27	0.24	0.08	0.13

Table 3.2: Global genetic diversity statistics sorted by locus related to 84 Cannabis individuals based on 20 SSR marker loci.

N: sample size of individual genotypes; pi: frequency of the most common marker allele; Na: number of observed alleles per locus; PIC: polymorphism information content; Ho: observed heterozygosity; He: expected heterozygosity; Ha: average heterozygosity; Fis and Fit: Wright's inbreeding coefficients; Fst: fixation index; Nm: gene flow.

In total, 301 alleles were detected among 11 varieties with several observed alleles per locus ranging between 3 (SSR_6-4) and 28 (SSR_X-1). According to Botstein et al. (1980) [149], 16 of 20 SSR loci were highly informative (PIC >0.5), and 4 SSR loci were reasonably informative (0.5 > PIC > 0.25). It is likely that the high number of alleles per locus (Na) and, therefore, the resulting PIC values reflect the extension of the geographical area of varietal origin. A similar average PIC value (0.71) was obtained by Maria et al. (2015) [151] by analyzing samples collected from different European regions with 16 SSR loci. In contrast, limiting analyses to samples derived exclusively from Turkey with 22 markers [152] obtains a much lower average PIC (0.28). The frequency of the most common marker allele (pi) was low when the observed number of marker alleles was high and vice versa. For instance, SSR_6-4 and SSR_X-1, had Na values equal to 3 and 28, showed pi values of 0.67 and 0.13, respectively.

Although extremely variable (loci ranging from 0.07 in SSR_2-1 to 0.84 in SSR_8-2), the observed heterozygosity (Ho), was always lower than the expected heterozygosity (He), highlighting an excess of homozygosity equally distributed across the genome. This result was further confirmed by the positive inbreeding coefficient (Fis) values and equal to 0.41 on average (Table 3.3). Similarly, Fit (heterozygosity between subpopulations) and Fst (fixation index) indices were highly variable with values ranging from 0.10 to 0.89 and from 0.06 to 0.37, respectively. Nm (gene flow measure) ranged from 0.23 to 2.78 with an average value of 1.02 ± 0.13, showing the presence of gene flow (Nm>1) for 9 loci out of 20. Table 3.3 summarizes the statistics related to the 104 accessions analyzed in the present study, and they are grouped into the 11 varieties of *C. sativa* sampled.

Variety	N	General Statistics				H-statistics				F-statistics			Nm
		P	I	SM (%)	GS (%)	No	Ne	Ho	He	Ha	Fis	Fit	Fst	
ITA1	10	0.90	1.19 ± 0.13	65.68 ± 4.42	88.68 ± 1.71	4.90 ± 0.56	3.11 ± 0.32	0.59 ± 0.05	0.62 ± 0.06	0.26 ± 0.16	0.13	0.12	0.56	0.19
ITA2	9	0.95	1.15 ± 0.12	64.21 ± 4.80	89.16 ± 1.62	4.40 ± 0.48	3.15 ± 0.36	0.58 ± 0.05	0.63 ± 0.06	0.20 ± 0.17	0.27	0.36	0.68	0.12
ITA3	10	1.00	1.49 ± 0.09	67.96 ± 3.76	86.62 ± 1.55	6.15 ± 0.47	4.00 ± 0.36	0.70 ± 0.03	0.74 ± 0.03	0.25 ± 0.14	0.26	0.29	0.64	0.14
HUN1	11	1.00	1.15 ± 0.11	67.50 ± 5.67	89.96 ± 1.67	4.60 ± 0.46	3.03 ± 0.32	0.59 ± 0.05	0.62 ± 0.05	0.19 ± 0.13	0.31	0.40	0.70	0.11
HUN2	10	1.00	1.35 ± 0.13	67.36 ± 3.31	87.09 ± 1.47	5.75 ± 0.63	3.80 ± 0.49	0.64 ± 0.04	0.68 ± 0.05	0.22 ± 0.13	0.24	0.35	0.68	0.12
FIN	11	0.85	1.41 ± 0.17	69.76 ± 2.97	87.54 ± 1.27	6.15 ± 0.74	4.43 ± 0.55	0.63 ± 0.07	0.66 ± 0.07	0.22 ± 0.17	0.34	0.32	0.66	0.13
NED	13	0.90	1.15 ± 0.14	65.67 ± 5.88	89.37 ± 1.77	4.60 ± 0.50	3.29 ± 0.40	0.57 ± 0.06	0.59 ± 0.06	0.18 ± 0.15	0.35	0.38	0.69	0.11
POL	8	1.00	1.04 ± 0.12	64.16 ± 4.91	89.91 ± 1.34	3.95 ± 0.41	2.77 ± 0.32	0.54 ± 0.05	0.58 ± 0.05	0.21 ± 0.16	0.22	0.26	0.63	0.15
FRA1	10	0.95	1.22 ± 0.14	63.57 ± 6.13	88.27 ± 2.15	4.70 ± 0.51	3.57 ± 0.42	0.60 ± 0.06	0.64 ± 0.06	0.18 ± 0.16	0.31	0.45	0.73	0.09
FRA2	9	1.00	1.39 ± 0.12	63.36 ± 4.30	86.26 ± 1.81	5.45 ± 0.56	4.02 ± 0.50	0.67 ± 0.04	0.72 ± 0.04	0.21 ± 0.13	0.32	0.40	0.70	0.11
FRA3	3	0.65	0.56 ± 0.11	49.86 ± 9.43	91.08 ± 2.13	2.10 ± 0.25	1.89 ± 0.21	0.34 ± 0.06	0.48 ± 0.09	0.17 ± 0.16	0.02	0.09	0.54	0.21

Table 3.3: Descriptive genetic diversity statistics of the Cannabis accessions calculated for individual varieties, including heterozygosity degrees, similarity degrees, inbreeding coefficients and gene flow estimates.

Variety: code identifying the Cannabis variety; N: number of individuals per variety; P: proportion of polymorphic loci (P = npj/ntot); I: estimates of Shannon's information index of phenotypic diversity; Simple matching-based genetic similarity (SM): the genetic similarity coefficient or simple matching coefficient calculated within each variety without considering all the others; MGS: mean genetic similarity calculated using simple matching coefficient calculated considering all the varieties; No: average number of observed alleles per locus; Ne: effective number of alleles; Ho: observed heterozygosity; He: expected heterozygosity; Ha: average heterozygosity; Fis and Fit: Wright's inbreeding coefficients; Fst= fixation index; Nm= gene flow.

The number of polymorphic loci was high among all varieties analyzed (at least 90% polymorphic loci), except for the FRA3 variety. Nevertheless, it should be emphasized that this result may be affected by the reduced number of individuals considered in this variety compared to the others (only 3 in contrast to the 8–13 of the other varieties). A similar result was observed for the Shannon index (I) of phenotypic diversity and simple matching-based genetic similarity, which showed high values in all varieties (1 to 1.5 and 63 to 69%, respectively), except for the FRA3 variety.

On the other end, the mean genetic similarity (MGS) among all varieties was more uniform with values ranging between 86 and 91%. The mean number of observed and effective alleles per locus ranged from 2.10 to 6.15 and from 1.89 to 4.43, respectively. Despite the different geographical origins and the different sexual behaviors, the observed heterozygosity was comparable among all varieties, and on average, it was equal to 0.58 ± 0.09, consistent with the allogamous reproductive system. When compared to the expected heterozygosity, all varieties showed an excess of homozygosity as confirmed by Wright's statistics. Both Fis and Fit were higher than 0 in all varieties, suggesting repeated crosses of genetically related individuals and thus strong selective pressure operated by breeders. Estimates of the fixation index (Fst) changed considerably (from 0.54 to 0.73) among the 11 varieties, indicating an unbalancing contribution of the investigated populations to the total genetic variation. Accordingly, our estimates of inbreeding coefficients suggested that these varieties are characterized by a relatively high degree of genetic differentiation with approximately 64% (average Fst = 0.64) of the genetic variation found among varieties and 36% of the total genetic variation expressed within varieties. As expected, the gene flow (Nm) estimate derived from Fst was low (ranging from 0.09 to 0.21), demonstrating the absence of gene flow among the varieties.

Private alleles, beyond estimating genetic diversity, are also useful for varietal registration and traceability. The 301 alleles were screened for private markers by considering the core collections from different perspectives. The first classification was made between dioecious and monoecious samples, and the second classification considered the 11 varieties as distinct groups. In addition, the third classification was based on the geographical origin (Italy, France, Netherlands, Poland, Hungary and Finland). In the first case, 22 alleles were found to be private for dioecious samples and 13 for monoecious samples which showed significant genetic differentiation between the groups. By considering the 11 groups separately, the highest number of private alleles was found in the ITA3 variety (23 out of 123). Finally, the third division of the samples based on their geographical origin found the highest number of private alleles in the Italian, Finnish and Hungarian samples with 13 out of 175, 12 out of 124 and 10 out of 153, respectively. However, it should be noted that the frequency of each

private allele was low with few highly frequent alleles. For example, in the case of the ITA3 variety, the private allele frequency ranged from 5% (loci 8–4) to 41.7% (loci 5–2) with 16 out of 23 private alleles having a frequency ≤10%. Although this finding would discourage the use of single SSR loci for traceability or registration purposes, the combination of multiple private alleles at different loci may still represent an excellent tool for varietal identification [153]. The fact that a high number of private alleles was assigned to Italian and Hungarian clusters and varieties suggests a higher degree of isolation and local adaption (or selection) to environmental conditions for the southern samples, and at the same time, it may indicate that this area is the European domestication center of Cannabis. In fact, it is conceivable that the longer a species is present in each geographical area, the higher the number of allelic variants (albeit rare) that have accumulated.

3.2.6 Genotyping Cannabis Drug-type varieties

After testing our developed SSR markers on hemp samples we were able to verify their efficacy also on drug-type Cannabis DNA samples.

All samples, aside from the one coded as CBD-1, were collected from Canadian varieties currently used in industrial production.

The sample list is detailed in the next table (Table 3.4):

Sample Name	Identification Code	Additional Information
Lost Coast Hashplant	LCH	From drug-type TCH variety, claimed to be unrelated with all other varieties
Sapphire Scout 3	SS-3	From drug-type TCH variety, sample from the Sapphire Scout variety
Sapphire Scout 8	SS-8	From drug-type TCH variety, sample from the Sapphire Scout variety
THC-CAD-1	THC-CAD-1	From drug-type TCH variety, sample from variety claimed to be related with Sapphire Scout
The New P11	TN-P11	From drug-type TCH variety, sample from the The New P variety
The New P17	TN-P17	From drug-type TCH variety, sample from the The New P variety
White Siberian	WS	From drug-type TCH variety, sample from variety claimed to be related with The New P
THC-CAD-2	THC-CAD-2	From drug-type TCH variety, claimed to be unrelated with all other varieties
Purple Orange CBD	PO-CBD	From drug-type CBD variety with 5:10 CBD:THC ratio
CBD-1	CBD-1	From drug-trype CBD variety of non-Canadian origin

Table 3.4: Information on samples of drug-type varieties analyzed with the developed SSR markers.

All the same methods applied for the hemp-type varieties, starting from DNA extraction to PCR amplification ad finally bioinformatic data analysis, were used for the genotyping of these additional varieties,

3.3 Results and discussion

As stated, an SSR marker protocol was developed during my first PhD year and, after testing on 11 hemp varieties, the markers demonstrated to be successful in accessing heterozygosity/ homozygosity, genetic uniformity/stability, and genetic variation within and among varieties. During the third year of my PhD, we were also able to test these markers on drug-type varieties, these tests were essential with the prospect of developing new drug-type Cannabis strains with Marker Assisted Breeding (MAB).

The SSR markers were revealed to also work extremely well with the new samples, as expected the Canadian samples were shown to be significantly genetically different from the European hemp previously genotyped. In this section results for the experiment on hemp genotyping will be presented first, followed by the ones on the drug-type samples and finally, in **Chapter 3.4**: *Conclusions*, comparisons between all examined samples will be made.

3.3.1 Genotyping Hemp with SSR markers

Thanks to our developed SSR markers we were able to study the genetic structure of the germplasm collection of hemp samples, the data was investigated to infer population relationships among individuals (Figure 3.2).

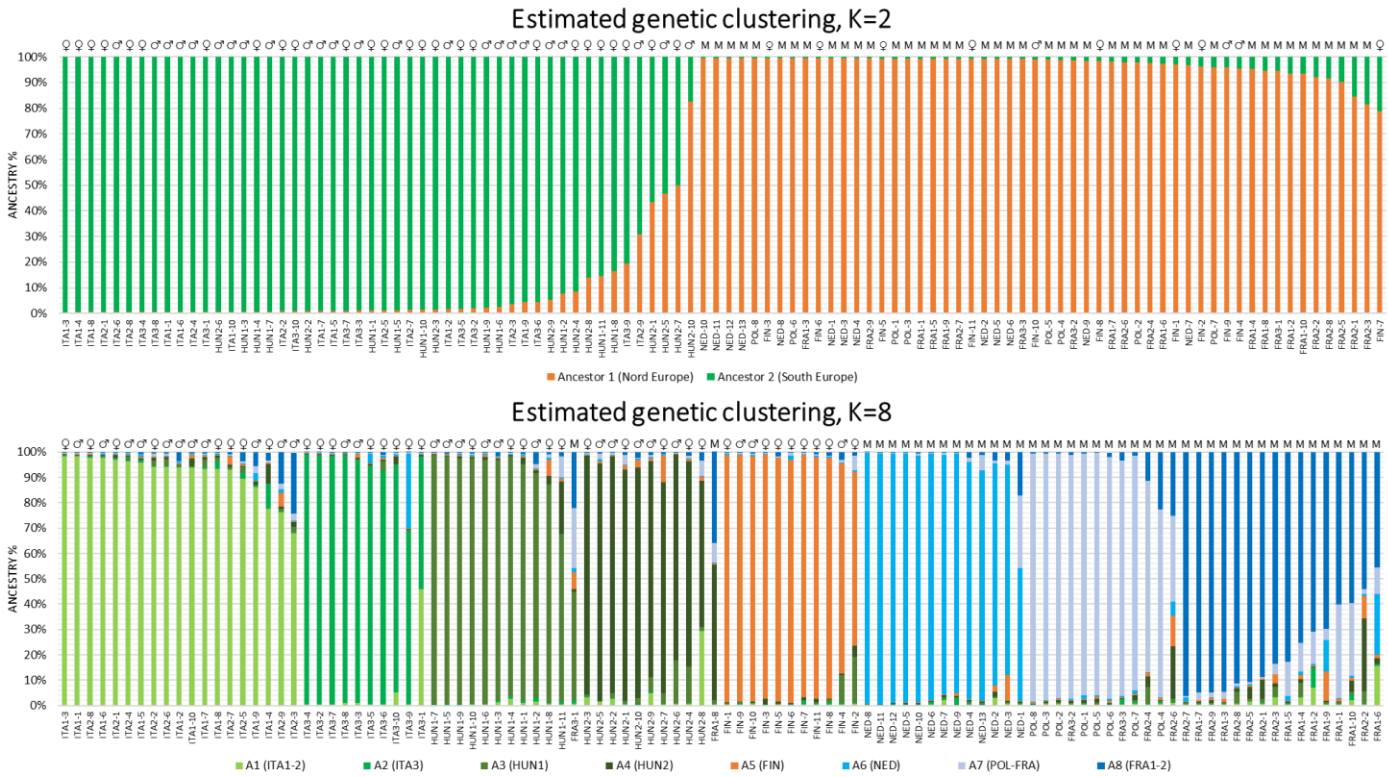

Figure 3.2: Population structure with K=2 (top) and K=8 (bottom) of the 104 Cannabis accessions based on the 20 selected SSR marker loci. Symbols: ♀: dioecious female sample, ♂: dioecious male sample; M: monoecious sample.

Following the procedure of Evanno et al. (2005) [147], a maximum ΔK value was found at K = 2 and K = 8. For K=2 (Fig. 3.2, top panel), samples were clearly divided based on geographical origin into two groups as follows: one consisting of varieties from southern Europe (Italian and Hungarian Origin) and the other from northern Europe (France, Poland, Netherlands and Finland). The clustering of genotypes revealed that 92 of 104 samples showed strong ancestry association (>90%), and on average, the membership of the southern and northern samples to their ancestral group was 92.3% and 96.9%, respectively. In the southern samples, 12 showed admixed ancestry, and in particular, HUN2-10 was the only sample with a higher percentage membership in the opposite group (northern group). Given the geographical distances between these two groups, the most likely hypothesis is that the

admixed samples are the intended result of hybridization events among southern and northern genotypes performed in the past by impassionate amateurs or expert geneticists to introduce genetic variability and constitute new lines [39]. It should also be emphasized that the subdivision into two large groups is not only geographic but also based on the reproductive system. In fact, the southern group contained only dioecious plants (i.e., both male and female plants), while the northern group contained only monoecious plants with few exceptions (FIN group).

Historical information on the introduction of Cannabis in Europe is vague. If the center of origin is in Central or East Asia [154], the first data certifying its presence in Europe date back to approximately 10,000 years ago. In particular, paleobotanical evidence shows the presence of wild Cannabis/Humulus-type pollen in the southeastern part of Europe and, in particular, in Hungary, Romania and Bulgaria approximately 10200–8500 years BP [154] as well as in Italy (Lake Albano, Rome) 11000 BCE. Although the most accredited hypothesis [11] suggests that dioecious plants (such as those belonging to the South group in this study) descended from monoecious plants (such as those belonging to the northern group). However, our molecular data along with the paleobotanical evidence mentioned above discourage the hypothesis that the Italian-Hungarian group evolved from the Northern Europe group. It is more likely that the introduction of Cannabis in northern Europe and the Mediterranean basin occurred independently at different times. The clear division between the two European clusters was also clearly visible with PCoA analysis (Figure 3.3), in which the southern samples were positioned in the top-right part and the northern samples were positioned in the bottom-left part.

When analyzing the germplasm collection with an additional level of population structure (K=8), sample clustering by geographical area becomes even more detailed. Plant materials from ITA1 and ITA2 were clearly assigned to a single, distinct cluster by STRUCTURE software (Fig. 3.2, bottom panel) with an average membership value of 91.3%. The PCoA (Fig. 3.3) analysis further confirmed this finding with these two varieties visibly separated from the others and occupying the top-right quadrant.

Figure 3.3: Principal component analysis (PCoA) displaying centroids of individual plants across all Cannabis varieties plotted according to the first two dimensions based on the 20 selected SSR marker loci. The same coloring scheme was used as in Fig 3.2 for the 8-population division of samples.

Instead, the third group of Italian varieties (ITA3) formed a separate cluster in the structure analysis. This sample group showed the lowest levels of similarity with all the varieties analyzed and clustered separately in the PCoA (Fig. 3.3). Additionally, the average similarity values with ITA1 (85.4%) and ITA2 (85.1%) were lower than that calculated between ITA1 and ITA2 (88.4%). Since the introduction of Cannabis in Italy, it could be hypothesized that the ITA3 group has remained moderately isolated both from the ITA1/ITA2 group and from

all other European varieties. This hypothesis was further supported by the number of private alleles found in ITA3, which was the highest among the 11 varieties.

In particular, the HUN1 and HUN2 dioecious samples were assigned to two different clusters by STRUCTURE software. Members of the HUN1 variety showed, on average, higher membership percentages to the cluster (91.9%) compared to the HUN2 ones (85.3%), which was clearly visible in the PCoA with the HUN1 samples tightly clustered in the bottom-right quadrant and the HUN2 samples dispersed around the previous ones. Similarly, samples from Finland (FIN), Poland (POL) and Netherlands (NED) were assigned to three different clusters with an average membership equal to 92.6%, 93.2% and 91.2%, respectively. The results observed for the three French varieties (FRA1, FRA2 and FRA3) were unexpected. FRA1 and FRA2 formed a cluster apart from an important share of admixed samples and a consistent representation of the Polish ancestor, suggesting hybridization events between French and Polish genotypes. Within the FRA3 variety, FRA3-1 resulted admixed with uncertain ancestry, while FRA3-2 and FRA3-3 clustered with high ancestry membership (96.4% and 93.1%, respectively) with the Polish cluster. The close phylogenetic relationship existing between French and Polish samples was also evident from the PCoA and from the similarity matrix. The three French groups showed an average degree of similarity with the POL group oscillating between 86.6% (FRA2) and 88.3% (FRA3). In support of this, a recent study conducted on the occurrence of the CBDA-synthase gene (CBDAS), THCA-synthase gene (THCAS) and two CBDAS pseudogenes across 110 Cannabis accessions has demonstrated the tight relationship existing among three French and two Polish genotypes, suggest a recent common ancestor [155].

Overall, this is the first study in which a robust and novel panel of SSRs has been successfully used to investigate the geographical origin of samples derived from different European countries. Previous studies have mainly concerned Asian accessions [93], while the only other study focusing on accessions collected from the same countries considered here (i.e., Italy, Hungary, France and Finland) failed to reconstruct the geographical origin of the samples [156].

3.3.2 Genotyping Drug-type Cannabis with SSR markers

With the analysis of our SSR marker data on ten drug-type Cannabis samples we were able to observe that our markers proved to be not only functional but also informative on these varieties too.

The obtained results, which can be summarized thanks to the Genetic Similarity Matrix (Table 3.5) and the Principal Component Analysis (PCoA) (Table 3.5) were found to be in line with the claimed origins and supposed relationships between samples.

