Integrated Classification of Cannabis Strains Marketed in Canada for Medical Purposes Based on Genetic, Chemical Profiles, and Morphological Traits

by

Dan Jin

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biomedical Engineering
University of Alberta

Abstract

The vernacular naming convention commonly used by the cannabis community ("Sativa" and "Indica") is inadequate for identifying or selecting strains for clinical research and medicinal production because they are arbitrary and inconsistent. In addition, they mostly contain tetrahydrocannabinol (THC) dominant strains while cannabidiol (CBD) dominant and intermediate strains (THC ≈ CBD), which are getting increased attention due to CBD's use as a therapeutic, have not been systematically studied nor compared to THC strains. This study series proposed a new, reliable, and scientific classification system for medicinal purposes based on genetics, chemical fingerprinting, and morphological traits of 23 strains grown in a common garden experiment. Leveraging the recent release of the 10-chromosome cannabis genome map, this study sequenced the whole genome of 23 cannabis strains and identified 137,858 genome-wide SNPs that provided insight into the distribution of genetic diversity and population structure in modern cannabis sold in Canada. This study identified 344 multiallelic SNPs that were able to separate CBD dominant, intermediate, and THC dominant strains using discriminant analysis of principal components (DAPC). Using canonical correlation analysis, this study tested the goodness of fit between this genotypic clustering (aligned with chemotypes) and the chemotypic variation by quantifying secondary metabolites in various plant parts of the same set of strains. Canonical correlation analysis assigned individual plants into their chemotypes with 100% accuracy. Other than THC and CBD, minor cannabinoids, terpenoids, and flavonoids showed differentiation power between CBD dominant, intermediate, and THC dominant chemotypes. In phenotyping, this study tested the goodness of fit between the genotypic clustering (aligned with chemotypes) and the morphological variation using 30 traits measured during the vegetative stage, at the end of flowering, and on harvested flowers of the same set of strains. Canonical correlation analysis

assigned individual plants to their preassigned genotypes with 92.9% accuracy. Both qualitative and quantitative traits showed differentiation power between CBD dominant, intermediate, and THC dominant chemotypes. In summary, this integrated investigation of Canadian cannabis strains showed that CBD dominant, intermediate, THC dominant strains can be separated at whole genome level and that the separation is further supported by chemotypic and phenotypic variation. This study series developed a set of classification rules for sorting strains into groups using identified traits or markers, individually or in tandem, that will facilitate strain identification and selection for research and clinical studies.

Preface

Each chapter (except **Chapter 1**) in this thesis is an independent manuscript which has been published (**Chapter 2, Chapters 3, Chapter 4,** and **Chapter 5**) in peer-reviewed journals.

Chapter 2 of this thesis has been published in the journal *Scientific Reports*: Jin, D., Dai, K., Xie, Z. & Chen, J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Scientific Reports* **10**, 3309 (2020). Dan Jin was responsible for project concept, experimental design, experimentation, data collection & analysis, and manuscript writing. Kaiping Dai and Dr. Zhen Xie assisted with method development and validation. Dr. Jie Chen was the supervisory author and monitored the research progress, provided suggestions on the structure of the manuscript, and finalized the manuscript.

Chapter 3 of this thesis has been published in the journal *PLOS ONE*: Jin, D., Henry, P., Shan, J. & Chen, J. Classification of Cannabis strains in the Canadian market with Discriminant Analysis of Principal Components Using Genome-wide Single Nucleotide Polymorphisms. *PLOS ONE* **16**, e0253387 (**2021**). Dan Jin conceived the project, designed the experiments, preformed the experiments, collected & analyzed the data, and wrote the manuscript. Dr. Philippe Henry contacted a licensed producer who provided cannabis strains and a commercial greenhouse for this project. He also provided vital direction for analyzing the data, assisted with data analysis, and software programming. Dr. Jacqueline Shan provided funding, provided suggestions, and proofread the manuscript. Dr. Jie Chen was the supervisory author and monitored the research progress, provided suggestions, and finalized the manuscript.