Drug-type Cannabis samples Genetic Similarity Matrix (SM)										
Lost Coast Hashplant	X									
Sapphire Scout 3	64.04%	X								
Sapphire Scout 8	60.67%	86.24%	X							
THC-CAD-1	60.67%	78.64%	76.70%	X						
The New P11	64.04%	75.23%	77.98%	71.84%	X					
The New P17	69.66%	71.56%	74.31%	66.02%	85.32%	X				
White Siberian	70.59%	67.68%	69.70%	64.49%	76.77%	80.81%	X			
THC-CAD-2	50.56%	65.14%	60.55%	63.96%	64.22%	56.88%	58.88%	X		
Purple Orange CBD	56.63%	65.05%	68.93%	61.90%	69.90%	69.90%	75.25%	56.76%	X	
CBD-1	65.17%	65.14%	64.22%	63.06%	71.56%	66.06%	68.22%	57.26%	62.16%	X
	LCH	SS-3	SS-8	THC-CAD-1	TN-P11	TN-P17	WS	THC-CAD-2	PO-CBD	CBD-1

Table 3.5: Genetic Similarity Matrix (GSM) between ten Drug-type Cannabis samples analyzed with our developed SSR molecular markers, the matrix was obtained with simple matching (SM).

As can be visualized in Table 3.5, highest scores of genetic similarity were observed between the Sapphire Scout 3 / Sapphire Scout 8 and The New P11 / The New P17 samples with 86.24 and 85,32% respectively. Results were also moderately high between THC-CAD-1 compared to the Sapphire Scout samples (with 77 to 79 % similarity) and the White Siberian samples compared to “The New P11/17” ones (77 to 81 % similarity). Close relationships between these samples were also indicated by the owner of the plant collection and are also clearly visible in the 3D Principal component analysis (PCoA) (Fig 3.4) where they cluster together.

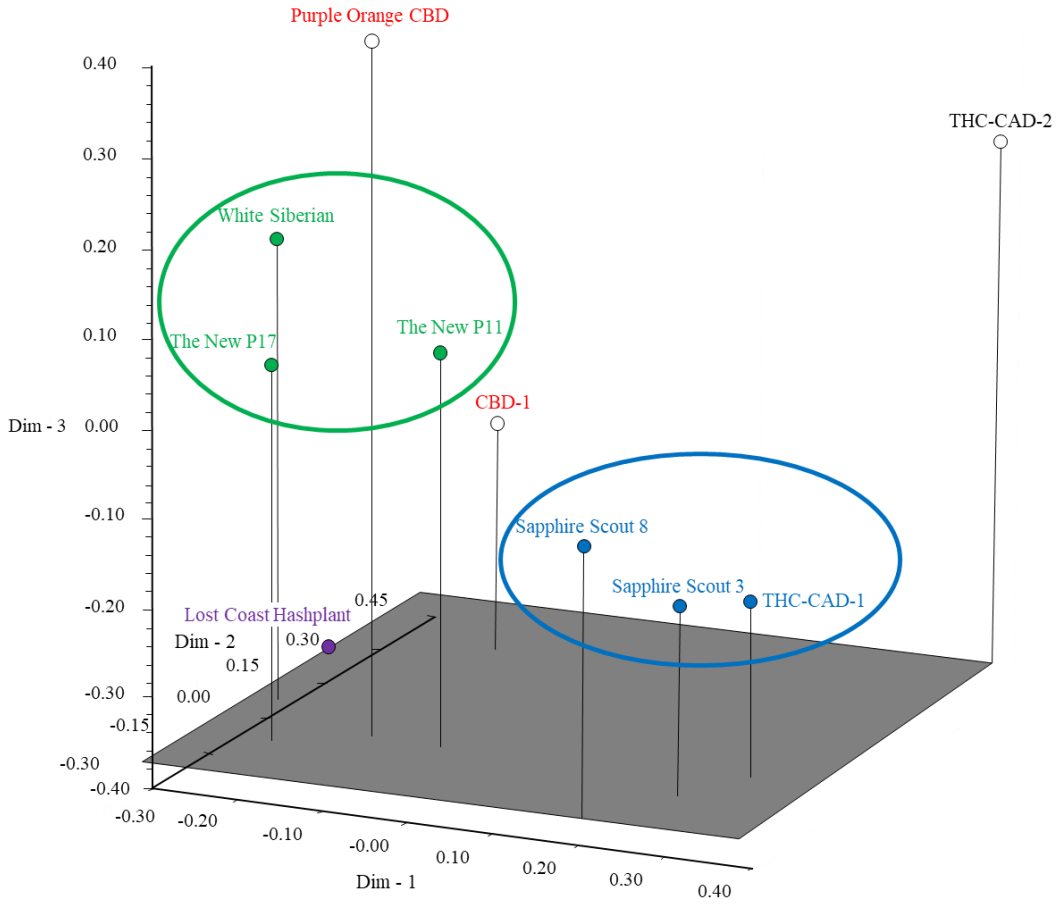

Fig 3.4: Principal component analysis (PCoA) displaying centroids of individual plants across all Cannabis varieties plotted according to 3 dimensions based on the 20 SSR marker loci utilized for the analysis. Samples within the Green and Blue were found to have the highest scores of genetic similarity. Samples from CBD-dominant varieties (in red) and the remaining two (THC-CAD-2 and Lost Coast Hashplant) were found to be genetically distant from all other samples.

Samples from the Sapphire Scout 3/8 and THC-CAD-1 cluster (colored in blue) occupy the bottom right part of the graph whereas else the White Siberian and The New P11/17 samples (colored in green) clustered together in the left part of the graph.

All other samples, including two from CBD dominant varieties (in red), were shown to be more genetically distant from all others (with 60 to 70 % maximum similarity); this again was also concordant with the information we were given on their origins and relations.

3.4 Conclusions

Our research focused on the implementation and validation of an informative, reliable, and reproducible multilocus genotyping systems in *C. sativa* that may be useful both for marker-assisted breeding of new varieties and for unequivocally identifying specific varieties already on the market or assessing their genetic identity. This is also crucial for protecting plant breeders' rights and Cannabis derivatives' consumers.

All analyzed samples (hemp and drug-type) have been included in the PCoA plot in Figure 3.5, all drug type varieties (in red) were significantly genetically distant and distinguishable from all hemp varieties, which was to be expected but still important to confirm.

All drug type samples were clearly clustered in the left portion of the plot where none of the fiber-type samples were present.

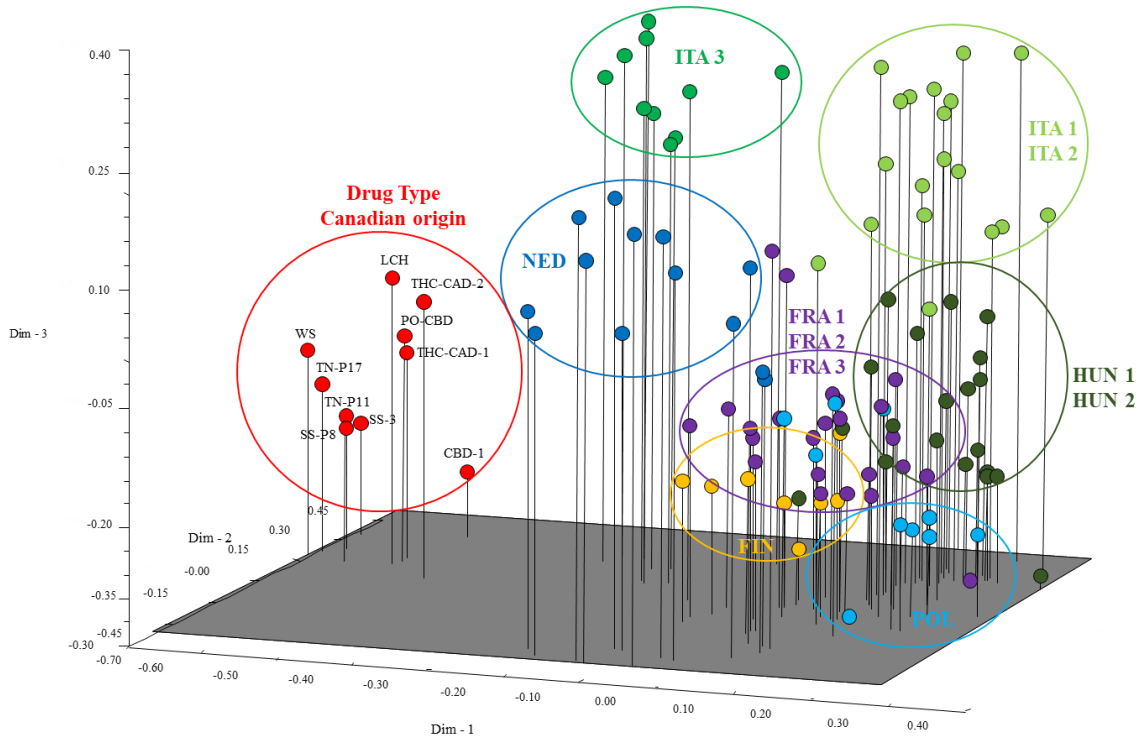

Fig 3.5: Principal component analysis (PCoA) displaying centroids for all analyzed Cannabis samples plotted according to 3 dimensions based on the 20 SSR marker loci utilized for the analysis.

Some varieties, or groups of ones, such as the ITA 1 - ITA 2 group or the NED and ITA 3 varieties have their samples clearly clustered together and distinguishable from all others. Samples from the varieties originating from France, Netherlands and Poland are instead more dispersed and mixed between each other.

In conclusion, this genotyping protocol proved to be successful in differentiating between varieties and this information could be exploited for planning crosses and predicting heterosis as an expression of hybridity in experimental F1 populations based on the allelic divergence and genetic distance of the selected parental lines.

These results were also compared with further genotyping analyses conducted with a chloroplast sequencing protocol as shown in the following chapter (**Chapter 4:** *Chloroplast sequencing markers in Cannabis sativa*).

CHAPTER 4: Chloroplast sequencing markers in Cannabis sativa

4.1 Introduction

Following the establishment of the SSR markers genotyping protocol described in the previous chapter, our focus shifted to finding a means to compare our samples with others documented in the literature. While such a comparison was unfeasible using our SSR panel, it became achievable through the analysis of sequenced chloroplast polymorphic regions.

The development of chloroplast markers may help in deciphering the geographical origin of Cannabis varieties, using DNA sequencing of maternally transmitted markers in the chloroplast genome. This data can help us identify statistically different maternal genetic backgrounds, which contributes to authentication, and improvement of breeding programs for *C. sativa*. To test this method sequencing was conducted on both hemp- and drug- type Cannabis samples with remarkable results.

To create an informative chloroplast sequencing protocol, we searched for chloroplast polymorphic regions in Cannabis literature, finding seven regions where polymorphic loci were clustered.

Studies on the subject had already started in the early 2000 the finding of intraspecific variation in *C. sativa* found in intergenic regions of the Chloroplast DNA [157].

In 2015 genetic identification of *C. sativa* using the chloroplast gene trnL-F were performed and sequencing was used to discriminate *C. sativa* DNA samples from other members of the Cannabaceae family [158].

In the last years whole plastome [128,159,160] and gene-specific sequencing [161–163] studies in numerous Cannabis samples found seven main hotspot regions for polymorphic locii: trnK-matK-trnK, rps16, trnS-trnG, ycf3, accD-psaI, clpP, and rpl32-trnL.

These regions were then investigated with the development of a primer panel for Sanger Sequencing which was first developed *in silico* and later tested on hemp- and drug- type Cannabis samples.

The results from the chloroplast sequencing analysis shown in this chapter were presented at the Cannabis Scientific Symposium: *From Plants to People* hosted by the QAQCC and the McGill Research Centre for Cannabis In Montreal, Qc, Canada, May 4-5, 2023; where I participated as a speaker to the congress with a presentation in Session 1-Plants on: "Overcoming challenges in Cannabis sativa breeding".

4.2 Materials and Methods

The in-silico development of PCR primers for Sanger Sequencing for the Cannabis plastome started from identifying the polymorphic regions reported in literature [161–163]. The investigation on the chloroplast circular DNA was done by using the Geneious Prime® 2023.2.1 software and, as a chloroplast reference, the Cannabis variety of Yoruba Nigeria (NC_027223) was chosen.

Sanger Sequencing primers require specific characteristics such as: small PCR amplification products (usually under one-thousand base pairs) and closeness between sequencing primer and target sequence of interest for maximum sequencing quality.

4.2.1 Chloroplast Sequencing primer design

To obtain a collection of primers with the same melting temperatures and capable of sequencing the maximum number of polymorphic loci with the minimum number of Sanger Sequencing reactions necessary, new primers were developed in silico based on the work by Roman et al. 2019 [164].

28 new PCR primers were designed for the Sanger Sequencing of 49 polymorphic loci located in these seven regions. The localization of the regions and loci can be visualized in the following figures (Fig 4.1), on the chloroplast circular DNA (left) and within the polymorphic regions (right).

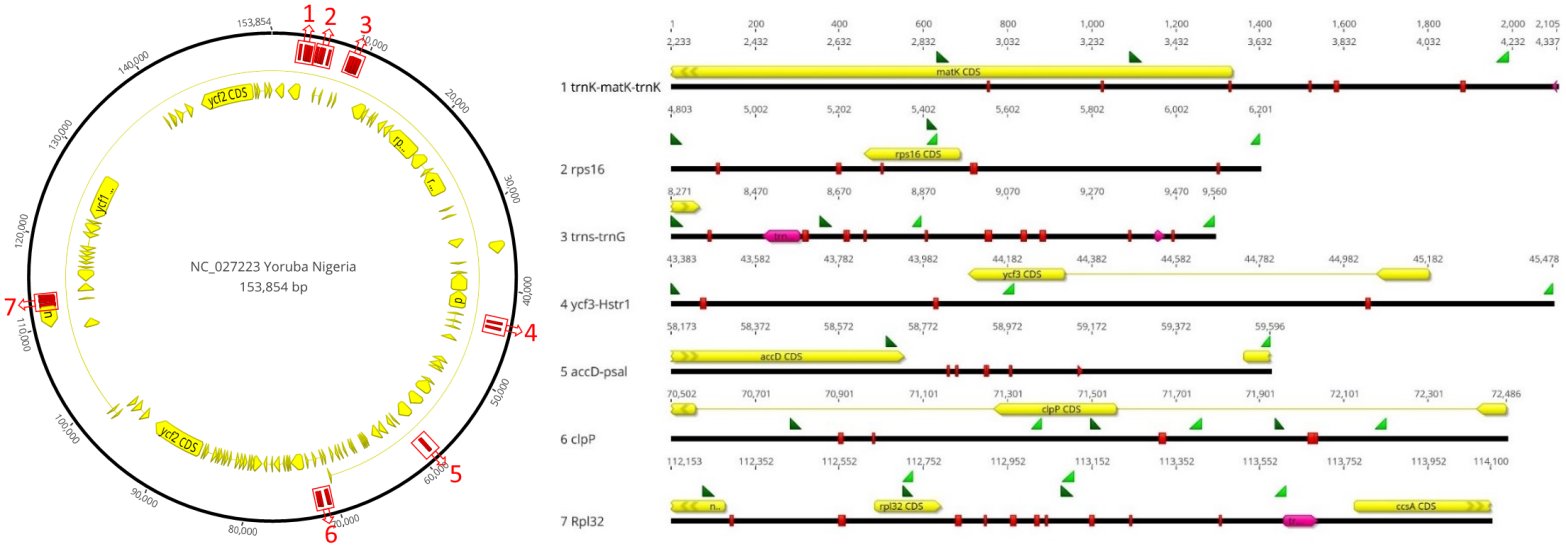

Figure 4.1: on the left: the polymorphic regions are highlighted on the chloroplast circular DNA by red rectangles; on the right: the position of primers (in green directional triangles) and pololymorfic loci (in red rectangles) located in the seven regions, coding sequences are represented by yellow rectangles.

The primer list is presented in *Supplementary Table S2: Chloroplast sequencing primer list*. To test this sequencing method, samples from eleven hemp-type Cannabis varieties (**Chapter 2.2.1.2***: Experiments on hemp samples* Table 2.1) and nine more of Canadian origin were tested (for more information on the Canadian varieties refer to Table 2.2 in **Chapter 2.2.2***: Drug-type Cannabis cultivation*).

4.2.2 Chloroplast Sequencing primer testing

For all samples DNA was extracted as described in **Chapter 3.2.2**: *Plant materials of Cannabis* and then PCR amplification was performed with the following components: 0.2 µl High fidelity Polymerase (Thermo Scientific ®), 5 µl 10X Buffer, 2 µl $MgSO_4$, 1 µl dNTPs, 2 µl Primers (Forward and Reverse mix 20µM), 4 µl DNA (20 ng/µl) and finally 35.8 µl of H_2O to a final volume of 50 µL.

PCR reactions were performed in a Veriti® 96 Well Thermal Cycler (Applied Biosystems®) as follows: 94 °C for 5 min; 35 cycles at 94 °C for 15 s, annealing with a 60° to 56° touchdown for 60 s and 72 °C extension with time depending on product size. Five cycles of touchdown from 60 to 56 °C followed by 30 cycles at 56 °C were utilized to increase specificity of the reaction.

The quality of the PCR products was then assessed by electrophoresis on a 2% (w/v) agarose gel stained with 1X SYBR® Safe™ DNA Gel Stain (Life Technologies) using Tris–acetate-EDTA (TAE) running buffer.

Sanger Sequencing was then performed with a 3730xl DNA Analyzer (Thermo Scientific ®) and results were analyzed with the Geneious Prime® 2023.2.1 software. Sequences were trimmed to obtain only high-quality sequencing products and results were aligned between each other and others already deposited in literature that were complete for all loci and with an indicated geographic origin. This made possible to include the following samples:

- AK Royal Automatic (MK878538),
- Brazuka (MK878537),
- Cheungsam(KR184827),
- Dagestani (KR779995),
- Chinese Hemp (KY084475),
- sd130 (MH118118),
- Yunma 7 (MW013540),
- Carmagnola (NC_026562),
- Yoruba Nigeria (NC_027223) [159,160].

Two *Humulus* genome chloroplasts sequences were also used as related outgroups for the analysis: Humulus Hallertauer (MG573060) and Humulus japonicus (NC_064346).

It must be noted that all samples from the previously analyzed (with SSR markers) varieties resulted to have identical sequencing results within the varieties, therefore each variety, or group of identical ones, will be considered as one sample for the following analysis.

For example, in some cases, like between two Italian and two French hemp-varieties, all sequenced samples were identical, and so considered as one.

In total 21 different sequences, each representing a single sample, or multiple identical ones have been obtained starting from 31 sequenced varieties; these sequences were sourced from the literature (9), from hemp-type varieties cultivated at Gruppo Padana (11) or from Canadian varieties (11).

After doing this, for each variety or group of identical ones the seven polymorphic regions were concatenated and aligned with all others with Geneious Prime® 2023.2.1 software (Figure 4.2).

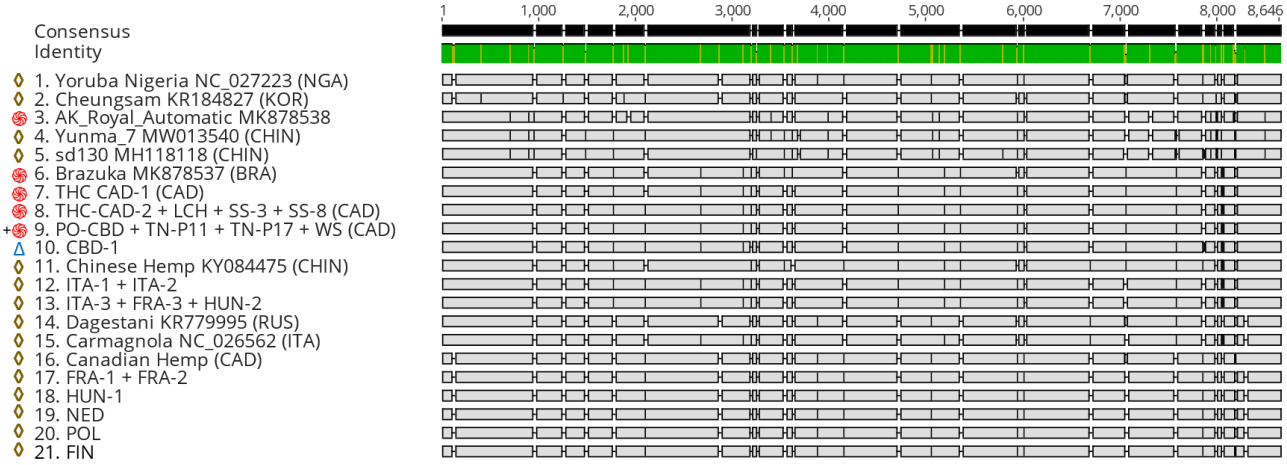

Figure 4.2: alignment between all sequenced Cannabis chloroplast polymorphic regions, polymorphic loci are visible as gaps in the consensus. Samples are labeled as THC-Drug Type (), CBD-Drug Type (Δ) or fiber-type (◊).

The developed primers proved successful in sequencing the investigated loci and polymorphisms were observed between all samples are represented by gaps in the consensus in Figure 4.2.

4.3 Results and discussion

The alignment data between all our sequenced samples' regions with their 49 polymorphic loci was used for a phylogenetic analysis investigation.

This alignment was used to generate a genetic tree with the Geneious Tree Builder tool with the Geneious Prime® 2023.2.1 software using Tamura-Nei as a genetic distance model and UPGMA as a tree building method.

In the resulting phylogenetic tree distances between samples were directly correlated with branch length on the x-axis (Fig 4.3), as an example all samples in the central orange cluster (group 2) comprised only of hemp-type Cannabis samples were found to have very similar sequences, where only 1 to 10 nucleotides differences were detected out of the 8647 that were aligned, the only exception for the group was the Cheungsam variety from South Korea which had on average around 20 nucleotides differences compared to the other samples in the cluster. Complete information on nucleotide differences between the samples' chloroplast sequencing can be found in: *Supplementary Table S3: Nucleotide differences between Chloroplast polymorphic regions*.