Chapter 4 of this thesis has been published in the journal *Frontiers in Plant Science*: Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Chemotypic Markers in Three Chemotype Categories of Cannabis Using Secondary Metabolites Profiled in Inflorescences, Leaves, Stem Bark, and Roots. *Front. Plant Sci.* **12**, (**2021**). Dan Jin conceived the project, designed the experiments, preformed the experiments, collected & analyzed the data, and wrote the manuscript. Dr. Philippe Henry contacted the licensed cultivator for this project and proofread the manuscript. Dr. Jacqueline Shan provided funding, provided suggestions, and proofread the manuscript. Dr. Jie Chen was the supervisory author and monitored the research progress, provided suggestions, and finalized the manuscript.

Chapter 5 of this thesis has been published in the journal *HortScience*: Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Phenotypic Characteristics in Three Chemotype Categories in the Genus Cannabis. *HortScience* **1**, 1–10 (2021). Dan Jin conceived the project, designed the experiments, performed the experiments, collected and analyzed the data, and wrote the manuscript. Dr. Philippe Henry found the resources of cannabis cultivars and a commercial greenhouse for this project and proofread the manuscript. Dr. Jacqueline Shan provided funding, provided suggestions, and proofread the manuscript. Dr. Jie Chen was the supervisory author and monitored the research progress, provided suggestions, and finalized the manuscript.

The papers that have been published are the following:

1. (**Published**) Jin, D., Jin, S., Yu, Y., Lee, C. & Chen, J. Classification of Cannabis Cultivars Marketed in Canada for Medical Purposes by Quantification of Cannabinoids and Terpenes Using HPLC-DAD and GC-MS. *J Anal Bioanal Tech* **8**, 2 (**2017**).

2. (**Published**) Jin, D., Jin, S. & Chen, J. Cannabis Indoor Growing Conditions, Management Practices, and Post-Harvest Treatment: A Review. *American Journal of Plant Sciences* **10**, 925–946 (**2019**).

3. (**Published**) Jin, D., Dai, K., Xie, Z. & Chen, J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Scientific Reports* **10**, 3309 (**2020**).

4. (**Published - Coauthor**) Henry, P. *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. *Journal of Cannabis Research* **2**, 26 (**2020**).

5. (**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Classification of Cannabis strains in the Canadian market with Discriminant Analysis of Principal Components Using Genome-wide Single Nucleotide Polymorphisms. *PLOS ONE* **16**, e0253387 (**2021**).

6. (**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Chemotypic Markers in Three Chemotype Categories of Cannabis Using Secondary Metabolites Profiled in Inflorescences, Leaves, Stem Bark, and Roots. *Front. Plant Sci.* **12**, (**2021**).

7. (**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Phenotypic Characteristics in Three Chemotype Categories in the genus Cannabis. *HortScience* **1**, 1–10 (**2021**).

Acknowledgements

I am sincerely grateful to my advisor Dr. Jie Chen for his unconditional support and constant encouragement through my studies every step of the way. I am sincerely grateful to Dr. Robert Burrell, Dr. Raimar Löbenberg, Dr. Liang Li, and Dr. Zhehui Jin for chairing my committee, providing me guidance, and bringing valuable suggestions for my project through meetings and emails. I am grateful to Dr. Youbin Zheng from the University of Guelph for being my external examiner of my final defense. I am grateful to Dr. Thomas Graham from the University of Guelph for reviewing my thesis proposal and academic papers, and providing horticultural advice for the design of the common garden experiment. I am grateful to Dr. Xiujie Susie Li from InnoTech Alberta, and Dr. David Hawley for reviewing my thesis proposal and providing valuable suggestions. I am also grateful to Dr. Tiequan Zhang from Agriculture and Agri-Food Canada, who reviewed and provided insightful comments on the papers I published.