In the obtained Phylogenetic tree, three significantly different groups of samples can be observed, high BOOTSTRAP values were assuring the branching into the three clusters.

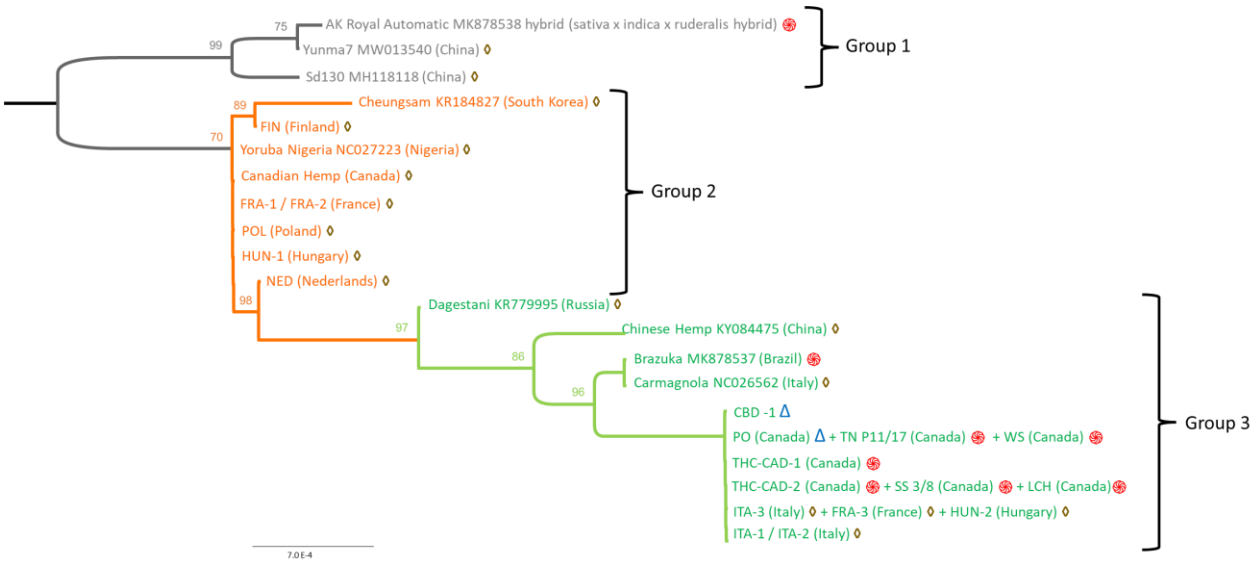

Figure 4.3**:** Genetic tree resulting from the alignment of chloroplast polymorphic regions; Samples are labeled as THC-Drug Type (), CBD-Drug Type (Δ) or fiber-type (◊).
Three main branches, supported by high bootstrap values can be identified: In the top (group 1) there are 3 samples originated from China (in grey), the middle (group 2) comprehends only hemp samples (in orange) and, on the bottom (group 3) hemp and drug-type samples are present.

Comparing the 3 resulting groups with the geographical origin it can be seen that the first group was clearly assigned to samples claimed to be originating from the Yunman region in southern China, the second group instead was comprised only of hemp samples from northern Europe, Canada, Nigeria, and South Korea and finally the third group included hemp samples from southern Europe, Southern Russia and China but also drug-type samples from Brazil and Canada.
These results showed that all 31 Cannabis varieties initially analyzed were derived from only 3 main maternal ancestors: the first was clearly associated with a geographical origin (Yunman region in southern China), the second was only correlated with hemp-type Cannabis samples and the third was interestingly shared by both hemp and drug-type varieties.

4.3.1 Finding a MatK polymorphism between Cannabis varieties.

While comparing the sequenced chloroplast polymorphic regions an unexpected polymorphism located in the coding sequence of a Maturase K protein was found.
Whilst all common polymorphic loci studied occurred in intergenic (and so non-coding) sequences, one, found in nearly half of the sequenced samples, consisted in a gap of 6 nucleotides in a intron for the Mat K gene.
As this mutation is in a coding sequence, the resulting protein will have an insertion or deletion of two amino acids between different Cannabis varieties as can be seen in Fig. 4.4 resulting in the existence of two different Mat K protein variants in *C. sativa*, one with 507 aa and the other with 509 aa.

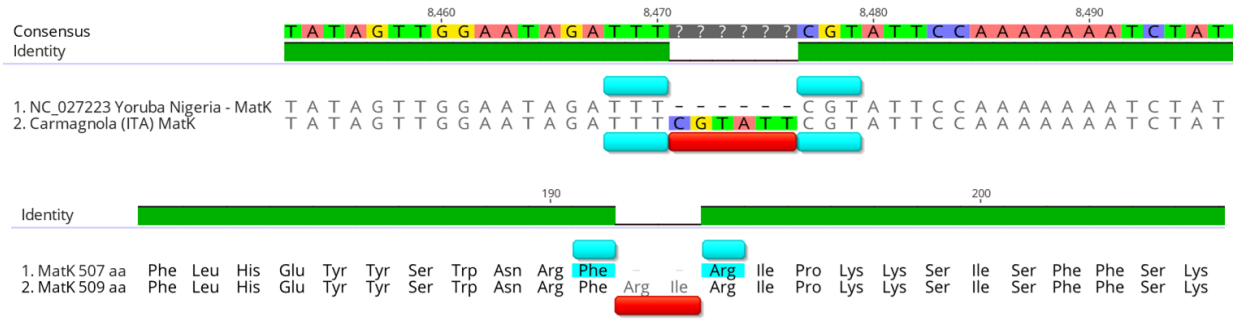

Figure 4.4: Top = Mutation on the MatK gene coding sequence, 6 nucleotide in/del was detected between the target varieties; Bottom = The mutation results in the difference of two amino acids in position 191 between variants: Arginine and Isoleucine.

By locating the mutation on a predicted 3D structure of the MatK protein with UCSF Chimera 1.16 software [165] we were able to observe that it would result in the modification of an alfa helix secondary structure as can be observed in Figure 4.5.

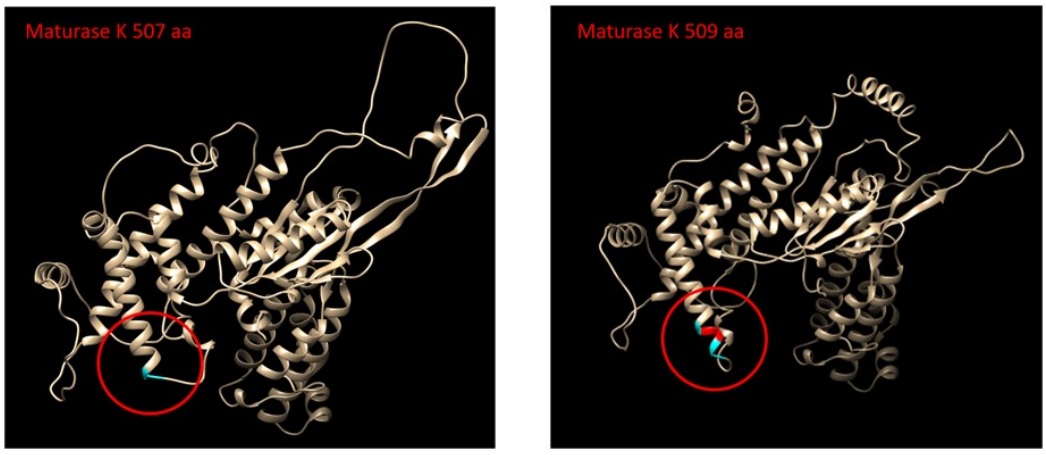

Figure 4.5: predicted 3D model structure of the Maturase K protein with 507 (on the left) or 509 (on the right) amino acids caused by the found mutation.

The presence of the 509 aa variant has not yet been reported for *C. sativa* where else the 507 aa MatK has been [166] (https://www.uniprot.org/uniprotkb/Q95BY0/entry).

At this stage it cannot yet be known if this mutation could alter the Maturase K function of RNA processing, this could be very noteworthy as this mutation was detected in nearly half of the sequenced samples, importantly it was exclusively present in fiber-type varieties but not universal for all of them, it was unique for all samples in the Group 2 cluster (Fig 4.3) which were the: Cheungsam, FIN, Yoruba Nigeria, Canadian Hemp (CAD), FRA-1, FRA-2, POL, HUN-1 and NED samples.

4.4 Conclusions

We were able to observe that the developed chloroplast markers proved successful in calculating genetic distances between varieties and assigning them to phylogenetic clusters. These results will be confronted and compared with the ones from the SSR markers genotyping analysis in **Chapter 6:** *Conclusions*.

Chloroplast sequencing couldn't detect any differences between samples belonging to the same variety or closely related ones, but significant variations were discovered between more genetically distant samples; among these a mutation on the MatK gene coding sequence could have relevant implications on the protein functions.

Sequencing additional samples especially from more Drug Type Cannabis and landraces of proven origin would be extremely useful for increasing the usefulness of the analysis for traceability tests and breeding programs.

CHAPTER 5: Flower development studies

5.1 Introduction

Cannabis is an increasingly important crop with several applications, when grown for cannabinoid extraction only female plants are cultivated due to the high concentration of glandular trichomes on their flowers [167]. These glandular trichomes are the site of cannabinoid biosynthesis and storage [168]. Cannabidiolic acid (CBDA), Tetrahydrocannabinolic acid (THCA), and cannabichromeneic acid (CBCA) are the three most abundant cannabinoids found in Cannabis [61,109,169].

In this chapter, flower development in *Cannabis sativa* will be investigated not only to study cannabinoids expression pathways and accumulation in glandular trichomes of female flowers but also to investigate the sex-determination mechanism in *C. sativa*.

Flower development was studied by integrating two approaches: specific transcriptomic and untargeted metabolomics analysis.

For the specific transcriptomic analysis, we monitored in three fiber-type varieties (HUN-1, HUN-2 and ITA-3) the pattern of:

- genes of the cannabinoid synthesis pathway at different development stages;
- genes putatively involved in the ethylene pathway.

Starting from comparisons between RNA-seq reads experiments PRJNA669389 [170] and PRJNA549804 [171] these genes were claimed to be involved in the sex-reversion molecular mechanism in *C. sativa* [172]. To verity this hypothesis the transcription expression levels of the target genes was confronted between flowering samples collected from Female Flowers (FF), Male Flowers (MF) and chemically Induced Male Flowers (IMF) [58].

Untargeted metabolomics analysis and cannabinoid identification and quantification experiments were conducted on the same flowering samples by an Ultra-Performance Liquid Chromatography–High-Resolution Mass Spectrometry (UPLC-HRMS).
All transcriptomic and metabolomics experiments were performed at the Department of Biotechnology of the University of Verona where, for legal reasons, only hemp flowering samples could be analyzed.

5.2 Materials and methods

5.2.1 Flowering samples

Flowering samples for all the following experiments have been collected from hemp-type Cannabis plants as described in **Chapter 2.3.2***: Flowering experiments*.
All samples were collected and subsequently immediately powdered in liquid nitrogen (-80 °C) using a mortar and pestle. At least 200 mg of powder for each sample was stored in -80 °C for future metabolomics and transcriptomic experiments.
The hemp varieties, HUN-1, HUN-2 and ITA-3, previously analyzed by SSR and chloroplast sequencing genotyping methods, were selected for these studies *(Table 2.1,* **Chapter 2.2.1.2:** *Genotyping experiments on hemp samples)* because the two Hungarian varieties HUN-1 and HUN-2 belong to the same genetic cluster (as resulted from our SSR analysis), while the ITA-3 variety because significantly genetically different from the first two.
The 3 selected are dioecious, essential for sex determination studies and they have controllable flowering conditions: they all flower by reducing the photoperiod from 20h to 12h of light per day.
For the metabolomics and cannabinoid pathway transcriptomic experiments female flower samples were collected from the HUN-1, HUN-2 and ITA-3 varieties at 14, 21 and 28 days after flower initiation, additional flowering samples from the HUN-1 variety were also collected at 35, 49 and 63 days after flower initiation to study cannabinoids accumulation.

Instead, for the ethylene pathway transcriptomic experiment samples were collected only at 14 days after flower initiation from the same varieties (HUN-1, HUN-2 and ITA-3). Samples included Female Flowers (FF), the same as those analyzed in the previous experiments, but also of Male Flowers (MF) and Induced Male Flowers (IMF); IMF were obtained with a Silver ThioSulfate (STS) treatment as described in **Chapter 2.3.2.1:** *Sex-reversion experiments*.

5.2.2 Metabolomic analysis of flowering *C. sativa* samples

5.2.2.1 Sample preparation for untargeted metabolomics analysis

Cannabis flowering samples were collected and powdered in liquid nitrogen using a mortar and pestle. About 100 mg frozen powder was extracted with 50 volumes (w/v), i.e. about 5 mL, of LC-MS grade methanol (Honeywell, Seezle, Germany). The tubes were vortexed for 30 s and sonicated in an ice-cold water bath at 40 kHz in a Sonica Ultrasonic Cleaner ultrasonic bath (SOLTEC, Milano, Italy) for 15 min. The samples were then centrifuged at 14,000 × *g* for 10 min and the supernatant was recovered. Prior to the metabolomics analysis with Ultra-Performance Liquid Chromatography–High-Resolution Mass Spectrometry (UPLC-HRMS), the extracts were diluted 1:5 (v/v) with LC-MS grade water and passed through Minisart RC4 filters with 0.2 µm pores (Sartorius, Göttingen, Germany).

5.2.2.2 UPLC-HRMS untargeted metabolomics analysis

To allow the separation and detection of the metabolites extracted from *C. sativa* flowers, 1 µL of each diluted extract was injected in a UPLC-PDA-ESI-QqToF system.

The UPLC consisted of an ACQUITY I CLASS system (Waters, Milford, MA, USA) equipped with a reverse phase BEH C18 column (2.1 mm × 100 mm, 1.7 µm) maintained at 30 ◦C. The reverse-phase chromatography was carried out by establishing a gradient of a mobile phase composed of water acidified with 0.1% v/v formic acid (solvent A) and acetonitrile (solvent B).

The gradient previously proposed by Aizpurua-Olaizola *et al.* (2014) [173] was adopted to separate and profile the phytocannabinoids in the samples according to the following operating conditions:

I. initial condition 50% B;
II. gradient till 15 min from 50% B to 100% B;
III. isocratic 15-17 min 100% B;
IV. gradient 17-19 min from 100% B to 50% B;
V. isocratic 100 % B till 22 min (re-equilibrium to restore the initial condition)

The flux of the mobile phase was set at 0.250 ml/min.

Another analysis was performed to profile other medium polar specialized metabolites present in *C. sativa* samples by using the following gradient:

I. initial condition 1% B;
II. isocratic 1% B for 1 min;
III. gradient 1-10 min from 1% B to 40% B;
IV. gradient 10-13.5 min from 40% B to 70% B;
V. gradient 13.5-15 min from 70% B to 90% B;
VI. gradient 15-16.5 min from 90% B to 100% B;
VII. isocratic 100% B till 20 min;
VIII. gradient 20-20.1 min from 100% B to 1% B;
IX. isocratic 1% B till 25 min (re-equilibrium to restore the initial condition)

The flux of the mobile phase was set at 0.350 mL/min.

The UPLC system was linked to an Acquity PDA detector (Waters) followed by a Xevo G2-XS qTOF mass spectrometer (Waters). The PDA carried an eLambda detector able to detect the absorbance of the compounds eluting from the column within the range 190-800 nm (UV-vis). The mass spectrometer was equipped with an electrospray ionization (ESI) source operating in either positive or negative ionization modes. The scan range was set to 50–2000

m/z and the scan time at 0.3 s. Argon gas was used to perform CID (collision induced fragmentation) and the collision energy was fixed at 35 V. In order to check the accuracy of the mass spectrometer, a solution of 100 pg/µL leucine-enkephalin was injected with a flow rate of 10 µL/min and generating a signal of 556.2771 *m/z* in positive mode and 554.2615 *m/z* in negative mode. Masslynx v4.1 software (Waters) was used to manage all the UPLC, PDA and ESI/QqTOF instrument functions. The raw MS data files were processed using Masslynx v4.1 software (Waters).

Authentic standards of CBG, CBGA, THCV, THCVA, THCA-A, CBC, CBCA, CBD, CBDA, CBN, CBNA, CBL, CBDV and CBDVA were analyzed in the same operative conditions to allow their identification within the *C. sativa* samples by comparison of their retention times, *m/z* and fragmentation patterns.

The identification of other specialized metabolites was performed by using the information collected in a proprietary library of authentic standards and in public spectral databases (e.g. MassBank, Human Metabolome Database, Pubchem, Lotus Database, etc.).

The extrapolation of peak area values for the relative comparison of the levels of the identified metabolites was performed through the software Progenesis QI (Waters).

CBDA and CBGA were quantified in UV-vis with PDA by building a calibration curve of their corresponding standards.

5.2.3 Transcriptomic analysis of flowering *C. sativa* samples

5.2.3.1 RNA extraction, quantification, and retro transcription.

Total RNAs from Cannabis flowers was isolated using the SpectrumTM Plant Total RNA Kit (Sigma-Aldrich) following the manufacturer instructions. Briefly, for each sample, around 100 mg of frozen powdered flowering tissue was mixed with lysis solution supplemented with 2-mercaptoethanol, was filtered, and then bonded to an RNA column.

Protocol A was chosen for RNA binding as it was recommended for tissues with high water content including flower samples, after three column washes RNA was finally eluted in 50

µL of RNAse free H_2O water. RNA was then quantified using a NanoDrop 2000c UV–Vis Spectrophotometer (Thermo Scientific, Pittsburgh, PA, USA). For most samples very high yields of total RNA was achieved with concentrations around 450 ng/µL; only in late male flowers collected at least 24 days after flower initiation yields were significantly lowered, for these samples RNA concentrations were measured between 100 to 200 ng/µL.

Lower concentrations of RNA in aging male *C. sativa* flowers were expected because they develop faster and enter the senescence phase much earlier comparted to their female counterparts.

Following quantification, for each sample, 2 µg of RNA were treated with DNase with the TURBO™ DNase Kit (Invitrogen) following the manufacturer instructions. Afterwards total RNA was retro-transcripted into cDNA with SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific®) following the manufacturer recommendations.

These cDNA samples would later be used in multiple qRT-PCR experiments such as cannabinoids and ethylene pathways' gene expression studies.

5.2.3.2 Cannabinoid pathway gene expression studies

The cannabinoid pathway genes expression were examined by developing specific primers in silico, although PCR primers for all the involved genes had already been designed previously and are available in literature [174,175] new primers pairs have been designed in order to have them all with the same PCR characteristics like annealing temperature (set at 60 °C) and product sizes (between 100 to 200 bp). Information on all developed qRT-PCR primers is presented in: *Supplementary Table S4: Developed Primers for qRT-PCR*.

For the cannabinoid pathway transcriptome analysis, the following genes have been studied:

- Olivetolic Acid prenyltransferases = responsible for the synthesis of cannabigerolic acid (CBGA) two genes: prenyltransferase 1 (CsPT1) and prenyltransferase 4 (CsPT4) have been reportedly implicated in the CBGA biosynthesis [176].

- Tetrahydrocannabinolic acid synthase (THCAS) = responsible for catalyzing the formation of THCA from CBGA [17,177].
- Cannabidiolic-acid synthase (CBDAS) = responsible for the oxidative cyclization of CBGA into CBDA [17,89,178]
- Cannabichromenic acid synthase (CBCAS) = responsible for the cyclization of the monoterpene moiety in CBGA leading to the production of cannabichromenic acid (CBCA) [17,179].
- Additionally, Cannabis housekeeping genes were needed for the qRT-PCR analysis, four were selected for testing in our samples: Actin, EF2, F-BOX and Histone-3 [180,181].

In Fig 5.1 the cannabinoid pathway under study and the relevant investigated genes and their chemical products are presented.

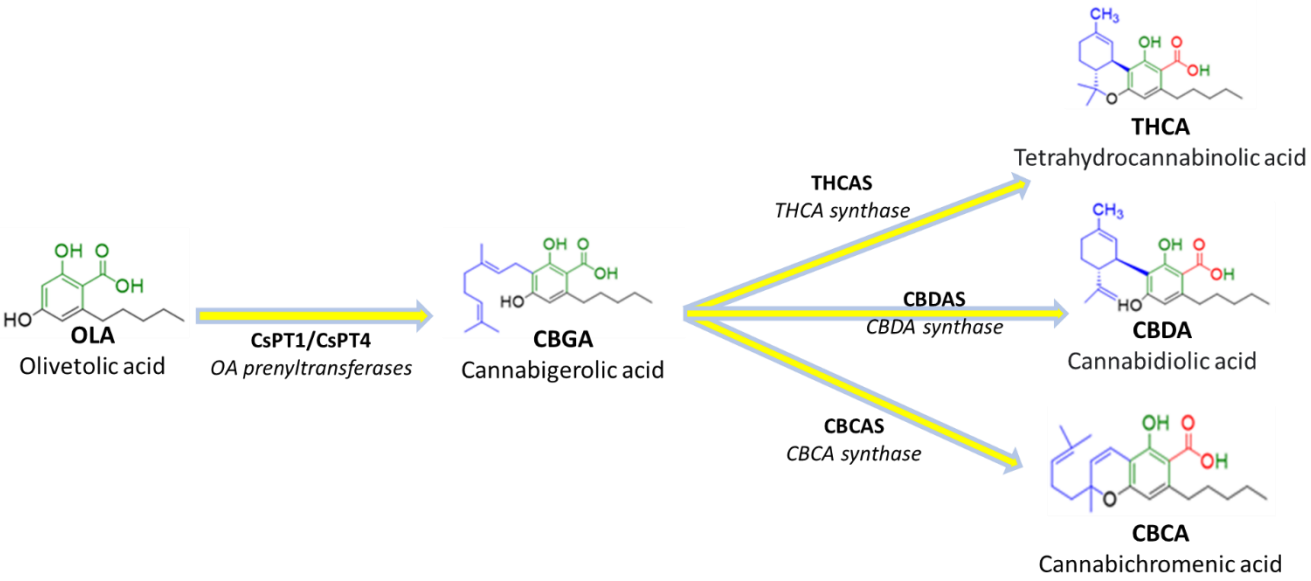

Fig 5.1: Cannabinoid pathway under study with our transcriptomic analysis. qRT-PCR primers have been developed for the relevant genes: CsPT1, CsPT4, THCA synthase, CBDA synthase and CBCA synthase.