I am sincerely grateful to Dr. Philippe Henry, adjunct professor at the University of Northern British Columbia. He is an excellent researcher in agricultural genomics and a great mentor that selflessly shared his knowledge. I am grateful to Barry Kuypers, who owns and operates the Emerald Flower Farm based in Kelowna, BC and who agreed to provide strains and growing facilities.

I would like to thank Dr. Jacquelin Shan from PBG BioPharma Inc. (Leduc, AB, Canada) who helped me throughout the project and provided financial support for sample collection, genome sequencing, travel expenses, and data analysis. I am also grateful to Labs-Mart Inc. for providing funding, facilities, and instrumentation support for chemical testing.

I am grateful to my parents-in-law, Dr. Changdong Jin and Chun Piao for their help and support in many ways throughout my project. I would like to thank my parents, Haiyuan Jin and Meishan Yin, who have always been supportive and optimistic as the source of energy and the light of my life. Thanks to my husband, Shengxi Jin, who have been the best partner who helped and supported me in my studies. Special thanks to my son, Dash, who brought laughter and courage to my life. We have been great partners.

Table of Contents

List of Tables

List of Figures

Chapter 1 – General Introduction

1.1 The history of cannabis as a medicine

Originating from Central Asia, cannabis was cultivated in China since 4000 BC for its stem fiber and seed[1]. Sometime between the third and second millennia BC, Emperor Shen-Nung first described the medical properties of cannabis in the world's oldest pharmacopoeia, The Classic of Herbal Medicine (*Shen-nung Pen-ts'ao Ching*), which indicated cannabis for rheumatic pain, intestinal constipation, disorders of the female reproductive system, malaria, and others[2]. The *Pen-ts'ao Ching* also contains the first record of cannabis's psychoactive effects: "if taken in excess will produce visions of devils. …Over a long term, it makes one communicate with spirits and lightens one's body…". However, there are limited records regarding the use of cannabis as a hallucinogen in ancient China[3]. The recreational use of cannabis was of more importance in India, possibly due to its direct association with religion. The psychoactive effects of cannabis were broadly known in India as a result of its preparation – breaking the glands (trichomes), which contain considerable amounts of active cannabinoids, and ensuring the presence of active cannabinoids in the finished product[4]. In the same period as recreational use, medical use of cannabis in India likely started around 1000 BC[5] with a wide range of applications, including as an analgesic, anticonvulsant, hypnotic, tranquilizer, anesthetic, anti-inflammatory, antibiotic, anti-parasitic, antispasmodic, digestive, appetite stimulant, diuretic, aphrodisiac or anaphrodisiac, antitussive, and expectorant[1]. There is also evidence suggesting cannabis was used medically in Assyria, Persia, and Scythia[1]. The medical use of cannabis spread from India to the Middle East around the 11th century and to Africa around the 15th century[6]. Cannabis was introduced to South America in the 16th century[7]. During this period, cannabis was cultivated in Europe for fiber, but not for medicinal uses[8]. Cannabis was introduced into western medicine in 1839 by Dr. William Brooke O'Shaughnessy, an Irish physician who worked in India and therein became familiar with cannabis. His publication "*On the preparations of the Indian hemp, or gunjah*" comprehensively described cannabis use: "*The narcotic effects of Hemp are popularly known in the south of Africa, South America, Turkey, Egypt, Middle East Asia, India, and the adjacent territories of the Malays, Burmese, and Siamese. In all these countries, Hemp is used in various forms, by the dissipated and depraved, as the ready agent of a pleasing intoxication. In the popular medicine of these nations, we find it extensively employed for a multitude of affections. But in Western Europe, its use either*

as a stimulant or as a remedy is equally unknown"[6]. He described many popular preparations for cannabis, evaluated its toxicity in animals, and tested its effects on patients. O'Shaughnessy's contribution greatly impacted on Western medicine and, around this period, hundreds of scientific articles were published in Europe and United States about the medicinal usage of cannabis. An upsurge in interest in cannabis research occurred in the late 19th and early 20th centuries. Cannabis was listed in the United States Pharmacopeia as a legitimate medical compound in 1851[9]. Main medical indications were summarized in *Sajous's Analytic Cyclopedia of Practical Medicine* in 1924[8].