Following in-silico design of our primer panel, all primer pairs were tested on cDNA obtained from Cannabis flower samples.

Firstly, PCR amplification was tested with gel electrophoresis to confirm the functionality of the primers and the correct amplification of the expected product sizes.

To do so, for each primer pair test, a PCR mix was prepared containing: 0.08 µL Taq Polymerases (Thermo Scientific™), 4 µL Buffer 5X, 0.8 µL primer mix 20 µM (forward + reverse), 0.4 µL dNTPs, 1 µL cDNA (500 ng), 13.72 µL H_2O up to a total volume of 20 µL. PCR reactions were performed in a Veriti® 96 Well Thermal Cycler (Applied Biosystems®) as follows: 94 °C for 5 min; 35 cycles at 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s; followed by a final extension of 7 min at 72 °C. The quality of the PCR products was then assessed by electrophoresis on a 2% (w/v) agarose gel stained with 1X SYBR® Safe™ DNA Gel Stain (Life Technologies) using Tris–acetate-EDTA (TAE) running buffer.

Following primer functionality testing quantitative RT-PCR was performed with a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) using GoTaq® qPCR Master Mix (Promega, Madison, Wisconsin, USA). Each reaction was carried out for 40 amplification cycles, in technical triplicates using the primers reported in *Supplementary Table S4: Developed Primers for qRT-PCR.*; data was normalized using the proper endogenous reference gene, which resulted to be F-BOX for this study as can be seen in *Supplementary Figure S5: amplification plot of endogenous reference genes during female flower development*. Finally, data analysis was performed using the $2^{-\Delta\Delta CT}$ method [182].

5.2.3.3 Sex reversion in *C. sativa*: ethylene pathway studies

Following the work currently ongoing by Adrian S. Monthony et al. [172] which found differential expression in the analysis of ethylene-related genes (ERGs) between RNA-seq reads experiments PRJNA669389 [170] and PRJNA549804 [171] qRT-PCR primers were designed to verify ERGs expression in Male Flowers (MF), Female Flowers (FF) and Induced Male Flowers (IMF). IMF were obtained with Silver ThioSulfate (STS), a potent inhibitor of ethylene signaling, as described in **Chapter 2.3.2.1:** *Sex-reversion experiments*.

Our goal was to confirm the hypothesis that ethylene absence during female flower development (caused by the STS treatment) is correlated to sex reversion in *Cannabis sativa* and results in the production of male flower in genetically female plants [41,172].

The analyzed samples were collected 14 days after flower initiation, this was done in order to have comparable results to the differential expression analysis performed by Monthony et al. [172] which was based on RNA-seq datasets from samples collected at the same time point [170,171].

Differential expression analysis by Monthony et al. [172] showed five cases of ERGs with floral-organ-concordant (FOC) expression, FOC refer to genes whose expression pattern is shared between organisms with the same floral organ, but not necessarily the same sex-chromosome karyotype; for example, transcription levels for male flowers on XY plants and induced male flowers on XX plants were found to be statistically identical. Additionally, two case of genes where expression in the sex-changed plant does not match either that of the plant with which is shares the same sex chromosome karyotype or the plant with which is shares the same floral organs were found, these were classified as unique ethylene-related genes (uERG).

In order to find general expression patterns for the *Cannabis sativa* specie, in contrast to variety-specific ones, three different varieties have been analyzed in the qRT-PCR experiment. For each variety sampling was performed from three different plants for each flower type (FF, MF and IMF).

qRT-PCR primers compatible with the previous cannabinoid pathway gene expression study for the five floral-organ-concordant (FOC) and two unique ethylene-related genes (uERG) were then developed and tested.

The following floral-organ-concordant (FOC) genes were investigated:

- ACS2 (LOC115707397): involved in the ethylene biosynthesis pathway, found to be implicated in ethylene production during immature fruit development [183];
- ETP2 (LOC115720323): involved as a suppressor in the ethylene signaling pathway [184];

- FYF2 (LOC115706939) and FYF4 (LOC115717309): part of the FOREVER YOUNG FLOWER family, involved in an ethylene adjacent pathway, acts as a suppressor of floral abscission [185];
- MTN (LOC115704327): involved in the Yang Cycle, *CsMTN* reportedly associated with masculinity and sexual plasticity in cucumber, melon [186] and Cannabis [172].

The following unique ethylene-related genes (uERG) were investigated:

- ACO5 (LOC115699408): involved in the ethylene biosynthesis pathway, responsible for ethylene production during immature fruit development in ripening tomatoes [183] as well as for sex determination in cucumber and melon [186];
- FYF1 (LOC115700576): part of the FOREVER YOUNG FLOWER family, involved in an ethylene adjacent pathway, acts as a suppressor of floral abscission [185].

All developed primers are listed in *Supplementary Table S4: Developed Primers for qRT-PCR* in the Ethylene Related Genes (ERGs) section, a test to confirm the functionality of the primers and the subsequent qRT-PCR analysis was performed as described in *5.2.3.2 Cannabinoid pathway gene expression studies*, the only difference was that the best endogenous reference gene for qRT-PCR normalization was found to be Elongation Factor 2 (EF2) for this analysis.

Results on the expression levels of the target genes was confronted with the ones obtained previously with differential expression analysis [172] and their biological significance was evaluated by considering their function the ethylene synthesis pathway [187,188].

5.3 Results and Discussion

5.3.1 Metabolomic analyses results

A protocol for the separation and proofing of phytocannabinoids with a UPLC-HRMS untargeted metabolomics analysis was tested on hemp flowering samples.

Numerous different cannabinoids and other metabolites (such as polyphenols, procyanidins and other polar compounds) were identified with a combination of pure standards and fragmentation patterns.

Cannabinoid identification was obtained with the gradient previously proposed by Aizpurua-Olaizola *et al.* (2014) [173] and the comparison with pure standards and is visualized for the three tested hemp varieties (HUN-1, HUN-2 and ITA-3) in Fig 5.2.

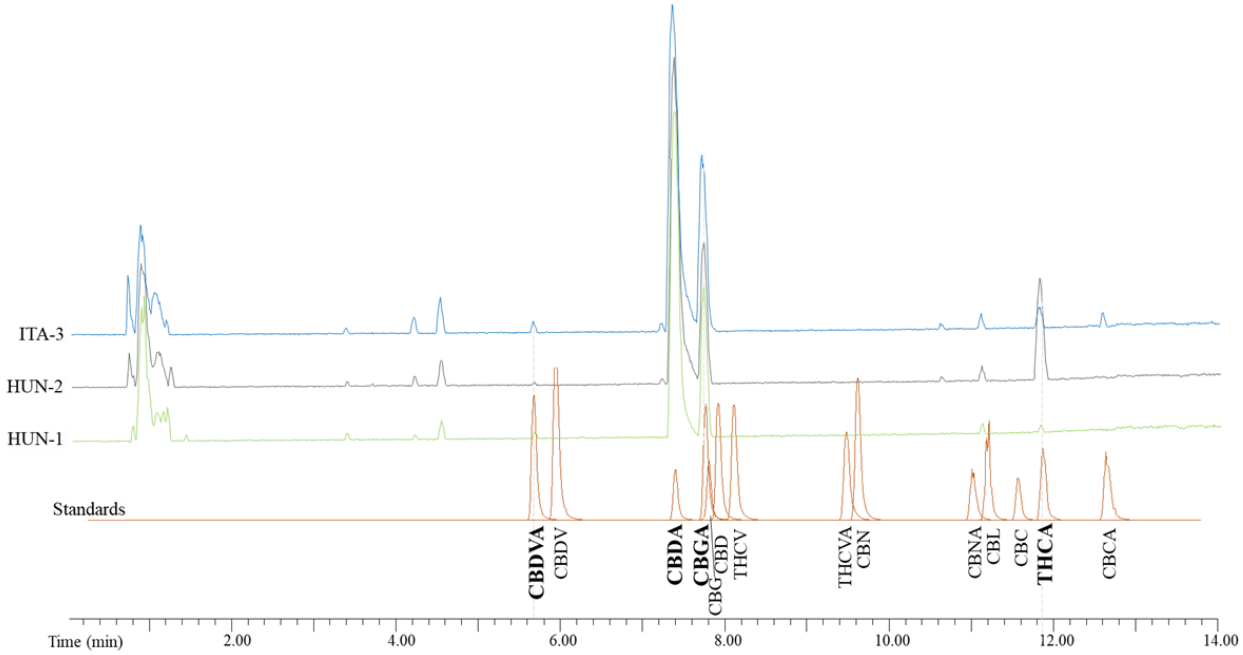

Fig 5.2: Scan chromatogram for the HUN-1, HUN-2 and ITA-3 samples collected at 14 days after flower initiation compared to the following cannabinoid standards: CBDVA, CBDV, CBDA, CBGA, CBG, CBD, THC, THCVA, CBN, CBNA, CBL, CBC, THCA, CBCA.

The following cannabinoids were identified with peak comparison with our standards: CBDVA, CBDA, CBGA, CBG, CBD, and THCA. Additionally, CBG and CBD, whose peak was covered by the much more intense CBGA's one, were detected by their fragmentation patterns.

From this identification analysis it was also observed that noticeable differences in cannabinoids concentrations were present between our tested varieties, as peak intensity is directly correlated to concentration. For example, CBDVA, CBDA and CBGA showed

nearly double the intensity in the ITA-3 samples compared to the other two; at the same time THCA was found to be up to five times higher in HUN-2 contrasted to the others, notably, THCA was still present in trace amounts considering that the samples were hemp-type.

On the same samples, aside from cannabinoids, other metabolites were detected in our UPLC-HRMS untargeted metabolomics analysis, a relevant group of polyphenols and other medium polar metabolites were detected in our *C. sativa* samples.

Glycosylated flavones and flavonoids as well as procyanidins and flavan-3-ols have been identified in our samples, these molecules are notable as they are reportedly correlated with various several health beneficial effects by acting as antioxidant, anticarcinogen, cardio-preventive, antimicrobial, anti-viral, and neuro-protective agents [189–192].

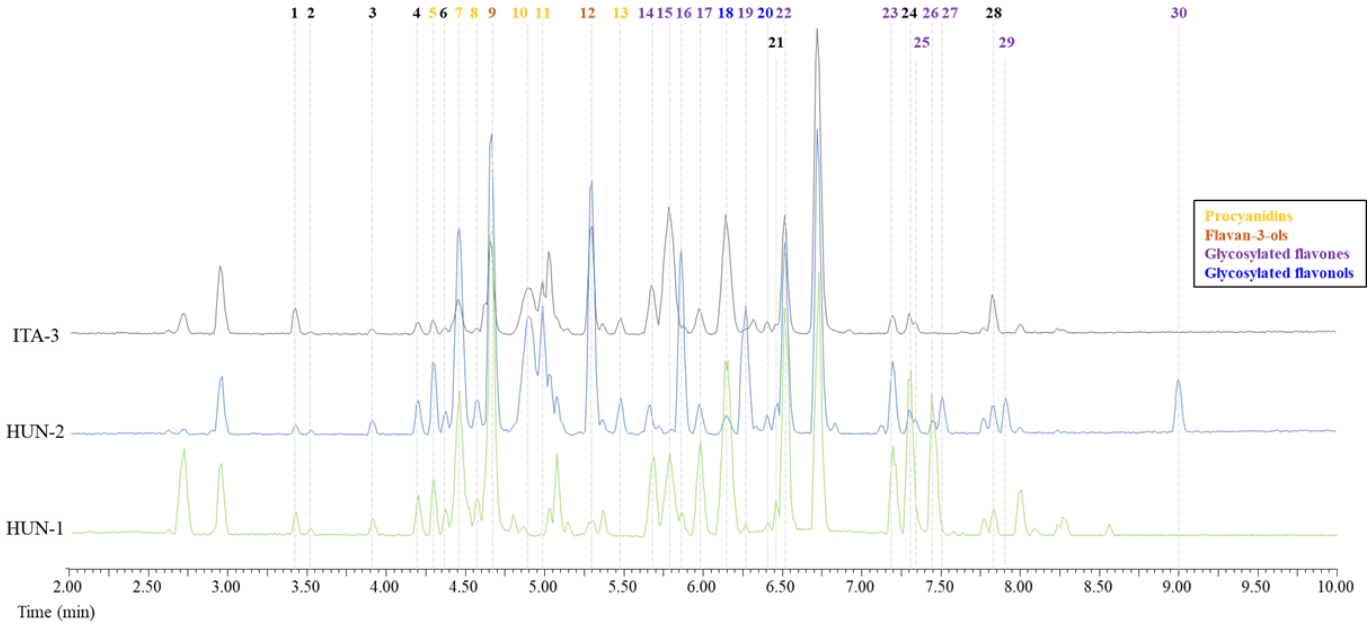

Fig 5.3: Scan chromatogram for the HUN-1, HUN-2 and ITA-3 samples collected at 14 days after flower initiation, the following glycosylated flavones and flavonoids, procyanidins, flavan-3-ols and other polar molecules have been identified: (1) Hydroxybenzoic acid-O-hexoside, (2) Pantothenic acid, (3) Galactaric/Glucaric acid-O-deoxyhexoside, (4) Galacturonic/Gluconic acid-O-deoxyhexoside, (5) B-type procyanidin dimer, (6) Galacturonic/Gluconic acid-O-deoxyhexoside, (7) B-type procyanidin dimer, (8) B-type procyanidin trimer, (9) Catechin, (10+11) B-type procyanidin dimer, (12) Epicatechin, (13) B-type procyanidin trimer, (14)

(iso)orientin derivative, (15) (iso)orientin derivative, (16) (iso)orientin, (17) (iso)orientin derivative, (18) Kaempferol-di-O-deoxyhexoside, (19) Vitexin, (20) Quercetin-3-O-glucoside, (21) Phloretin-di-C-glucoside, (22) Luteolin-O-glucuronide, (23) Apigenin-O-glucuronide, (24) Oligosaccharide derivative, (25) O-methyl luteolin glucuronide isomer, (26) O-methyl luteolin glucuronide isomer, (27) Acacetin-di-C-hexoside, (28) Polyketide, (29) Acacetin-O-hexoside, (30) Wogonoside.

As observed with the identified cannabinoids the polar molecules presence was also significantly highly variable between varieties (as shown in Fig 5.3) and collection timing (data not shown), for example the flavonoid Wogonoside (30) was detected only on the HUN-2 variety.

Afterwards, a study on cannabinoid accumulation during flower development was conducted with all collected female flowers samples from the HUN-1, HUN-2, and ITA-3 hemp-type Cannabis varieties, the experiment was conducted on flowering samples collected from 14 to 28 days after flower initiation.

Results from this quantification experiment showed a significant reduction in the main cannabinoids concentrations (CBDA and CBGA) during flower development in our hemp samples as shown in Fig 5.4.

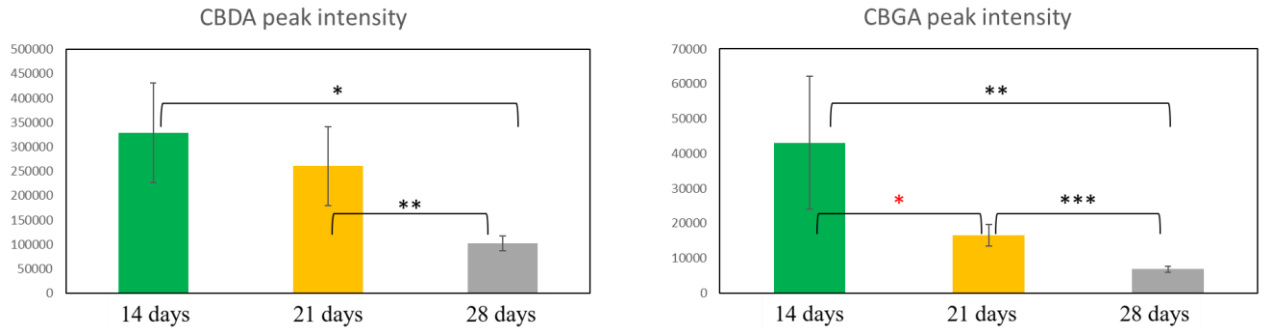

Fig 5.4: Variations of the major metabolites observed with our analysis between time points, samples collected at: 14 days, 21 days, and 28 days. At each time point data is an average between the three sampled varieties (HUN-1, HUN-2 and ITA-3). Level of significance (P value from T test: * > 95% * > 90%).

The metabolomic analysis clearly showed a downward trend with the flowering progression (between 14 to 28 days after flower initiation) moving towards seed formation.
Reduction of CBGA content during flower development was also reported in literature [193,194] and was expected as it is expended for the production of the downstream cannabinoids in the pathway. CBDA concentrations instead are usually reported to increase with flower maturation [193,195,196], the decrease in our samples could be explained by two hypotheses:

- Our samples were not dried before extraction as they also needed to be used for transcriptomic analysis, in contrast those in the literature were always dried beforehand [193–196]. Higher water content in the later time points may have artificially reduced the metabolites content.
- Cannabinoid biosynthesis could not cope with the high mass growth of the flowering samples observed in the third week of development: primordial apical flower buds collected 14 days after initiation weighted on average under 5 mg, in contrast, developing flower buds collected 28 days after initiation measured on average over 20 mg per sample.

Commonly, cannabinoid quantification is performed at harvest time which, depending on the variety, generally occurs from 7 to 10 weeks after flower initiation, cannabinoids analysis in early flower development is mostly an under-studied process.
Metabolomic analyses in later stages of flower development were then performed with our HUN-1 hemp variety, CBDA was quantified at 35, 49 and 63 days after flower initiation (Fig 5.5) and bioaccumulation was observed as expected for this stages [193–196].

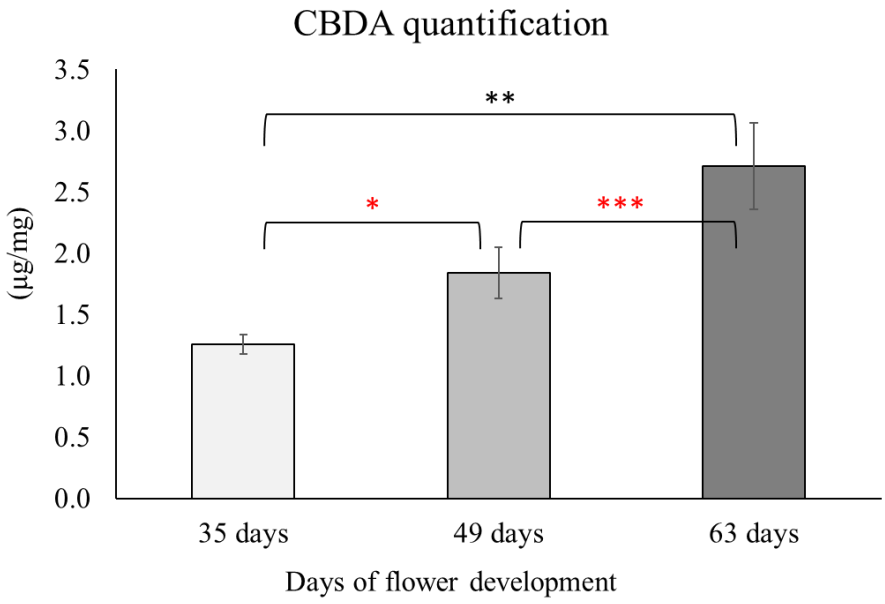

Fig 5.5: Variations of CBDA content between time points, samples of HUN-1 collected at: 35 days, 49 days, and 63 days. At each time point data has been averaged between three biological replicates of our HUN-1 variety. Level of significance (P value from T test: * > 99% * > 95%).

CBDA content was shown to be around 0.3% (0.271 ± 0.035) at flower maturation which occurred at 7 to 8 weeks after flower initiation. Considering that flowers were not exsiccated before being weighed, results were lower but comparable to those of similar fiber-type varieties [197]; in the reported study by Pavlovic et al. CBDA content (in dry weight) for two hemp varieties, Finola and Futura, was at 2.614 ± 0.580 % and 0.56 ± 0.049 % respectively.

5.3.2 Cannabinoid gene expression analysis

Regarding the cannabinoid genes expression analysis, all primers were verified for the correct amplification of cDNA samples, as expected in our hemp-type flowering samples no expression was detected for the THCAS and CBCAS genes;

Additional testing conducted during my stay at the University of Saskatchewan, where drug-type Cannabis samples were available, showed that the THCAS and CBCAS primers were

fully functional on those varieties; gel electrophoresis analysis following PCR amplification with the developed primers showed the expected results for the TCH and CBC genes in these drug-types varieties.