However, the cannabis boom cooled significantly in the beginning of the 20th century. The most direct cause was that the clinically important effects could not be replicated reliably due to the inconsistency of source material – different strains, parts, ages, regions, and/or countries of original introduce variability in effects and efficacy[1]. As a result of a campaign by the Federal Bureau of Narcotics, legal restrictions limited the use of cannabis in the United States through an exorbitant tax and severe enforcement thereof. Cannabis was finally removed from the United States Pharmacopoeia in 1941[5].

Despite the restricted medical use of cannabis, recreational consumption spread in Europe, Brazil, Mexico, and other western countries during the second half of the 20th century. The percentage of young adults in the United States who had consumed cannabis at least once grew rapidly from 5% to 64% between 1971 and 1982[10]. Increased consumption drew scientific attention to cannabis once again, and research began to peak. Using more advanced techniques and sophisticated equipment, several active principal components of cannabis were isolated and studied. This began with the isolation and structural identification of the psychoactive compound Δ^9-THC ((-) trans-delta-9-tetrahydrocannabinol) in 1964[11]. Later in the 1990s, the endocannabinoid system within the nervous system was discovered, including its constituent cannabinoid receptors (CB_1 and CB_2), endogenous cannabinoids such as anandamide (AEA) and 2-AG (2-arachidonoylglycerol), and associated enzymes[12]. The endocannabinoid system is one of the most important physiologic systems involved in establishing and maintaining human health, including the direct regulation of appetite, pain, inflammation, thermoregulation, muscle control, motivation, mood, memory, etc. As such, there has been a resurgence in studying the therapeutic effects of cannabis, especially isolated cannabinoids. In the 21st century, medical cannabis has become a burgeoning industry

worldwide. Despite this, since 1970, the United States Controlled Substances Act has listed cannabis under Schedule I, reserved for drugs which have "no currently accepted medical treatment use".

The legal restriction on cannabis is relaxing worldwide. As of June 2021, forty-one countries have legalized the medical use of cannabis[13]. In the United States, 36 states, 4 territories, and the District of Columbia have legalized the medical use, but at the federal level, its use remains prohibited for any purpose[14]. In 2001, Medical Marijuana Access Regulations (MMAR) were established in Canada, granting licenses to people who are terminally ill, with severe spinal cord injury, arthritis, or multiple sclerosis to produce and possess cannabis for medical reasons[15]. The MMAR was replaced by Marijuana for Medical Purposes Regulations (MMPR) in 2013[15] and was further replaced by the Access to Cannabis for Medical Purposes Regulation (ACMPR) in 2016[16]. On October 17, 2018, the *Cannabis Act* came into force and removed cannabis from Controlled Drugs and Substance Act[17]. The *Cannabis Act* regulations apply to industrial hemp (THC < 0.3%) and cannabis (THC > 0.3%).

Two categories of cannabinoids-based products for medical purposes are available in Canada. One is phytocannabinoid-dense botanicals (dried marihuana), which can be purchased from Health Canada authorized licensed producers[18]. The second category is prescription drug in pill forms and are approved by the *Food and Drug Regulations* (FDR) in Canada[19]. These include Marinol® (dronabinol, which is synthetic Δ^9-THC), Cesamet® (nabilone, a THC-derivative), and Sativex® (nabiximols, wherein THC and CBD are in a 1:1 ratio in a liquid form of extracts used as an oromucosal spray). Marinol® is indicated for the treatment of nausea or vomiting associated with chemotherapy, and anorexia associated with AIDS-related weight loss[20]. Cesamet® is indicated for the treatment of nausea or vomiting associated with chemotherapy when symptoms are unresponsive to conventional therapy[21]. Sativex® is indicated as adjunctive treatment for spasticity and neuropathic pain associated with multiple sclerosis and intractable cancer pain[22–24].

1.2 A short review of cannabis taxonomy and classification system

The word cannabis derives from the Greek κάνναβις (kánnabis)[25]. The current botanical classification of cannabis[26,27] is listed in **Table 1.1**.

Table 1.1 Current botanical classification

Item	Name
Division	Angiosperms
Class	Dicotyledon
Subclass	Archichlamydeae
Order	Urticales
Family	Cannabinaceae
Genus	*Cannabis*
Species	*C. sativa* L.

The classification of species in the genus *Cannabis* had been a heated debate in the 1970s. Although many putative cannabis species were proposed for morphologically distinguishing cannabis, only three are widely accepted. These are *C. sativa, C. indica* and *C. ruderalis*. The genus *Cannabis* was first proposed by Linnaeus in 1753, who considered the genus to be monotypic and with a single species, *Cannabis sativa* L., which has loose inflorescences covered with sparse trichomes and resembles a northern European fiber-type landrace (domesticated and locally adapted)[28]. Later in 1785, de Lamarck described a second species or subspecies, *Cannabis indica* Lam., which was collected in India and had dense trichomes, narrower leaflets, branchier growth, poorer fibre quality but stronger psychoactive effects[29]. In 1924, either a third variety of *C. sativa* or a separate species was recognized in central Russia, named *Cannabis ruderalis* Janisch. after Russian botanist Janichevsky[30]. Schultes et al. (1974) supported the polytypic concept of the genus *Cannabis* and insisted that *Cannabis* should be divided into three species[31]: *C. sativa* L.*, C. indica* Lam.*,* and *C. ruderalis* Janisch., with the first two considered as domesticated phases and the latter a wild phase. He pointed out that the leaflet morphologies have close correlation with the three distinctive growth forms of the species, which was later further supported by Anderson (1980), shown in **Figure 1.1**[32]:

1) *Cannabis sativa*: Plants relatively tall, 5-18 feet tall or more, laxly branched; leaflet narrowly lanceolate.
2) *Cannabis indica*: Plants short, 2-4 feet tall, pyramidal, compactly branched; leaflet wide, oblanceolate.
3) *Cannabis ruderalis*: Plants (female) very short, 0.5-2 feet tall, usually unbranched; leaflet elliptic.

These modified concepts of *sativa* and *indica* by Schultes and Anderson deviated from the original botanical nomenclature by Linnaeus and de Lamarck, but these concepts and the illustration have been widely used by cannabis growers and breeders since the 1980s.