Regarding the endogenous reference genes testing F-BOX, a gene involved in many crucial processes such as embryogenesis, hormonal responses, seedling development, floral organogenesis and senescence, was found to have the most consistent transcriptional levels during flower development along with high PCR efficacy when compared with the other tested housekeeping genes: EF-2, HIS-3 and Actin (*Supplementary Figure S5*).

From our quantitative real-time PCR (qRT-PCR) results on female flowering samples gene expression for the cannabinoid pathway showed a clear and statistically significant downward trend for all involved genes correlated with the advancing of flower development as shown in Figure 5.6.

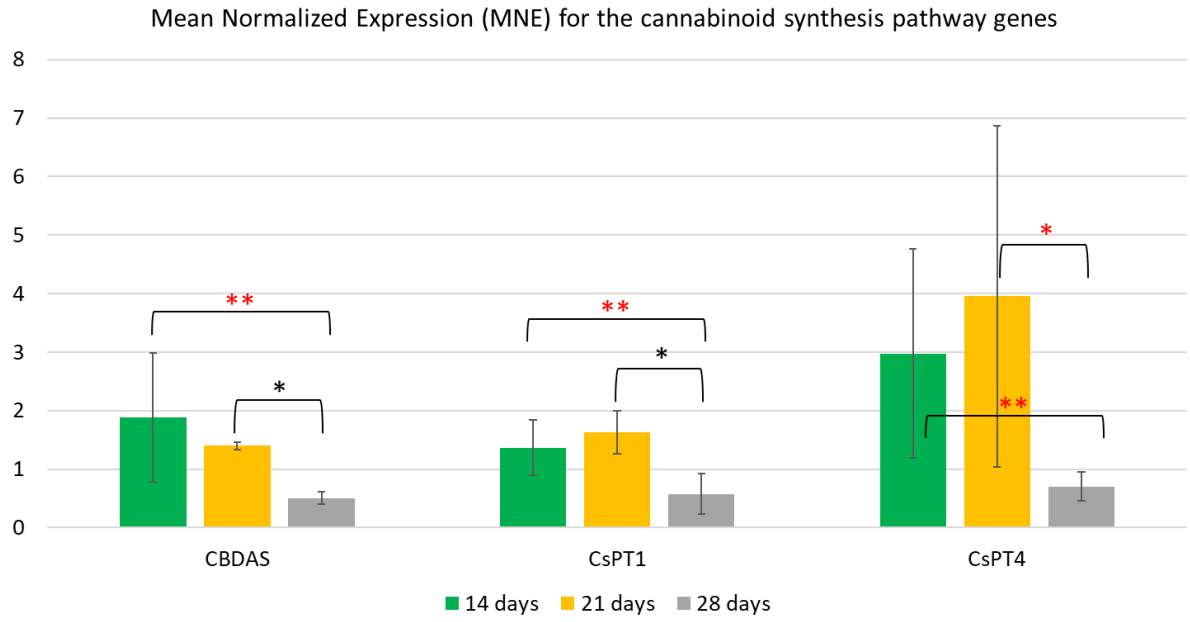

Fig 5.6: Mean Normalized Expression (MNE) of CBDAS, CsPT1 and CsPT4 and its observed variation between time points 1 (14 days), 2 (21 days) and 3 (28 days). Level of significance (P value from T test: * > 95% * > 90%).

High error values were observed in this study, these may have been due to multiple factors such as: each data value is an average of samples from three different hemp varieties, the selected varieties were found to have slightly different flower development speeds, cannabinoid synthesis is known to be inherently variable even inside an individual *C. sativa* plant depending on position external factors [109–111].

Importantly, even when expression values differed between samples the same transcription patterns were observed in all three tested varieties.

As we were working with fiber-type Cannabis varieties known for low cannabinoid synthesis and accumulation in flowers, which was observed during our metabolomic analysis, a reduction in expression levels for both genes responsible for the synthesis of CBG (CsPT1 and CsPT4) as well as for the CBDAS were expected.

This results are in line with others in the literature that shows significant reductions in expression levels for CBDAS, CsPT1 and CsPT4 after 4 to 5 weeks following flower initiation in low-CBD content varieties [198].

It must be noted that, contrary to what we have seen in our fiber-type Cannabis samples, CBDAS and THCAS genes in drug-type Cannabis strains have usually relatively higher expression in the latter stages of female lower development (mature floral buds) compared to immature floral buds as cannabinoids are known to accumulate during development in these varieties [17,179,199].

5.3.3 Ethylene biosynthesis genes' expression analysis results

During our experiment on ethylene biosynthesis genes, the correct amplification of all developed primers was verified on cDNA samples collected from Female Flowers (FF), Male Flowers (MF) and Induced Male Flowers (IMF).

During this qRT-PCR analysis the endogenous reference gene was EF-2, responsible for coding an essential protein catalyzing ribosomal translocation during protein synthesis. The house-keeping gene previously utilized for the cannabinoid pathway study (F-BOX) as well

as all other tested endogenous reference genes (HIS-3 and Actin) were found to have highly variable transcription levels between different flowering sample types.

Results of the qRT-PCR experiment are presented in Figure 5.7 where the Mean Normalized Expression (MNE) for each gene has been calculated confronting different flower samples (FF, MF and IMF). The first five presented genes (ACS2, ETP2, FYF2, FYF4, and MTN) were hypothesized to have floral-organ-concordant (FOC) expression, the last two (ACO5 and FYF1) were instead classified as unique ethylene-related genes (uERG).

Each MNE value in Fig 5.7 is an average of the expression levels of the samples from the three analyzed varieties. It is significant that the same expression patterns were still observed even if each sample was considered singularly.

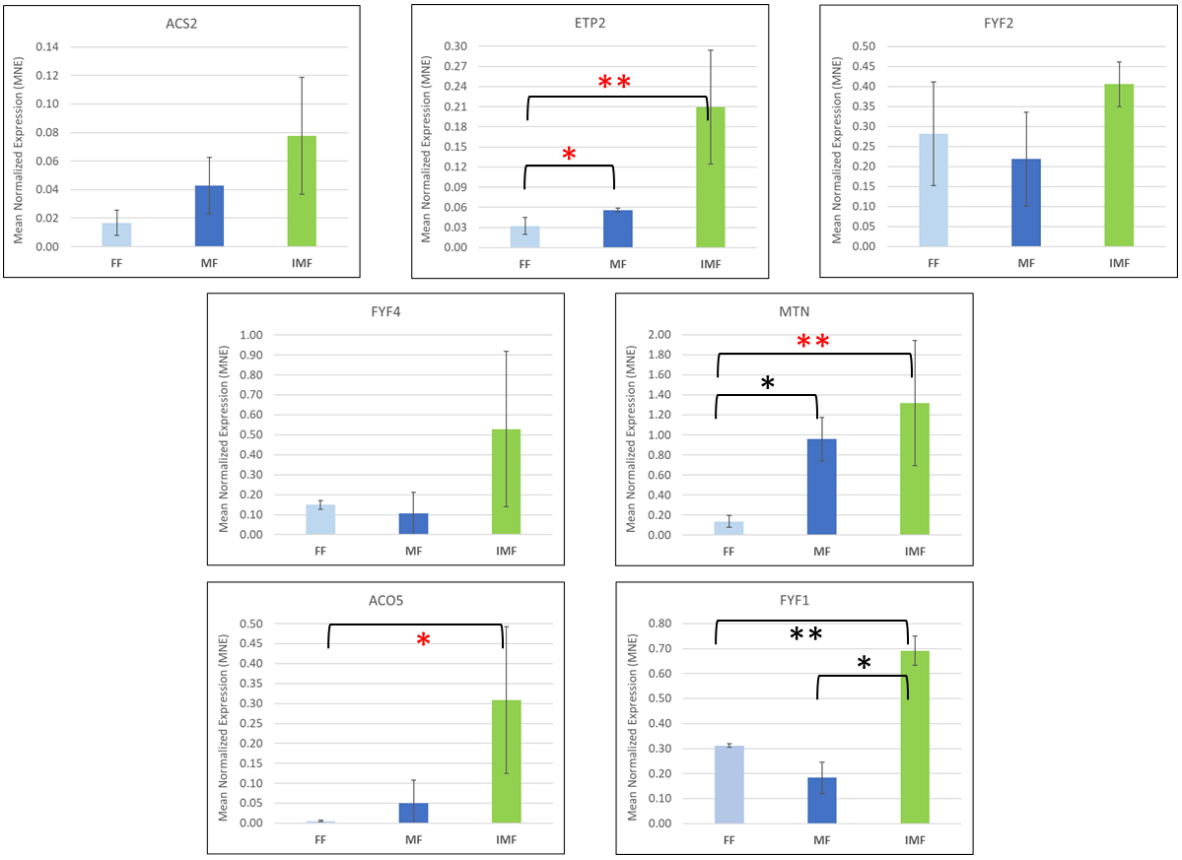

Fig 5.7: Mean Normalized Expression (MNE) of ethylene related genes (ERG) in Female Flowers (FF), Male Flowers (MF) and Induced Male Flowers (IMF). Level of significance (P value from T test: * > 95% * > 90%).

It must be noted that, due to very high variation in the gene expression between sampled varieties (ITA-3, HUN-1 and HUN-2) and the low number of biological samples available for this analysis, for three of examined genes: ACS2, FYF2 and FYF4, no statistically significant variations in expression levels could be observed.
For the rest of the genes' expression levels statistically relevant differences were found with at least a level of significance higher than 90% (*) or 95% (*) (P value from T test).
Our objective was to verity the effect of STS (a known potent inhibitor of ethylene signaling) on genetically female plants, the treatment is known to result in sex reversion with the production of male flowers (IMF) [55]. To understand its molecular effect the expression of the following genes (those with statistical significance in Fig 5,7) was compared between flowering samples (FF, MF, and IMF):

- For the ETP2 gene significantly higher expression levels were found in all male flowers (MF and IMF) following the same floral-organ-concordant (FOC) expression found by the ongoing research by Monthony et al.[172], these results also confirms that in the absence of ethylene (in MF and IMF samples) ETP2 is expressed and can act as a suppressor to the ethylene signaling pathway [184].

- MTN transcripts were found to be significantly more expressed in MF and IMF than in FF, this suggests that MTN is associated with masculinity and sexual plasticity in *Cannabis sativa,* the same expression pattern for the gene was found by Monthony et al.[172].

- In the ACS5 gene a significantly higher expression was found on the IMF compared to the FF but, unfortunately, due to high variation between MF samples no statistically relevant difference could be found between it and the others; in the analysis by Monthony et al.[172] the same pattern (much higher expression in IMF compared to

FF) was assessed for this gene. Disruptions of ACO genes were found to results in decreased ethylene production and sex reversion in cucumber and melon [186].

- Finally, in the FYF1 gene expression levels were statistically different between both FF vs IMF and MF vs IMF, as found by Monthony et al. [172] IMF expression was higher compared to MF but, opposite to their results, FF expression level of FYF1 resulted lower than in IMF. FYF genes are reportedly expressed in young flowers prior to pollination and acts as a suppressor of floral abscission [185], lower expression levels on our female flower samples (compared to by Monthony et al. [172]) may be the result of an early pollination event that ensued in our female plants.

Future experiments should include a higher sample size as in this analysis, conducted only with flowering samples from three varieties, differences in transcription levels between strains were high and led to low levels of significance for many expression levels comparisons between different flower types (FF, MF and IMF).

5.4 Conclusions

Experiments conducted on flowering Cannabis samples gave us valuable results on metabolites identification, cannabinoid quantification, cannabinoid pathway expression levels as well as insights on the correlation between the ethylene pathway and sex reversion. Metabolomic and transcriptomic analysis have proven to be an essential component for the study of this increasingly relevant plant.

During our whole metabolomic analysis various detected molecules of interest such as flavones, flavonoids, procyanidins and flavan-3-ols have been identified in our samples, their presence was also notably shifting between varieties and flower collection timing.

For legal reasons, only hemp samples could be used for cannabinoids identification and quantification, as expected the mayor detected cannabinoids were CBGA and CBDA, small

or trace amounts of CBDVA, CBG, CBD, and THCA were found as expected for this hemp-Cannabis varieties.

Predictably, at flower harvest time CBDA concentration was calculated only at 0.3% (w/w) as opposed to the ~20% that can be surpassed by drug-type Cannabis (CBDA or THCA, depending on which drug-type).

During flower development CBDA concentration showed a reduction from 14 to 28 days after flower initiation, but also an increase from 35 to 63 days reaching flower maturation in a different experiment; it was hypothesized that the decrease at 28 days may have been due to high vegetative growth at that stage paired with low CBDAS expression levels.

In fact, during our transcriptomic analysis, expression levels for genes of the cannabinoid pathway (CBDAS, CsPT1 and CsPT4) showed a downward trend from 14 to 28 days of flower development.

It must be noted that if the experiment were to be conducted on drug-type Cannabis expression levels should have increased going towards flower maturation.

Finally, a transcriptomic study aimed at better understanding sex determination was conducted, we focused our attention on validating speculated differential gene expressions linked to the ethylene pathway between FF, MF, and IMF samples. The observed gene expression changes supported the hypothesis that the STS treatment leads to variations in the ERGs expressions with the biological effect of inhibiting the ethylene biosynthesis; and this results in the production of male flowers in genetically female plants.

Understanding sex determination is extremely important as a scientific goal, for breeding purposes and for cannabinoid production in industrial scale. As only unfertilized female flowers are harvested for cannabinoid extraction.

CHAPTER 6: Concluding Remarks

Cannabis sativa is a high-demand crop with a long human history of medicinal, agricultural, industrial, and recreational uses. However, following a long history of criminalization and prohibitions, there is a lack of knowledge of this valuable plant in many fields of science, including the use of genomics for breeding new varieties and in the tracing of their commercial derivatives.

Our research focused on the implementation and validation of two informative, reliable, and reproducible genotyping protocols in *C. sativa* that may be useful both for marker-assisted breeding of new varieties and for unequivocally identifying specific varieties already on the market or assessing their genetic identity.

First, we demonstrated the effectiveness of our novel set of SSR marker loci for assessing the genetic distinctiveness, uniformity, and stability of individual varieties as well as for estimating heterozygosity/homozygosity statistics of single plants or populations and the extent of genetic variation within and genetic relationship among different *C. sativa* varieties. This information could be exploited for planning crosses and predicting heterosis as an expression of hybridity in experimental F1 populations based on the allelic divergence and genetic distance of the selected parental lines.

Afterwards we developed and tested a chloroplast sequencing protocol which was able to analyze samples from our own cultivated hemp and drug-type varieties as well as from sequences of varieties form the literature and assign them to common maternal ancestor clusters.

Comparing the two methods we determined that the SSR marker's one was more effective in calculating genetic distances between samples, especially for closely related ones; the chloroplast sequencing method was instead essential for the comparison of our samples with more in the literature.

The conducted SSR analysis results showed the mutual genetic distance between our samples (**Chapter 3.4:** *Conclusions*, Figure 3.5) with a Principal component analysis (PCoA) and the

chloroplast sequencing analysis showed it thought a phylogenetic tree that also included additional samples from the literature (**Chapter 4.3:** *Results and discussion*, Figure 4.3). With the comparison of the results (as shown in Fig 6.1) we observed that our cultivated varieties analyzed with the SSR marker method (top figure) were assigned to two different clusters in the chloroplast sequencing phylogenetic tree (bottom figure).

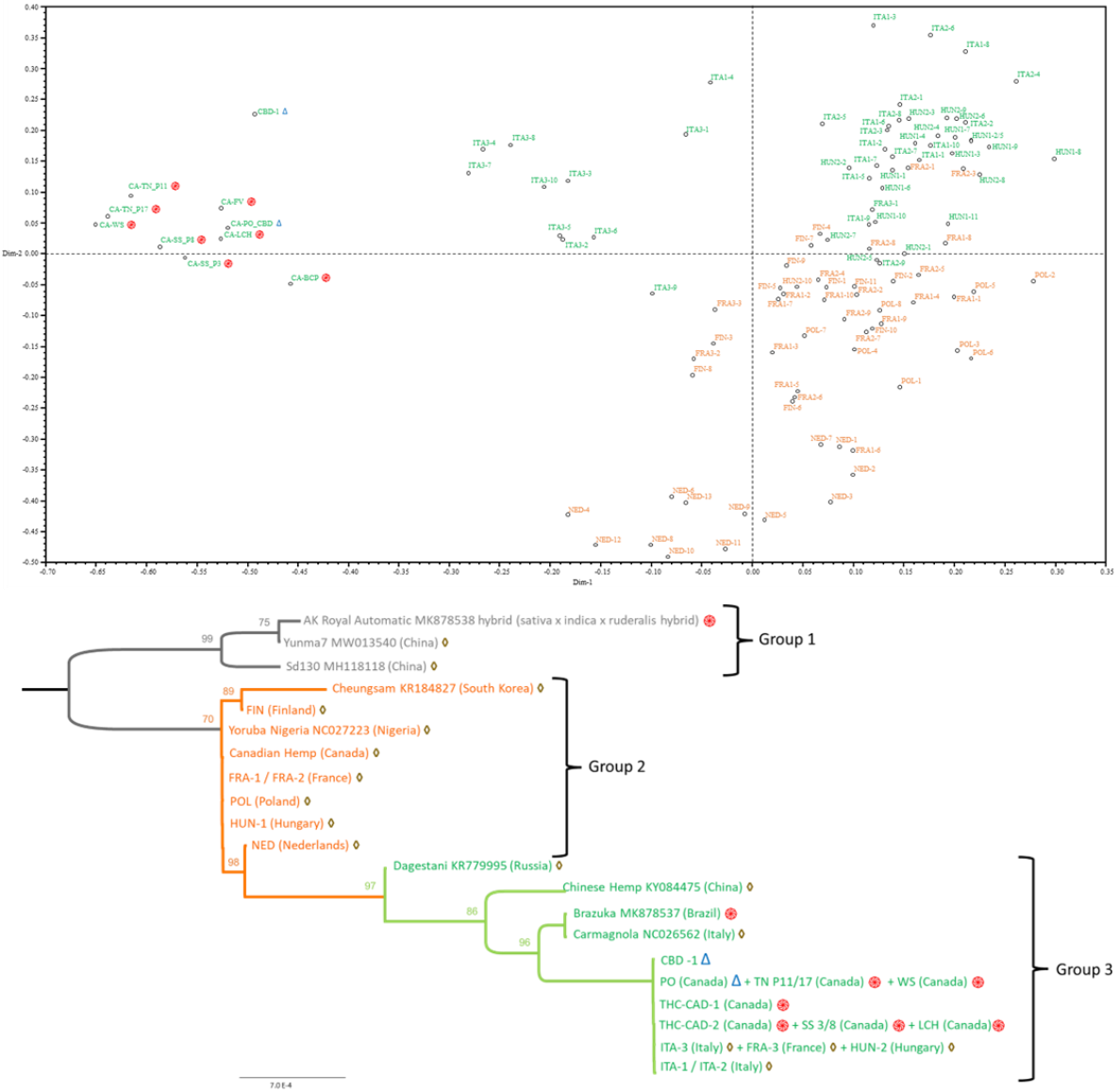

Fig 6.1: Comparison between the results from the PCoA from SSR genotyping analysis (top) and the phylogenetic tree from the chloroplast sequencing analysis (bottom). Samples from groups 2 and 3 are all color-coded in orange and green respectively.

No samples from group 1 (grey in the chloroplast sequencing analysis) are present in the SSR genotyping analysis (top) as they were all collected from chloroplast sequencing databases in the literature. Samples are labeled as THC-Drug Type (), CBD-Drug Type (Δ) or fiber-type (◊ or no labeling).

Our samples from the bottom-right section in the PCoA plot were all assigned to the Group 2 cluster in the phylogenetic tree and all samples from the top-right and top-left section in the PCoA plot were assigned to the Group 3 cluster in the phylogenetic tree. These results showed mutual corroboration between our genotyping protocols. The SSR genotyping analysis was also clearly able to differentiate between samples of the same variety; they were also capable of visibly distinguishing between THC/CBD-Drug Type samples and fiber type samples among the group 3 varieties with the drug-types located on the far-left side of the PCoA and the fiber types on the right side.

The chloroplast sequencing analysis showed that a common maternal ancestor was shared between all samples on the Group 3 cluster which was interestingly comprised of both drug and fiber type varieties.

We are confident that the DNA-based labeling of varieties covered by patent protection represents the only solution in the Cannabis market for managing intellectual property rights in the Cannabis seed industry and protecting professional producers of Cannabis.

Numerous flowering experiments were conducted to model flowering behaviors and sex reversion protocols; several hemp- and drug-type varieties were tested, and, during our flowering development experiments, the evolution of major cannabinoids concentrations and expression of target genes of interest were studied.

A metabolomic protocol was tested and found able to detect numerous metabolites extracted from *C. sativa* flowers, in particular numerous cannabinoids, and other molecules of interest (such as flavones, flavonoids, procyanidins and flavan-3-ols) were identified.

In our hemp-type Cannabis samples the mayor identified cannabinoid was CBDA as expected, its concentration was quantified during flower development.

Results from our cannabinoid pathway gene expression transcriptomic experiment conformed with the ones from the metabolic quantification showing a consistent decrease in the CBDAS gene expression alongside CBDA concentrations for all our flowering hemp samples in the 14 to 28 days of flower development period. On the contrary CBDA accumulation was observed from 35 to 63 days samples as they were reaching flower maturation.

Consecutively, our transcriptomic experiment investigating hypothesized relations between the poorly known sex reversion phenomenon and the ethylene pathway in *C. sativa* was successful in validating relevant differential gene expressions between FF, MF, and IMF samples. The observed gene expression changes supported the hypothesis that the STS treatment leads to variations in the ERGs expressions with the biological effect of inhibiting the ethylene biosynthesis; and this results in the production of male flowers in genetically female plants.