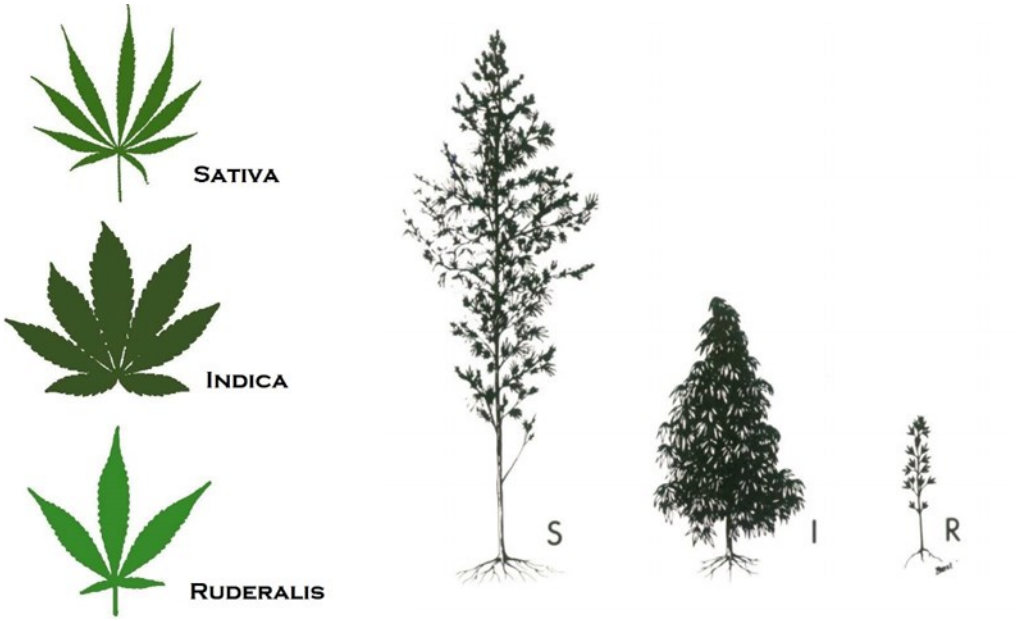


Figure 1.1 Cannabis taxonomy adapted from Anderson (1980)
(S=Sativa, I=Indica, R=Ruderalis)

On the other hand, Small and Cronquist (1976) considered the *Cannabis* genus to be monotypic. They studied two groups of cannabis from different origins, each with differing intoxicant potential and usage, and treated them as subspecies: *sativa* and *indica*, of the only species, *C. sativa* L., in the genus. Each subspecies would consist of one domesticated phase and one wild phase[33]. This conclusion, that *Cannabis* has only one highly variable species, is due to the lack of sterility barriers[34], based on the biological definition that one species is "*a group of living organisms consisting of similar individuals capable of exchanging genes or interbreeding*"[35]. However, there are examples in other genera showing that interspecific hybridization is possible, which implies it may be inaccurate to use breeding behavior to define species[36].

Hillig (2005) conducted a series of investigations on 157 cannabis accessions in an attempt to systematically classify them using genetic, morphological, and chemotaxonomic characteristics[27].

The results supported a polytypic concept of *Cannabis*. Genetic markers (allozymes) were visualized by starch gel electrophoresis and the divergence in allele frequencies revealed two major gene pools. One was a *sativa* gene pool, including fibre/seed landraces from Europe, Asia Minor, and Central Asia. The other was an *indica* gene pool, including fibre/seed landraces from eastern Asia, narrow-leafleted drug (NLD) landraces from southern Asia, Africa, and Latin America, wide-leafleted drug (WLD) landraces from Afghanistan and Pakistan, and feral populations from India and Nepal (**Figure 1.2, Table 1.2**). A third putative gene pool includes ruderal populations from Central Asia combined with a *C. sativa* feral biotype.

Figure 1.2 Map showing the countries of origin of accessions assigned to the *C. indica* and *C. sativa* gene pools.
The arrows suggest human-vectored dispersal from the presumed origin of Cannabis in Central Asia[27].

Table 1.2 Taxonomic circumscription of the *Cannabis* germplasm collection.

Putative Taxon	Description	Quantitative Differences (THC%, CBD%)*
C. *indica* hemp biotype	Hemp landraces from southern and eastern Asia	3rd THC level 5th CBD level 5th THC+CBD level
C. *indica* feral biotype	Feral populations from India and Nepal	4th THC level 4th CBD level 4th THC+CBD level
C. *indica* NLD biotype	Narrow-leaflet drug (NLD) strains from the Indian subcontinent and other drug producing regions	2nd THC level 7th CBD level 2nd THC+CBD level
C. *indica* WLD biotype	Wide-leaflet drug (WLD) strains from Afghanistan and Pakistan	1st THC level 6th CBD level 1st THC+CBD level
C. *sativa* hemp biotype	Hemp landraces from Europe, Asia Minor and central Asia	5th THC level 1st CBD level 3rd THC+CBD level

C. *sativa* feral biotype	Feral populations from eastern Europe	6th THC level 2nd CBD level 6th THC+CBD level
C. *ruderalis*	Ruderal populations from central Asia	7th THC level 3rd CBD level 7th THC+CBD level

*Taxonomic circumscription of the *Cannabis* germplasm collection based on allozyme allele frequencies and chemotaxonomic analysis of cannabinoid variation in Cannabis[27]. 1st means highest average content and 7th means lowest average content of THC or CBD between these seven biotypes.