Finally, our experiments on flowering samples, even if they have been conducted mostly on hemp-type Cannabis, have given us fundamental insights on how to properly analyze *C. sativa* flowers with metabolomic and transcriptomic analyses.

This knowledge will be essential in the future for the choice of the correct individual plants in breeding programs, it is in fact essential to be able to select plants in a crossing not only basing it on genetic distances results but also considering the analysis of their chemical composition and transcriptional capabilities.

References

1. Kovalchuk, I.; Pellino, M.; Rigault, P.; Van Velzen, R.; Ebersbach, J.; Ashnest, J.R.; Mau, M.; Schranz, M.E.; Alcorn, J.; Laprairie, R.B.; et al. The Genomics of Cannabis and Its Close Relatives. *Annu. Rev. Plant Biol.* **2020**, *71*, 713–739, doi:10.1146/annurev-arplant-081519-040203.
2. Schultes, R.E. *Random thoughts and queries on the botany of cannabis*; C. R. B. Joyce and S. H. Curry, Ed.; London, UK: J.A.Churchill, 1970;
3. Small, E.; Cronquist, A. *A Practical and Natural Taxonomy for Cannabis*; International Association for Plant Taxonomy (IAPT), Ed.; Taxon 25: Hoboken, New Jersey, 1976; Vol. 25;.
4. Schultes, R. E., and Hofmann, A. *The Botany and Chemistry of Hallucinogens.*; 2nd ed.; Springfield, Illinois, USA: Charles C Thomas Pub Ltd, 1980;
5. Gerarde, J. The Herball or General Historie of Plants. *John Norton, London* **1597**.
6. Clarke, C. *Marijuana Botany. An Advanced Study: The Propagation and Breeding of Distinctive Cannabis*; Ronin Publishing: Berkeley, California, 1981; Vol. 13; ISBN 9780914171782.
7. Ming R., Bendahmane A., & R.S.S. Sex chromosomes in land plants. *Ann. Rev. Plant Biol.* **2011**, *62*, 485–514.
8. Lamarck, J.B.P.A. *Encyclopédie méthodique. Botanique*; Paris, France: Panckoucke, 1785;
9. Hazekamp, A., Tejkalová, K., & Papadimitriou, S. Cannabis: from cultivar to chemovar II - a metabolomics approach to Cannabis classification. *Cannabis Cannabinoid Res.* **2016**, *1*, 202–215, doi:10.1089/can.2016.0017.
10. Piomelli, D., & Russo, E.B. The Cannabis sativa Versus Cannabis indica Debate: An Interview with Ethan Russo. *MD. Cannabis Cannabinoid Res* **2016**, *1*, 4–46, doi:10.1089/can.2015.29003.ebr.
11. Charlesworth, D.; Charlesworth, B.; Marais, G. Steps in the evolution of

heteromorphic sex chromosomes. *Heredity (Edinb).* **2005**, *95*, 118–128, doi:10.1038/sj.hdy.6800697.

12. Renner, S.S. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* **2014**, *101*, 1588–1596, doi:10.3732/ajb.1400196.
13. Sakamoto, K.; Akiyama, Y.; Fukui, K.; Kamada, H.; Satoh, S. Characterization; Genome sizes and morphology of sex chromosomes in hemp (Cannabis sativa L.). *Cytologia (Tokyo).* **1998**, *63*, 459–464, doi:10.1508/cytologia.63.459.
14. Faux, A.M.; Draye, X.; Flamand, M.C.; Occre, A.; Bertin, P. Identification of QTLs for sex expression in dioecious and monoecious hemp (Cannabis sativa L.). *Euphytica* **2016**, *209*, 357–376, doi:10.1007/s10681-016-1641-2.
15. Lynch, R. C., Vergara, D., Tittes, S., White, K., Schwartz, C., Gibbs, M. J., Ruthenburg, T. C., Decesare, K., Land, D. P., & Kane, N.C. Genomic and chemical diversity in cannabis. *Crit. Rev. Plant Sci.* **2016**, *35*, 349–363.
16. Peil, A.; Flachowsky, H.; Schumann, E.; Weber, W.E. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (Cannabis sativa L.). *Theor. Appl. Genet.* **2003**, *107*, 102–109, doi:10.1007/s00122-003-1212-5.
17. van Bakel, H.; Stout, J.M.; Cote, A.G.; Tallon, C.M.; Sharpe, A.G.; Hughes, T.R.; Page, J.E. The draft genome and transcriptome of Cannabis sativa. *Genome Biol.* **2011**, *12*, doi:10.1186/gb-2011-12-10-r102.
18. Divashuk, M.G.; Alexandrov, O.S.; Razumova, O. V.; Kirov, I. V.; Karlov, G.I. Molecular cytogenetic characterization of the dioecious Cannabis sativa with an XY chromosome sex determination system. *PLoS One* **2014**, *9*, 1–7, doi:10.1371/journal.pone.0085118.
19. Hennink, S. Optimisation of Breeding for Agronomic Traits in Fibre Hemp (Cannabis Sativa L.). *Euphytica* **1994**, *78*, 69–76, doi:10.1017/ CBO9781107415324.004.
20. Klaue, K. Drugs, addiction, deviance and disease as social constructs. *Bull. Narc.* **1999**, *51*, 45–55.
21. Kendell, R. Cannabis condemned: the proscription of Indian hemp. *Addiction* **2003**,

98, 143–151, doi:10.1046/j.1360-0443.2003.00273.x.

22. Bewley-Taylor, D.; Blickman, T.; Jelsma, M. The Rise and Decline of Cannabis Prohibition: the history of cannabis in the UN drug control system and options for reform. **2014**, 88.
23. Seddon, T.; Floodgate, W. *Regulating cannabis: A global review and future directions*; 2020; ISBN 9783030529277.
24. Decorte, T.; Pardal, M.; Queirolo, R.; Boidi, M.F.; Sánchez Avilés, C.; Parés Franquero, Ò. Regulating Cannabis Social Clubs: A comparative analysis of legal and self-regulatory practices in Spain, Belgium and Uruguay. *Int. J. Drug Policy* **2017**, *43*, 44–56, doi:10.1016/j.drugpo.2016.12.020.
25. Schlag, A.K. An Evaluation of Regulatory Regimes of Medical Cannabis: What Lessons Can Be Learned for the UK? *Med. cannabis cannabinoids* **2020**, *3*, 76–83, doi:10.1159/000505028.
26. Ferfuia, C.; Zuliani, F.; Danuso, F.; Piani, B.; Cattivello, C.; Dorigo, G.; Baldini, M. Performance and stability of different monoecious hemp cultivars in a multi-environments trial in north-eastern Italy. *Agronomy* **2021**, *11*, 1–15, doi:10.3390/agronomy11071424.
27. Hammond, C.T.; Mahlberg, P.G. Morphogenesis of Capitate Glandular Hairs of Cannabis Sativa (Cannabaceae). *Am. J. Bot.* **1977**, *64*, 1023–1031, doi:10.1002/j.1537-2197.1977.tb11948.x.
28. Andre, C.M.; Hausman, J.F.; Guerriero, G. Cannabis sativa: The plant of the thousand and one molecules. *Front. Plant Sci.* **2016**, *7*, 1–17, doi:10.3389/fpls.2016.00019.
29. Raman, V.; Lata, H.; Chandra, S.; Khan, I.A.; ElSohly, M.A. Morpho-anatomy of marijuana (Cannabis sativa L.). In *Cannabis sativa L. - Botany and Biotechnology*; New York City, 2017; pp. 123–136 ISBN 9783319545646.
30. Bergero, R.; Charlesworth, D. The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* **2009**, *24*, 94–102, doi:10.1016/j.tree.2008.09.010.
31. Henry, I.M.; Akagi, T.; Tao, R.; Comai, L. One Hundred Ways to Invent the Sexes: Theoretical and Observed Paths to Dioecy in Plants. *Annu. Rev. Plant Biol.* **2018**, *69*,

553–575, doi:10.1146/annurev-arplant-042817-040615.

32. Mandolino, G.; Carboni, A.; Forapani, S.; Faeti, V.; Ranalli, P. Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.). *Theor. Appl. Genet.* **1999**, *98*, 86–92, doi:10.1007/s001220051043.
33. Törjék, O., Bucherna, N., Kiss, E., Homoki, H., Finta-Korpelová, Z., Bócsa, I., Nagy, I., & Heszky, L. Novel male-specific molecular markers (MADC5, MADC6) in hemp. *Euphytica* **2002**, 209–218, doi:10.1023/A:1020204729122.
34. Di Stilio, V.S.; Kramer, E.M.; Baum, D.A. Floral MADS box genes and homeotic gender dimorphism in Thalictrum dioicum (Ranunculaceae) - A new model for the study of dioecy. *Plant J.* **2005**, *41*, 755–766, doi:10.1111/j.1365-313X.2005.02336.x.
35. Khadka, J.; Yadav, N.S.; Guy, M.; Grafi, G.; Golan-Goldhirsh, A. Epigenetic aspects of floral homeotic genes in relation to sexual dimorphism in the dioecious plant Mercurialis annua. *J. Exp. Bot.* **2019**, *70*, 6245–6259, doi:10.1093/jxb/erz379.
36. Pfent, C.; Pobursky, K.J.; Sather, D.N.; Golenberg, E.M. Characterization of SpAPETALA3 and SpPISTILLATA, B class floral identity genes in Spinacia oleracea, and their relationship to sexual dimorphism. *Dev. Genes Evol.* **2005**, *215*, 132–142, doi:10.1007/s00427-004-0459-4.
37. Sather, D.N.; Jovanovic, M.; Golenberg, E.M. Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. *BMC Plant Biol.* **2010**, *10*, doi:10.1186/1471-2229-10-46.
38. LaRue, N.C.; Sullivan, A.M.; Di Stilio, V.S. Functional recapitulation of transitions in sexual systems by homeosis during the evolution of dioecy in Thalictrum. *Front. Plant Sci.* **2013**, *4*, 1–10, doi:10.3389/fpls.2013.00487.
39. Barcaccia, G.; Palumbo, F.; Scariolo, F.; Vannozzi, A.; Borin, M.; Bona, S. Potentials and Challenges of Genomics for Breeding Cannabis Cultivars. *Front. Plant Sci.* **2020**, *11*, doi:10.3389/fpls.2020.573299.
40. Monthony, A.S.; Page, S.R.; Hesami, M.; Jones, A.M.P. The past, present and future of cannabis sativa tissue culture. *Plants* **2021**, *10*, 1–29, doi:10.3390/plants10010185.
41. Monthony, A.S.; Bagheri, S.; Zheng, Y.; Jones, A.M.P. Flower power: floral reversion

as a viable alternative to nodal micropropagation in Cannabis sativa. *Vitr. Cell. Dev. Biol. - Plant* **2021**, *57*, 1018–1030, doi:10.1007/s11627-021-10181-5.

42. Ram, H.Y.M.; Jaiswal, V.S. Induction of male flowers on female plants of Cannabis sativa by gibberellins and its inhibition by abscisic acid. *Planta* **1972**, *105*, 263–266, doi:10.1007/BF00385397.
43. Lubell, J.D.; Brand, M.H. Foliar sprays of silver thiosulfate produce male flowers on female hemp plants. *Horttechnology* **2018**, *28*, 743–747, doi:10.21273/HORTTECH04188-18.
44. Moon, Y.-H.Y.J.S.C.M.Y.C.J.-K.W.T. Effect of Timing of Ethephon Treatment on the Formation of Female Flowers and Seeds from Male Plant of Hemp (Cannabis sativa L.). *Korean J. Plant Resour.* **2020**, *33*, 682–688.
45. Liu, C.; Li, J.; Zhu, P.; Yu, J.; Hou, J.; Wang, C.; Long, D.; Yu, M.; Zhao, A. Mulberry EIL3 confers salt and drought tolerances and modulates ethylene biosynthetic gene expression. *PeerJ* **2019**, *2019*, 1–17, doi:10.7717/peerj.6391.
46. Charlesworth, D.; Mank, J.E. The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. *Genetics* **2010**, *186*, 9–31, doi:10.1534/genetics.110.117697.
47. Schlesinger, A.; Kramarsky-Winter, E.; Rosenfeld, H.; Armoza-Zvoloni, R.; Loya, Y. Sexual plasticity and self-fertilization in the sea anemone Aiptasia diaphana. *PLoS One* **2010**, *5*, 1–7, doi:10.1371/journal.pone.0011874.
48. Senthilkumaran, B. Pesticide- and sex steroid analogue-induced endocrine disruption differentially targets hypothalamo–hypophyseal–gonadal system during gametogenesis in teleosts – A review. *Gen. Comp. Endocrinol.* **2015**, *219*, 136–142, doi:10.1016/j.ygcen.2015.01.010.
49. Lambert, M.R.; Tran, T.; Kilian, A.; Ezaz, T.; Skelly, D.K. Molecular evidence for sex reversal in wild populations of green frogs (Rana clamitans). *PeerJ* **2019**, *2019*, 1–21, doi:10.7717/peerj.6449.
50. Perry, A.N.; Grober, M.S. A model for social control of sex change: Interactions of behavior, neuropeptides, glucocorticoids, and sex steroids. *Horm. Behav.* **2003**, *43*,

31–38, doi:10.1016/S0018-506X(02)00036-3.

51. Hayes, T.B.; Khoury, V.; Narayan, A.; Nazir, M.; Parka, A.; Brown, T.; Adame, L.; Chan, E.; Buchholz, D.; Stueve, T.; et al. Atrazine induces complete feminization and chemical castration in male African clawed frogs (Xenopus laevis). *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4612–4617, doi:10.1073/pnas.0909519107.
52. Ashby, R.; Forêt, S.; Searle, I.; Maleszka, R. MicroRNAs in Honey Bee Caste Determination. *Sci. Rep.* **2016**, *6*, 1–15, doi:10.1038/srep18794.
53. Cardoso-Júnior, C.A.M.; Fujimura, P.T.; Santos-Júnior, C.D.; Borges, N.A.; Ueira-Vieira, C.; Hartfelder, K.; Goulart, L.R.; Bonetti, A.M. Epigenetic modifications and their relation to caste and sex determination and adult division of labor in the stingless bee Melipona scutellaris. *Genet. Mol. Biol.* **2017**, *40*, 61–68, doi:10.1590/1678-4685-GMB-2016-0242.
54. Kurtz, L.E.; Brand, M.H.; Lubell-Brand, J.D. Production of tetraploid and triploid hemp. *HortScience* **2020**, *55*, 1703–1707, doi:10.21273/HORTSCI15303-20.
55. Flajšman, M.; Slapnik, M.; Murovec, J. Production of Feminized Seeds of High CBD Cannabis sativa L. by Manipulation of Sex Expression and Its Application to Breeding. *Front. Plant Sci.* **2021**, *12*, 1–12, doi:10.3389/fpls.2021.718092.
56. Coulter & Barnes *Botanical Gazette*; The University of Chicago Press, Ed.; 1901;
57. R Gane 1934_Gane_Production of ethylene by some ripening fruits. *Nature* 1934, *134*, 1008.
58. Chen, W.H.; Li, P.F.; Chen, M.K.; Lee, Y.I.; Yang, C.H. FOREVER YOUNG FLOWER negatively regulates ethylene response DNA-binding factors by activating an ethylene-responsive factor to control arabidopsis floral organ senescence and abscission. *Plant Physiol.* **2015**, *168*, 1666–1683, doi:10.1104/pp.15.00433.
59. Argueso, C.T.; Hansen, M.; Kieber, J.J. Regulation of ethylene biosynthesis. *J. Plant Growth Regul.* **2007**, *26*, 92–105, doi:10.1007/s00344-007-0013-5.
60. Dhakal, S.; Karki, M.; Subedi, P.; GC, A. Effect Of Ethephon Doses On Vegetative Characters, Sex Expression and Yield Of Cucumber (Cucumis sativus cv. Bhaktapur Local) In Resunga Municipality, Gulmi, Nepal. *Int. J. Appl. Sci. Biotechnol.* **2019**, *7*,

370–377, doi:10.3126/ijasbt.v7i3.25284.

61. Duggan, P.J. The Chemistry of Cannabis and Cannabinoids. *Aust. J. Chem.* **2021**, *74*, 369–387.

62. Spitzer-Rimon, B.; Duchin, S.; Bernstein, N.; Kamenetsky, R. Architecture and florogenesis in female Cannabis sativa plants. *Front. Plant Sci.* **2019**, *10*, doi:10.3389/fpls.2019.00350.

63. Bonini, S.A.; Premoli, M.; Tambaro, S.; Kumar, A.; Maccarinelli, G.; Memo, M.; Mastinu, A. Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history. *J. Ethnopharmacol.* **2018**, *227*, 300–315, doi:10.1016/j.jep.2018.09.004.

64. Gaoni, Y.; Mechoulam, R. Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish. *J. Am. Chem. Soc.* **1964**, *86*, 1646–1647, doi:10.1021/ja01062a046.

65. Shoyama, Y.; Yagi, M.; Nishioka, I.; Yamauchi, T. Biosynthesis of cannabinoid acids. *Phytochemistry* **1975**, *14*, 2189–2192, doi:10.1016/S0031-9422(00)91096-3.

66. Fischedick, J.T.; Hazekamp, A.; Erkelens, T.; Choi, Y.H.; Verpoorte, R. Metabolic fingerprinting of Cannabis sativa L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. *Phytochemistry* **2010**, *71*, 2058–2073, doi:10.1016/j.phytochem.2010.10.001.

67. Tuberosa, R. Marker-Assistedbreeding/breed, see also animal breedingmarker-assistedBreedingbreeding/breed, see also animal breedingin Crops BT - Encyclopedia of Sustainability Science and Technology. In; Meyers, R.A., Ed.; Springer New York: New York, NY, 2012; pp. 6425–6448 ISBN 978-1-4419-0851-3.

68. Palumbo, F., Galla, G., Vitulo, N., & Barcaccia, G. First draft genome sequencing of fennel (Foeniculum Vulgare Mill.): identification of simple sequence repeats and their application in marker-assisted breeding. *Mol. Breed* **2018**, *38*, 1–17, doi:10.1007/s11032-018-0884-0.

69. Patella, A., Palumbo, F., Galla, G., & Barcaccia, G. The molecular determination of hybridity and homozygosity estimates in breeding populations of lettuce (Lactuca

sativa L.). *Genes (Basel)* **2019**, doi:10.3390/genes10110916.

70. Ghedina, A.; Galla, G.; Cadalen, T.; Hilbert, J.L.; Caenazzo, S.T.; Barcaccia, G. A method for genotyping elite breeding stocks of leaf chicory (Cichorium intybus L.) by assaying mapped microsatellite marker loci Genetics. *BMC Res. Notes* **2015**, *8*, 1–12, doi:10.1186/s13104-015-1819-z.
71. Patella, A., Scariolo, F., Palumbo, F., & Barcaccia, G. Genetic structure of cultivated varieties of radicchio (Cichorium intybus l.): A comparison between f1 hybrids and synthetics. *Plants* **2018**, *8*, 213, doi:10.3390/plants8070213.
72. Nigam, R.K.; Varkey, M.; Reuben, D.E. Irradiation induced changes in flower formation in Cannabis sativa L. *Biol. Plant.* **1981**, *23*, 389–391, doi:10.1007/BF02877422.
73. NIGAM, R.K.; VARKEY, M.; REUBEN, D.E. Streptovaricin-induced Sex Expression in Male and Female Plants of Cannabis sativa L. *Ann. Bot.* **1981**, *47*, 169–172, doi:10.1093/oxfordjournals.aob.a085995.
74. Mohan Ram, H.Y.; Sett, R. Reversal of Ethephon-Induced Feminization in Male Plants of Cannabis sativa by Ethylene Antagonists. *Zeitschrift für Pflanzenphysiologie* **1982**, *107*, 85–89, doi:10.1016/s0044-328x(11)80012-7.
75. Gillan, R.; Cole, M.D.; Linacre, A.; Thorpe, J.W.; Watson, N.D. Comparison of Cannabis sativa by Random Amplification of Polymorphic DNA (RAPD) and HPLC of cannabinoids: a preliminary study. *Sci. Justice* **1995**, *35*, 169–177, doi:https://doi.org/10.1016/S1355-0306(95)72658-2.
76. Faeti, V.; Mandolino, G.; Ranalli, P. Genetic diversity of Cannabis sativa germplasm based on RAPD markers. *Plant Breed.* **1996**, *115*, 367–370, doi:https://doi.org/10.1111/j.1439-0523.1996.tb00935.x.
77. Jagadish, V.; Robertson, J.; Gibbs, A. RAPD analysis distinguishes Cannabis sativa samples from different sources. *Forensic Sci. Int.* **1996**, *79*, 113–121, doi:https://doi.org/10.1016/0379-0738(96)01898-1.
78. Forapani, S.; Carboni, A.; Paoletti, C.; Moliterni, V.M.C.; Ranalli, P.; Mandolino, G. Comparison of Hemp Varieties Using Random Amplified Polymorphic DNA

Markers. *Crop Sci.* **2001**, *41*, 1682–1689, doi:https://doi.org/10.2135/cropsci2001.1682.

79. Datwyler, S.L.; Weiblen, G.D. Genetic Variation in Hemp and Marijuana (Cannabis sativa L.) According to Amplified Fragment Length Polymorphisms*. *J. Forensic Sci.* **2006**, *51*, 371–375, doi:https://doi.org/10.1111/j.1556-4029.2006.00061.x.
80. Alghanim, H.J.; Almirall, J.R. Development of microsatellite markers in Cannabis sativa for DNA typing and genetic relatedness analyses. *Anal. Bioanal. Chem.* **2003**, *376*, 1225–1233, doi:10.1007/s00216-003-1984-0.
81. Gilmore, S.; Peakall, R.; Robertson, J. Short tandem repeat (STR) DNA markers are hypervariable and informative in Cannabis sativa: Implications for forensic investigations. *Forensic Sci. Int.* **2003**, *131*, 65–74, doi:10.1016/S0379-0738(02)00397-3.
82. Hsieh, H.M.; Hou, R.J.; Tsai, L.C.; Wei, C.S.; Liu, S.W.; Huang, L.H.; Kuo, Y.C.; Linacre, A.; Lee, J.C.I. A highly polymorphic STR locus in Cannabis sativa. *Forensic Sci. Int.* **2003**, *131*, 53–58, doi:10.1016/S0379-0738(02)00395-X.
83. Mendoza, M.A.; Mills, D.E.K.; Lata, H.; Chandra, S.; Elsohly, M.A.; Almirall, J.R. Genetic individualization of Cannabis sativa by a short tandem repeat multiplex system. *Anal. Bioanal. Chem.* **2009**, *393*, 719–726, doi:10.1007/s00216-008-2500-3.
84. Dufresnes, C.; Jan, C.; Bienert, F.; Goudet, J.; Fumagalli, L. Broad-Scale Genetic Diversity of Cannabis for Forensic Applications. *PLoS One* **2017**, *12*, e0170522.
85. Soler, S.; Gramazio, P.; Figàs, M.R.; Vilanova, S.; Rosa, E.; Llosa, E.R.; Borràs, D.; Plazas, M.; Prohens, J. Genetic structure of Cannabis sativa var. indica cultivars based on genomic SSR (gSSR) markers: Implications for breeding and germplasm management. *Ind. Crops Prod.* **2017**, *104*, 171–178, doi:10.1016/j.indcrop.2017.04.043.
86. Gao, C.; Xin, P.; Cheng, C.; Tang, Q.; Chen, P.; Wang, C.; Zang, G.; Zhao, L. Diversity analysis in Cannabis sativabased on large-scale development of expressed sequence tag-derived simple sequence repeat markers. *PLoS One* **2014**, *9*, doi:10.1371/journal.pone.0110638.

87. Soorni, A., Fatahi, R., Haak, D. C., Salami, S. A., & Bombarely, A. Assessment of genetic diversity and population structure in Iranian cannabis germplasm. *Sci. Rep* **2017**, *7*.

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

89. Weiblen, G.D.; Wenger, J.P.; Craft, K.J.; ElSohly, M.A.; Mehmedic, Z.; Treiber, E.L.; Marks, M.D. Gene duplication and divergence affecting drug content in Cannabis sativa. *New Phytol.* **2015**, *208*, 1241–1250, doi:https://doi.org/10.1111/nph.13562.

90. 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.; et al. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res.* **2019**, *29*, 146–156, doi:10.1101/gr.242594.118.

91. Maione, S.; Piscitelli, F.; Gatta, L.; Vita, D.; De Petrocellis, L.; Palazzo, E.; de Novellis, V.; Di Marzo, V. Non-psychoactive cannabinoids modulate the descending pathway of antinociception in anaesthetized rats through several mechanisms of action. *Br. J. Pharmacol.* **2011**, *162*, 584–596, doi:https://doi.org/10.1111/j.1476-5381.2010.01063.x.

92. Henry, P.; Khatodia, S.; Kapoor, K.; Gonzales, B.; Middleton, A.; Hong, K.; Hilyard, A.; Johnson, S.; Allen, D.; Chester, Z.; et al. A single nucleotide polymorphism assay sheds light on the extent and distribution of genetic diversity, population structure and functional basis of key traits in cultivated north American cannabis. *J. Cannabis Res.* **2020**, *2*, doi:10.1186/s42238-020-00036-y.

93. McKernan, K.J.; Helbert, Y.; Kane, L.T.; Ebling, H.; Zhang, L.; Liu, B.; Eaton, Z.; McLaughlin, S.; Kingan, S.; Baybayan, P.; et al. Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. *bioRxiv* **2020**, doi:10.1101/2020.01.03.894428.

94. Gao, S.; Wang, B.; Xie, S.; Xu, X.; Zhang, J.; Pei, L.; Yu, Y.; Yang, W.; Zhang, Y. A

high-quality reference genome of wild Cannabis sativa. *Hortic. Res.* **2020**, *7*, doi:10.1038/s41438-020-0295-3.

95. Bernstein, N.; Gorelick, J.; Zerahia, R.; Koch, S. Impact of N, P, K, and humic acid supplementation on the chemical profile of medical cannabis (Cannabis sativa L). *Front. Plant Sci.* **2019**, *10*, doi:10.3389/fpls.2019.00736.
96. Danziger, N.; Bernstein, N. Light matters: Effect of light spectra on cannabinoid profile and plant development of medical cannabis (Cannabis sativa L.). *Ind. Crops Prod.* **2021**, *164*, 113351, doi:10.1016/j.indcrop.2021.113351.
97. Kong, Y.; Zheng, Y. Phototropin is partly involved in blue-light-mediated stem elongation, flower initiation, and leaf expansion: A comparison of phenotypic responses between wild Arabidopsis and its phototropin mutants. *Environ. Exp. Bot.* **2020**, *171*, 103967, doi:https://doi.org/10.1016/j.envexpbot.2019.103967.
98. Ouzounis, T.; Rosenqvist, E.; Ottosen, C.-O. Spectral Effects of Artificial Light on Plant Physiology and Secondary Metabolism: A Review. *HortScience horts* **2015**, *50*, 1128–1135, doi:10.21273/HORTSCI.50.8.1128.
99. Hawkesford, M.; Horst, W.; Kichey, T.; Lambers, H.; Schjoerring, J.; Møller, I.S.; White, P. Chapter 6 - Functions of Macronutrients. In; Marschner, P.B.T.-M.M.N. of H.P. (Third E., Ed.; Academic Press: San Diego, 2012; pp. 135–189 ISBN 978-0-12-384905-2.
100. Saloner, A.; Bernstein, N. Nitrogen supply affects cannabinoid and terpenoid profile in medical cannabis (Cannabis sativa L.). *Ind. Crops Prod.* **2021**, *167*, 113516, doi:10.1016/j.indcrop.2021.113516.
101. Saloner, A.; Bernstein, N. Response of Medical Cannabis (Cannabis sativa L.) to Nitrogen Supply Under Long Photoperiod. *Front. Plant Sci.* **2020**, *11*, doi:10.3389/fpls.2020.572293.
102. Britto, D.T.; Kronzucker, H.J. NH4+ toxicity in higher plants: a critical review. *J. Plant Physiol.* **2002**, *159*, 567–584, doi:https://doi.org/10.1078/0176-1617-0774.
103. Saloner, A.; Bernstein, N. Nitrogen Source Matters: High NH4/NO3 Ratio Reduces Cannabinoids, Terpenoids, and Yield in Medical Cannabis. *Front. Plant Sci.* **2022**, *13*,

1–17, doi:10.3389/fpls.2022.830224.

104. Shiponi, S.; Bernstein, N. The Highs and Lows of P Supply in Medical Cannabis: Effects on Cannabinoids, the Ionome, and Morpho-Physiology. *Front. Plant Sci.* **2021**, *12*, 1–22, doi:10.3389/fpls.2021.657323.
105. Shiponi, S.; Bernstein, N. Response of medical cannabis (Cannabis sativa L.) genotypes to P supply under long photoperiod: Functional phenotyping and the ionome. *Ind. Crops Prod.* **2021**, *161*, 113154, doi:10.1016/j.indcrop.2020.113154.
106. Grzebisz, W.; Gransee, A.; Szczepaniak, W.; Diatta, J. The effects of potassium fertilization on water-use efficiency in crop plants. *J. Plant Nutr. Soil Sci.* **2013**, *176*, 355–374, doi:https://doi.org/10.1002/jpln.201200287.
107. Saloner, A.; Sacks, M.M.; Bernstein, N. Response of Medical Cannabis (Cannabis sativa L.) Genotypes to K Supply Under Long Photoperiod. *Front. Plant Sci.* **2019**, *10*, 1–16, doi:10.3389/fpls.2019.01369.
108. Saloner, A.; Bernstein, N. Effect of Potassium (K) Supply on Cannabinoids, Terpenoids and Plant Function in Medical Cannabis. *Agronomy* **2022**, *12*, doi:10.3390/AGRONOMY12051242.
109. Sarma, N.D.; Waye, A.; Elsohly, M.A.; Brown, P.N.; Elzinga, S.; Johnson, H.E.; Marles, R.J.; Melanson, J.E.; Russo, E.; Deyton, L.; et al. Cannabis Inflorescence for Medical Purposes: USP Considerations for Quality Attributes. *J. Nat. Prod.* **2020**, *83*, 1334–1351, doi:10.1021/acs.jnatprod.9b01200.
110. Danziger, N.; Bernstein, N. Too Dense or Not Too Dense: Higher Planting Density Reduces Cannabinoid Uniformity but Increases Yield/Area in Drug-Type Medical Cannabis. *Front. Plant Sci.* **2022**, *13*, 1–18, doi:10.3389/fpls.2022.713481.
111. Danziger, N.; Bernstein, N. Shape matters: Plant architecture affects chemical uniformity in large-size medical cannabis plants. *Plants* **2021**, *10*, doi:10.3390/plants10091834.
112. Corredor-Perilla, I.C.; Cuervo Andrade, J.L.; Olejar, K.J.; Park, S.H. Beneficial properties of soil bacteria from Cannabis sativa L.: Seed germination, phosphorus solubilization and mycelial growth inhibition of Fusarium sp. *Rhizosphere* **2023**, *27*,

100780, doi:10.1016/j.rhisph.2023.100780.

113. Conant, R.T.; Walsh, R.P.; Walsh, M.; Bell, C.W.; Wallenstein, M.D. Effects of a Microbial Biostimulant, Mammoth PTM, on Cannabis sativa Bud Yield. *J. Hortic.* **2017**, *04*, 1–5, doi:10.4172/2376-0354.1000191.

114. Ahmed, B.; Hijri, M. Potential impacts of soil microbiota manipulation on secondary metabolites production in cannabis. *J. Cannabis Res.* **2021**, *3*, doi:10.1186/s42238-021-00082-0.

115. Comeau, D.; Balthazar, C.; Novinscak, A.; Bouhamdani, N.; Joly, D.L.; Filion, M. Interactions Between Bacillus Spp., Pseudomonas Spp. and Cannabis sativa Promote Plant Growth. *Front. Microbiol.* **2021**, *12*, doi:10.3389/fmicb.2021.715758.

116. Punja, Z.K. Emerging diseases of Cannabis sativa and sustainable management. *Pest Manag. Sci.* **2021**, *77*, 3857–3870, doi:10.1002/ps.6307.

117. Mendel, P.; Lalge, A.B.; Vyhnanek, T.; Trojan, V.; Maassen, H.; Havel, L. Progress in Early Sex Determination of Cannabis Plant By Dna Markers. *MendelNet2016* **2016**, *1*, 731–735.

118. Hesami, M.; Pepe, M.; Alizadeh, M.; Rakei, A.; Baiton, A.; Phineas Jones, A.M. Recent advances in cannabis biotechnology. *Ind. Crops Prod.* **2020**, *158*, 113026, doi:10.1016/j.indcrop.2020.113026.

119. Zhang, L.; Yuan, M.; Tao, A.; Xu, J.; Lin, L.; Fang, P.; Qi, J. Genetic structure and relationship analysis of an association population in jute (Corchorus spp.) evaluated by SSR markers. *PLoS One* **2015**, *10*, 1–9, doi:10.1371/journal.pone.0128195.

120. Gonzaga, Z.J.; Aslam, K.; Septiningsih, E.M.; Collard, B.C.Y. Evaluation of SSR and SNP Markers for Molecular Breeding in Rice. *Plant Breed. Biotechnol.* **2015**, *3*, 139–152, doi:10.9787/pbb.2015.3.2.139.

121. Dufresnes, C., Jan, C., Bienert, F., Goudet, J., & Fumagalli, L. Broad-scale genetic diversity of cannabis for forensic applications. *PLoS One* **2012**.

122. Grassa, C. J., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., Schwartz, C., & Weiblen, G.D. A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *BioRxiv* **2018**,

doi:10/1101/458083.

123. Hillig, K.W. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genet. Resour. Crop Evol.* **2005**, *52*, 161–180.

124. Sawler, J., Stout, J. M., Gardner, K. M., Hudson, D., Vidmar, J., Butler, L., Page, J. E., & Myles, S. The genetic structure of marijuana and hemp. *PLoS One* **2015**, *10*.

125. Hu, Z. G., Guo, H. Y., Hu, X. L., Chen, X., Liu, X Y., Guo, M. B., Zhang, Q. Y., Xu, Y. P., Guo, L. F., & Yang, M. Genetic diversity research of hemp (Cannabis sativa L) cultivar based on AFLP analysis. *Plant Gene. Res.* **2012**, *13*, 555–561.

126. Zhang, Q., Chen, X., Guo, H., Trindade, L. M., Salentijn, E. M., Guo, R., Guo, M., Xu, Y., & Yang, M. Latitudinal adaptation and genetic insights into the origins of Cannabis sativa L. *Front. Plant Sci.* **2018**, *9*.

127. Borin, M.; Palumbo, F.; Vannozzi, A.; Scariolo, F.; Sacilotto, G.B.; Gazzola, M.; Barcaccia, G. Developing and testing molecular markers in cannabis sativa (Hemp) for their use in variety and dioecy assessments. *Plants* **2021**, *10*, doi:10.3390/plants10102174.

128. Vergara, D.; White, K.H.; Keepers, K.G.; Kane, N.C. The complete chloroplast genomes of Cannabis sativa and Humulus lupulus. *Mitochondrial DNA Part A* **2016**, *27*, 3793–3794, doi:10.3109/19401736.2015.1079905.

129. White, K.H.; Vergara, D.; Keepers, K.G.; Kane, N.C. The complete mitochondrial genome for Cannabis sativa. *Mitochondrial DNA Part B* **2016**, *1*, 715–716, doi:10.1080/23802359.2016.1155083.

130. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422, doi:10.1007/s00122-002-1031-0.

131. Portis, E.; Lanteri, S.; Barchi, L.; Portis, F.; Valente, L.; Toppino, L.; Rotino, G.L.; Acquadro, A. Comprehensive characterization of simple sequence repeats in eggplant (Solanum melongena L.) genome and construction of a web resource. *Front. Plant Sci.* **2018**, *9*, 1–15, doi:10.3389/fpls.2018.00401.

132. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, 1–12, doi:10.1093/nar/gks596.
133. Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234, doi:10.1038/72708.
134. Yeh F. C. & Boyle T. J. B. Population Genetic Analysis of Codominant and Dominant Markers and Quantitative Traits. *Belgian J. Bot.* **1997**, *129*, 157–163.
135. Levene, H. On a matching problem arising in genetics. *Ann Math Stat.* **1949**, *20*, 1.
136. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 3321–3323, doi:10.1073/pnas.70.12.3321.
137. Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **1978**, *89*, 583–590, doi:10.1093/genetics/89.3.583.
138. Wright, S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution (N. Y).* **1965**, *19*, 395–420.
139. Wright, S. *Variability within and among natural populations. Evolution and the genetics of populations*; 4th ed.; Univ. of Chicago Press., 1978;
140. Lewontin, R.C. *The genetic basis of evolutionary change*; Columbia U.; New York, 1974;
141. McDonald, B.A.; McDermott, J.M. Population genetics of plant pathogenic fungi. *Bioscience* **1993**, *43*, 9.
142. Rohlf, F.J. *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System ver. 2.2*; Exeter Publishing: New York, 2009; ISBN 0925031313.
143. Dice, L.R. Measurement of the amount of ecological association between species. *Ecology* **1945**, *26*, 297–302.
144. Swofford, D.L. *PUAP: Phylogenetic Analysis Using Parsimong, ver. 4.0b*; Sinauer Associates, Inc: Sunderland, Massachusetts, USA., 2000;
145. Øyvind H., Harper D.A.T., Ryan P.D. PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 9.
146. Falush D., Stephens M., & P.J.K. Inference of population structure using multilocus

genotype data: linked loci and correlated allele frequencies. *Genetics* **2003**, *131*, 87.

147. Evanno G, Regnaut S, & G.J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **2005**, *14*, 20.
148. Peakall, R., & Smouse, P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic analysis for teaching and research-an update. *Bioinformatics* **2012**, *28*, 2537–2539.
149. Botstein D., White R.L., Skolnick M., & D.R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.
150. Maria, P.; Štiasna, K.; Vyhnánek, T.; Trojan, V.; Mrkvicova, E.; Hřivna, L.; Havel, L. Analysis of microsatellite markers in hemp (Cannabis sativa l.). In Proceedings of the MendelNet; International Ph.D. Students Conference on MendelNet 2015: Fac Agron, Brno, Czech Republic, 2015.
151. Maria, P., Štiasna, K., Vyhnánek, T., Trojan, V., Mrkvicova, E., Hřivna, L., & Havel, L. Analysis of microsatellite markers in hemp (Cannabis sativa l.). In Proceedings of the International Ph.D. Students Conference on MendelNet; 2015.
152. Kayis, S.A.; Hakki, E.E.; Pinarkara, E. Comparison of effectiveness of ISSR and RAPD markers in genetic characterization of seized marijuana (Cannabis sativa L.) in Turkey. *African J. Agric. Res.* **2010**, *5*, 2925–2933.
153. Palumbo, F., Galla, G., & Barcaccia, G. Developing a Molecular Identification Assay of Old Landraces for the Genetic Authentication of Typical Agro-Food Products: The Case Study of the Barley "Agordino." *Food Technol. Biotechnol.* **2017**, *55*, 29–39.
154. Pisanti, S., & Bifulco, M. Medical Cannabis: A plurimillennial history of an evergreen. *J. Cell. Physiol.* **2019**, *234*, 8342–8351.
155. Ren, G.; Zhang, X.; Li, Y.; Ridout, K.; Serrano-Serrano, M.L.; Yang, Y.; Liu, A.; Ravikanth, G.; Nawaz, M.A.; Mumtaz, A.S.; et al. Large-scale whole-genome sequencing unravels the domestication history of Cannabis sativa. *Sci. Adv.* **2021**, *In Press*, 1–13.
156. Vyhnánek, T., Nevrtalová, E., Bjelková, M., & Balgová, B. SSR loci survey of

technical hemp cultivars: The optimization of a cost-effective analyses to study genetic variability. *Plant Sci.* **2020**, *298*.

157. Kohjyouma, M.; Lee, I.J.; Iida, O.; Kurihara, K.; Yamada, K.; Makino, Y.; Sekita, S.; Satake, M. Intraspecific variation in Cannabis sativa L. based on intergenic spacer region of chloroplast DNA. *Biol. Pharm. Bull.* **2000**, *23*, 727–730, doi:10.1248/bpb.23.727.
158. Dias, V.H.G.; Ribeiro, A.S.D.; Mello, I.C.T.; Silva, R.; Sabino, B.D.; Garrido, R.G.; Seldin, L.; Moura-Neto, R.S. Genetic identification of Cannabis sativa using chloroplast trnL-F gene. *Forensic Sci. Int. Genet.* **2015**, *14*, 201–202, doi:10.1016/j.fsigen.2014.10.003.
159. Matielo, C.B.D.O.; Lemos, R.P.M.; Sarzi, D.S.; Machado, L. de O.; Beise, D.C.; Dobbler, P.C.T.; Castro, R.M.; Fett, M.S.; Roesch, L.F.W.; Camargo, F.A. de O.; et al. Whole Plastome Sequences of Two Drug-Type Cannabis: Insights Into the Use of Plastid in Forensic Analyses. *J. Forensic Sci.* **2020**, *65*, 259–265, doi:10.1111/1556-4029.14155.
160. Deng, G.; Yang, M.; Zhao, K.; Yang, Y.; Huang, X.; Cheng, X. The complete chloroplast genome of Cannabis sativa variety Yunma 7. *Mitochondrial DNA Part B Resour.* **2021**, *6*, 531–532, doi:10.1080/23802359.2021.1873709.
161. Roman, M.G.; Houston, R. Investigation of chloroplast regions rps16 and clpP for determination of Cannabis sativa crop type and biogeographical origin. *Leg. Med.* **2020**, *47*, doi:10.1016/j.legalmed.2020.101759.
162. Cheng, Y.C.; Houston, R. Evaluation of the trnK-matK-trnK, ycf3, and accD-psal chloroplast regions to differentiate crop type and biogeographical origin of Cannabis sativa. *Int. J. Legal Med.* **2021**, *135*, 1235–1244, doi:10.1007/s00414-021-02518-x.
163. Roman, M.G.; Gutierrez, R.; Houston, R. Massively parallel sequencing of Cannabis sativa chloroplast hotspots for forensic typing. *J. Cannabis Res.* **2022**, *4*, doi:10.1186/s42238-022-00123-2.
164. Roman, M.G.; Gangitano, D.; Houston, R. Characterization of new chloroplast markers to determine biogeographical origin and crop type of Cannabis sativa. *Int. J.*

Legal Med. **2019**, *133*, 1721–1732, doi:10.1007/s00414-019-02142-w.

165. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612, doi:10.1002/jcc.20084.
166. Song, B.H.; Wang, X.Q.; Li, F.Z.; Hong, D.Y. Further evidence for paraphyly of the Celtidaceae from the chloroplast gene matK. *Plant Syst. Evol.* **2001**, *228*, 107–115, doi:10.1007/s006060170041.
167. Yeo, H.C.; Reddy, V.A.; Mun, B.G.; Leong, S.H.; Dhandapani, S.; Rajani, S.; Jang, I.C. Comparative Transcriptome Analysis Reveals Coordinated Transcriptional Regulation of Central and Secondary Metabolism in the Trichomes of Cannabis Cultivars. *Int. J. Mol. Sci.* **2022**, *23*, doi:10.3390/ijms23158310.
168. Happyana, N.; Agnolet, S.; Muntendam, R.; Van Dam, A.; Schneider, B.; Kayser, O. Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR. *Phytochemistry* **2013**, *87*, 51–59, doi:10.1016/j.phytochem.2012.11.001.
169. Capriotti, A.L.; Cannazza, G.; Catani, M.; Cavaliere, C.; Cavazzini, A.; Cerrato, A.; Citti, C.; Felletti, S.; Montone, C.M.; Piovesana, S.; et al. Recent applications of mass spectrometry for the characterization of cannabis and hemp phytocannabinoids: From targeted to untargeted analysis. *J. Chromatogr. A* **2021**, *1655*, 462492, doi:10.1016/j.chroma.2021.462492.
170. Adal, A.M.; Doshi, K.; Holbrook, L.; Mahmoud, S.S. Comparative RNA-Seq analysis reveals genes associated with masculinization in female Cannabis sativa. *Planta* **2021**, *253*, 1–17, doi:10.1007/s00425-020-03522-y.
171. Prentout, D.; Razumova, O.; Rhoné, B.; Badouin, H.; Henri, H.; Feng, C.; Käfer, J.; Karlov, G.; Marais, G.A.B. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of Cannabis sativa. *Genome Res.* **2020**, *30*, 164–172, doi:10.1101/gr.251207.119.
172. Monthony, A.S.; Ronne, M. de; Torkamaneh, D. Exploring Ethylene-Related Genes in Cannabis sativa: Implications for Sexual Plasticity. *bioRxiv* **2023**,

2023.04.28.538750, doi:10.1101/2023.04.28.538750.

173. Aizpurua-Olaizola, O.; Omar, J.; Navarro, P.; Olivares, M.; Etxebarria, N.; Usobiaga, A. Identification and quantification of cannabinoids in Cannabis sativa L. plants by high performance liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2014**, *406*, 7549–7560, doi:10.1007/s00216-014-8177-x.

174. Stiasna, K.; Presinszka, M.; Vyhnanek, T.; Trojan, V.; Mrkvicova, E.; Hrivna, L.; Havel, L. Analysis of Genes From Cannabinoid Biosynthetic Pathway. *MendelNet 2015* **2015**, 6.

175. McGarvey, P.; Huang, J.; McCoy, M.; Orvis, J.; Katsir, Y.; Lotringer, N.; Nesher, I.; Kavarana, M.; Sun, M.; Peet, R.; et al. De novo assembly and annotation of transcriptomes from two cultivars of Cannabis sativa with different cannabinoid profiles. *Gene* **2020**, *762*, 145026, doi:10.1016/j.gene.2020.145026.

176. Sands, L.B.; Haiden, S.R.; Ma, Y.; Berkowitz, G.A. Hormonal control of promoter activities of Cannabis sativa prenyltransferase 1 and 4 and salicylic acid mediated regulation of cannabinoid biosynthesis. *Sci. Rep.* **2023**, *13*, 1–12, doi:10.1038/s41598-023-35303-4.

177. Sirikantaramas, S.; Taura, F.; Tanaka, Y.; Ishikawa, Y.; Morimoto, S.; Shoyama, Y. Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant Cell Physiol.* **2005**, *46*, 1578–1582, doi:10.1093/pcp/pci166.

178. Taura, F.; Sirikantaramas, S.; Shoyama, Y.; Yoshikai, K.; Shoyama, Y.; Morimoto, S. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. *FEBS Lett.* **2007**, *581*, 2929–2934, doi:10.1016/j.febslet.2007.05.043.

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

180. Deguchi, M.; Potlakayala, S.; Spuhler, Z.; George, H.; Sheri, V.; Agili, R.; Patel, A.; Rudrabhatla, S. Selection and validation of reference genes for normalization of qRT-PCR data to study the cannabinoid pathway genes in industrial hemp. *PLoS One* **2021**, *16*, 1–17, doi:10.1371/journal.pone.0260660.
181. Guo, R.; Guo, H.; Zhang, Q.; Guo, M.; Xu, Y.; Zeng, M.; Lv, P.; Chen, X.; Yang, M. Evaluation of reference genes for RT-qPCR analysis in wild and cultivated Cannabis. *Biosci. Biotechnol. Biochem.* **2018**, *82*, 1902–1910, doi:10.1080/09168451.2018.1506253.
182. Livak, K.J.; Schmittgen, T.D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **2001**, *25*, 402–408, doi:https://doi.org/10.1006/meth.2001.1262.
183. van de Poel, B.; Bulens, I.; Markoula, A.; Hertog, M.L.A.T.M.; Dreesen, R.; Wirtz, M.; Vandoninck, S.; Oppermann, Y.; Keulemans, J.; Hell, R.; et al. Targeted systems biology profiling of tomato fruit reveals coordination of the Yang cycle and a distinct regulation of ethylene biosynthesis during postclimacteric ripening. *Plant Physiol.* **2012**, *160*, 1498–1514, doi:10.1104/pp.112.206086.
184. Yang, C.; Lu, X.; Ma, B.; Chen, S.Y.; Zhang, J.S. Ethylene signaling in rice and arabidopsis: Conserved and diverged aspects. *Mol. Plant* **2015**, *8*, 495–505, doi:10.1016/j.molp.2015.01.003.
185. Chen, M.K.; Hsu, W.H.; Lee, P.F.; Thiruvengadam, M.; Chen, H.I.; Yang, C.H. The MADS box gene, FOREVER YOUNG FLOWER, acts as a repressor controlling floral organ senescence and abscission in Arabidopsis. *Plant J.* **2011**, *68*, 168–185, doi:10.1111/j.1365-313X.2011.04677.x.
186. Chen, H.; Sun, J.; Li, S.; Cui, Q.; Zhang, H.; Xin, F.; Wang, H.; Lin, T.; Gao, D.; Wang, S.; et al. An ACC Oxidase Gene Essential for Cucumber Carpel Development. *Mol. Plant* **2016**, *9*, 1315–1327, doi:10.1016/j.molp.2016.06.018.
187. Dubois, M.; Van den Broeck, L.; Inzé, D. The Pivotal Role of Ethylene in Plant Growth. *Trends Plant Sci.* **2018**, *23*, 311–323, doi:10.1016/j.tplants.2018.01.003.
188. Binder, B.M. Ethylene signaling in plants. *J. Biol. Chem.* **2020**, *295*, 7710–7725,

doi:10.1074/jbc.REV120.010854.

189. Hostetler, G.L.; Ralston, R.A.; Schwartz, S.J. Flavones : Food Sources , Bioavailability ,. **2017**, 423–435, doi:10.3945/an.116.012948.FIGURE.

190. Panche, A.N.; Diwan, A.D.; Chandra, S.R. Flavonoids: An overview. *J. Nutr. Sci.* **2016**, *5*, doi:10.1017/jns.2016.41.

191. Rue, E.A.; Rush, M.D.; van Breemen, R.B. Procyanidins: a comprehensive review encompassing structure elucidation via mass spectrometry. *Phytochem. Rev.* **2018**, *17*, 1–16, doi:10.1007/s11101-017-9507-3.

192. Aron, P.M.; Kennedy, J.A. Flavan-3-ols: Nature, occurrence and biological activity. *Mol. Nutr. Food Res.* **2008**, *52*, 79–104, doi:https://doi.org/10.1002/mnfr.200700137.

193. Aizpurua-Olaizola, O.; Soydaner, U.; Öztürk, E.; Schibano, D.; Simsir, Y.; Navarro, P.; Etxebarria, N.; Usobiaga, A. Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes. *J. Nat. Prod.* **2016**, *79*, 324–331, doi:10.1021/acs.jnatprod.5b00949.

194. Genotypes, C.L.; Burgel, L.; Hartung, J.; Pflugfelder, A.; Grae, S. Impact of Growth Stage and Biomass Fractions on Cannabinoid Content and Yield of Di ff erent Hemp. *Mol. Agron.* **2020**, *10*, 1–17.

195. Noppawan, P.; Bainier, C.; Lanot, A.; McQueen-Mason, S.; Supanchaiyamat, N.; Attard, T.M.; Hunt, A.J. Effect of harvest time on the compositional changes in essential oils, cannabinoids, and waxes of hemp (Cannabis sativa L.). *R. Soc. Open Sci.* **2022**, *9*, doi:10.1098/rsos.211699.

196. Linder, E.R.; Young, S.; Li, X.; Henriquez Inoa, S.; Suchoff, D.H. The Effect of Harvest Date on Temporal Cannabinoid and Biomass Production in the Floral Hemp (Cannabis sativa L.) Cultivars BaOx and Cherry Wine. *Horticulturae* **2022**, *8*, doi:10.3390/horticulturae8100959.

197. Pavlovic, R.; Panseri, S.; Giupponi, L.; Leoni, V.; Citti, C.; Cattaneo, C.; Cavaletto, M.; Giorgi, A. Phytochemical and Ecological Analysis of Two Varieties of Hemp (Cannabis sativa L.) Grown in a Mountain Environment of Italian Alps. *Front. Plant Sci.* **2019**, *10*, 1–20, doi:10.3389/fpls.2019.01265.

198. Apicella, P. V.; Sands, L.B.; Ma, Y.; Berkowitz, G.A. Delineating genetic regulation of cannabinoid biosynthesis during female flower development in Cannabis sativa. *Plant Direct* **2022**, *6*, 1–8, doi:10.1002/pld3.412.
199. Braich, S.; Baillie, R.C.; Jewell, L.S.; Spangenberg, G.C.; Cogan, N.O.I. Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis. *Sci. Rep.* **2019**, *9*, 1–12, doi:10.1038/s41598-019-53023-6.

Supplementary Materials

Supplementary Table S1: Multiplex SSR Primer List

Molecular Marker	Expected Size	Multiplex	Fluorescent Dye	Primer Annealing Temp. (°C)	Start	End	Motif type	Forward/ Reverse	Primer Sequence
SSR_6-3-F	180-200	1	M13	55 °C	35,062,092	35,062,261	(AAT)10	Forward	ATCTCATTTTCCGTACCTGTT
								Reverse	CTAATTCTCAACTTAACCGCG
SSR_2-2	250-270	1	M13	55 °C	27,019,093	27,019,345	(TGA)12	Forward	TAGTAGTAGTAGTGCCTGAGG
								Reverse	ACCTTAACAACACCACAACTA
SSR_X-1	390-450	1	M13	55 °C	12,090,959	12,091,352	(TC)40	Forward	TTGTCAAGGGAGCTTAGTTAG
								Reverse	ATGTGTATTTCTCGCCTGTTA
SSR_4-2	230-260	1	PAN1	55 °C	38,738,240	38,738,472	(AT)17	Forward	CAGAGTTTGGTCCTTTTCAAA
								Reverse	CACGGATTTTAAGCATTGGAT
SSR_2-3	350-410	1	PAN1	55 °C	49,240,375	49,240,744	(GA)22	Forward	CTCCCTGCCATTAGACAAATA
								Reverse	CCAGGAGGTAATTTTCTGCTA
SSR_7-3	230-280	1	PAN2	55 °C	51,776,452	51,776,692	(CT)22	Forward	ACTGTGAACTGTCCTTTTACA
								Reverse	AACAACCTGAAATCCGAAAAG
SSR_3-3	250-300	1	PAN3	55 °C	59,258,629	59,258,880	(AG)21	Forward	CAAAGAAAGCAGGCATTAGTT
								Reverse	CTCTCTGTGAATGTGATCTGT
SSR_2-1	240-260	2	M13	55 °C	15,695,145	15,695,388	(AAT)11	Forward	GGCAGGAAAAATCTCAAACAT
								Reverse	ACATTGGAATTAGACAGAGCA
SSR_4-1	230-270	2	PAN1	55 °C	3,414,697	3,414,947	(ATA)21	Forward	GTTGGTTATGTGTTAGGGTCT
								Reverse	GTTATGGACAAACAATGCATG
SSR_8-2	180-220	2	PAN2	55 °C	13,924,026	13,924,199	(CT)21	Forward	CATCACACCAGGTACCAATAT
								Reverse	CATGAAACAACGTTGGGTTAT
SSR_5-2	250-300	2	PAN2	55 °C	34,558,385	34,558,643	(CT)32	Forward	TGGCTGAAAGTAAGAAAAGAC
								Reverse	TTATCGCTCAAAACACTCAAC
SSR_6-1	200-270	2	PAN3	55 °C	3,764,859	3,765,058	(AT)17	Forward	ACTTCACATGAGATTGAGAACA
								Reverse	TCCTTTGGATTCATTAAGTTGT
SSR_X-3	280-350	2	PAN3	55 °C	71,305,129	71,305,410	(TC)41	Forward	ACAGTAGTTTTCAGGGTTGAA
								Reverse	TCACACCAATATCTATCAGCC
SSR_1-4	180-220	3	M13	55 °C	86,039,144	86,039,328	(TTA)17	Forward	TCAAGTTACGTAATCCCCAAA
								Reverse	CCTAAGCACAAGGTTAAATCAT
SSR_3-1	300-340	3	M13	55 °C	12,247,530	12,247,829	(TC)32	Forward	TGATTTTGCGACCCTTTTATG
								Reverse	CTTTTGCAGGTACATCCAAAA
SSR_8-4	280-330	3	PAN1	55 °C	50,925,135	50,925,396	(TC)22	Forward	TATGCATCCATTGTACCTGTT
								Reverse	TAATGTTTGTGTGTGTGCAAA
SSR_9-4	110-150	4	PAN1	57 °C	58,895,568	58,895,670	(CT)16	Forward	TTTCCTGCTCACCTTAAACC
								Reverse	AACCTATATTGAGACGAACCG
SSR_1-1	180-220	4	PAN1	57 °C	12,756,851	12,757,030	(TC)33	Forward	AAACTGACAGCTTAAGCATTC
								Reverse	TGGGCATGTACTCTATCACTA
SSR_5-5	270-290	4	PAN2	57 °C	82,565,436	82,565,719	(GA)18	Forward	AGAGGAAGGAAAGAGAGCTAT
								Reverse	CACGAGGGAGCCTTATTAATA
SSR_6-4	170-180	4	PAN3	57 °C	63,517,285	63,517,456	(CT)30	Forward	ACGAGACTTTACAGAGAACAA
								Reverse	AGATAGGGAAGAACACAACAC

Supplementary Table S2: Chloroplast sequencing primer list

Target region	Primer Name	Primer Sequence
trnK-matK-trnK	1-F1	AATGAGAAGATTGGTTACGGAGAAAA
	1-F2	TGGTTTGAATCATTAGCGGAAATAATC
	1-R	AAGCAGCATGTCGTATCAATAGAGAATT
rps16	2-F1	GAAAAGGGTGTAGACGAACGAAT
	2-F2	TCATAAAACCCCACTTTCCGA
	2-R1	TCGGAAAGTGGGGTTTTATGA
	2-R2	TCGTTTCTCGGAGGCAAGAAT
trns-trnG	3-F1	TTGTCTCTCTATTCATCTTTGGATTCC
	3-F2	CCTTATTTTGGCTTATGGAACCTTATG
	3-R1	TTGGAATAAAAGAGGCCCGG
	3-R2	CATTGAGAGGAAAAGGGTTAAATCC
ycf3	4-F	ACGGCTCAGCAGTCAAGTTC
	4-R1	CACAAGAAACGGAGAACATACGAAA
	4-R2	TTCGAAATTCATGAAAGGCCCC
accD-psal	5-F	CTCCATGCTTTCTCTCCTTTGAATC
	5-R	TGCCGGAAATACTAAGCCCA
clpP	6-F1	GGTCAATGATCCAATAACCACCC
	6-F2	TGAGATCTCGCTATCAACCTCTTG
	6-F3	CCGAGTAAAGATCTGCCCGA
	6-R1	TGCAATTTGTTCAACCAGACGT
	6-R2	AAACAGAAACAGACCAAAACTCATCTT
	6-R3	TGAAGTATCCAGGCTCCGTTTAGA
rpl32	7-F1	AACTGGAAGTGGAATAAAAGGTATAATCC
	7-F2	AGGATATTGGGCAGCATTAAAAGC
	7-F3	AATGTTTTATGTTTTGTTTTGAACTGATC
	7-R1	GCTTTTAATGCTGCCCAATATCCT
	7-R2	TGATCAGTTCAAAACAAAACATAAAACA
	7-R3	GCGGCTTGTTCGAAATCATAATA

Supplementary Table S3: Nucleotide differences between Chloroplast polymorphic regions

	Nucleotide differences between Chloroplast polymorphic regions																					
Variety N.																						
1	AK_Royal_Automatic MK878538	**Group 1**																				
2	Yunma_7 MW013540 (CHIN)	42																				
3	sd130 MH118118 (CHIN)	51	17																			
4	Cheungsam KR184827 (KOR)	45	82	87	**Group 2**																	
5	FIN	46	82	89	23																	
6	Yoruba Nigeria NC_027223 (NGA)	39	75	80	16	9																
7	Canadian Hemp (CAD)	41	76	81	18	9	2															
8	FRA-1 + FRA-2	45	81	88	24	1	8	8														
9	POL	44	80	87	25	2	9	9	1													
10	HUN-1	44	80	87	23	2	9	9	1	2												
11	NED	46	82	89	24	2	9	9	1	2	2											
12	Dagestani KR779995 (RUS)	61	74	79	50	41	45	46	40	39	40	40	**Group 3**									
13	Chinese Hemp KY084475 (CHIN)	69	80	87	64	64	58	59	63	62	64	63	34									
14	Brazuka MK878537 (BRA)	65	74	81	64	67	61	62	66	65	66	66	32	18								
15	Carmagnola NC_026562 (ITA)	59	72	79	60	51	57	58	50	49	50	50	14	30	22							
16	CBD-1	68	79	84	62	66	61	61	65	64	65	65	35	26	17	27						
17	PO-CBD + TN-P11 + TN-P17 + WS (CAD)	66	77	84	62	64	59	59	63	62	63	63	31	20	11	23	6					
18	THC CAD-1 (CAD)	64	75	82	60	64	57	59	63	62	63	63	30	21	12	22	6	2				
19	THC-CAD-2 + LCH + SS-3 + SS-8 (CAD)	65	76	83	61	63	58	58	62	61	62	62	30	21	12	22	5	1	1			
20	ITA-3 + FRA-3 + HUN-2	66	77	82	60	64	57	57	63	62	63	63	29	22	13	23	6	2	2	1		
21	ITA-1 + ITA-2	67	78	83	61	65	58	58	64	63	64	64	30	23	14	24	7	3	3	2	1	
	Variety N.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Supplementary Table S4: Developed Primers for qRT-PCR

Developed Primers for qRT-PCR		
Target genes	**Primer Name**	**Primer Sequence**
Housekeeping genes	Actin_F	TTGCTGGTCGTGATCTTACTGAT
	Actin_R	TGTCTCCATCTCCTGCTCAAA
	EF2_F	TAGGAACTGTGATCCTAATGGTCC
	EF2_R	AGCGAAAACACGACCGAA
	F-BOX_F	GGCCGAGAGATTTGAGTGC
	F-BOX_R	ACATCATACAATAAGCCCTTCCCT
	HISTONE3_F	TGAAGAAGCCTCATCGGTTC
	HISTONE3_R	TCTTGAGCGATTTCCCTGAC
Cannabinoid pathway genes	CBCAS_F1	CTTGGTGGAGTGGATAGTCTAGTTG
	CBCAS_R1	AAGATGGTTGTATCAATCCAGCTC
	CBDAS_F1	GGTGGAGGAGGCTATGGAC
	CBDAS_R1	ACCACCACGTAAAGCCCA
	CsPT1_F	ACTTTGGGAAGGCATGTTG
	CsPT1_R	GAATGCCTTGAACATCAGAGAC
	CsPT4_F	GGGATAATTTGGCCTCAGG
	CsPT4_R	TGCTAGAGCAAGCTCACGAGT
	THCAS_F	AAACTCGTATACACTCAACACGACC
	THCAS_R	TAAAATAGTTGCTTGGATATGGGAGTT
Ethylene Related Genes (ERGs)	1_ACS2_FOC_F	ATCACCAATCCATCAAACCCT
	1_ACS2_FOC_R	AATGCTAACAAAACTCGGCTTG
	2_ETP2_FOC_F	TTGGGATGGTACTTGCATCAT
	2_ETP2_FOC_R	TCTTTCCAAACCGCAATCTTT
	3_FYF2_FOC_F	CATGCTCCCTTGAAGAACTACA
	3_FYF2_FOC_R	TGCATTTTCAGATGTTAGGGCT
	4_FYF4_FOC_F	TCCCATTTCAGATTTTCGACC
	4_FYF4_FOC_R	CTTTCTTCAACAACCCACTTCG
	5_MTN_FOC_F	ATGTAGTTTGGCCAGGAAAAGA
	5_MTN_FOC_R	GGCTTTAATGCTTGGACAGAAG
	6_ACO5_Uni_F	TTTACTCTCATGGATGGTGCTG
	6_ACO5_Uni_R	CCCAAGTTCTCATCCATCACTT
	7_FYF1_Uni_F	CTATGGAGCACAGGAAGGAAAA
	7_FYF1_Uni_R	TCCCAAATGCTCTAAATCCTCC

Supplementary Figure S5: amplification plot of endogenous reference genes during female flower development. It can be noted that only the F-Box gene transcription levels showed high PCR efficacy (compared to Actin and HIS-3) along with consistency between samples during flower development (which was not observed for the EF-2 Gene).