1.3 Cannabis classification for forensic, industrial, and medical purposes

For forensic and industrial applications, cannabis plants can be split into two varieties based on the concentration of THC and CBD. The first variety is generally referred to as marijuana, which is the most widely used illicit drug worldwide and is cultivated to maximize psychoactive THC (5-20%). The second variety is hemp, which is a legal commercial crop containing low amounts of THC (<0.3%). Since the late 20th century, molecular analytical techniques such as Random Amplified Polymorphic DNA (RAPD)[37–44], Amplified Fragment Length Polymorphism (AFLP)[45–48], Simple Sequence Repeats (SSR)[49–54], and Inter-Simple Sequence Repeat (ISSR)[55–58] have been applied for distinguishing drugs from hemp, grouping varieties in accordance with geographic origin, or achieving early recognition of plant sex in hemp genetic improvement programs in agriculture (**Table 1.3**).

Table 1.3 Molecular analytical techniques applied in cannabis classification

Method	Advantages	Disadvantages	Applications	Ref
RAPD	Cheap and simple, need no prior information on the genome, random primers	Reproducibility problems (PCR thermal cycler ramp)	Hemp or drug accessions were grouped in accordance with their countries/regions/sources; determine sex specific markers	37–44
AFLP	Higher reproducibility than RFLP, need no prior information on the genome, random primers	Need to design primers, more complex and costly	Differentiate hemp from drug strains for forensic utilizations; hemp varieties grouping; determine sex specific markers	45–48
SSR	High reliability and repeatability, high polymorphism	Higher cost and longer time (primer design requires SSR flanking region sequencing)	Differentiate hemp from drug strains for forensic utilizations; Hemp varieties grouping in accordance with their geographic origin	49–54
ISSR	Higher reproducibility than RAPD, lower cost than AFLP		Differentiate cannabis from different origination; Discriminate hemp from	55–58

			drug strains for forensic utilizations	

For industrial and medical purposes, five chemotypes based on the content of major cannabinoids have been proposed (**Table 1.4**). Chemotype I (Drug type: high THC to CBD ratio) contains more than 0.3% THC and less than 0.5% CBD, Chemotype II (Intermediate type: THC to CBD ratio between 0.5 and 2) contains more than 0.3% THC and more than 0.5% CBD, and Chemotype III (Fiber type: low THC to CBD ratio) exhibits less than 0.3% THC[59,60]. Chemotype IV contains cannabigerol (CBG) (>0.30%) as the major cannabinoid and CBD as the single significant complementary cannabinoid[61]. Chemotype V has undetectable amounts of cannabinoids[62]. Chemotypes IV and V, with mainly CBG or no cannabinoids, have potential for textile and pharmaceutical use. An inheritance model of five chemotypes by cross inbreeding and RAPD analysis had been proposed, where the molecular mechanism of THC, CBD, cannabichromene (CBC), and CBG production was explored [42,63–66].

Table 1.4 Main cannabinoid characteristics for five chemotypes

Chemotype	Content	Quantitative criteria	Qualitative criteria
I	Prevalent THC	THC>0.3% d.w.* CBD<0.5% d.w.	THC > CBD
II	Intermediate	THC>0.3% d.w. CBD>0.5% d.w.	THC ≈ CBD
III	Prevalent CBD	THC<0.3% d.w. CBD>0.5% d.w.	THC < CBD
IV	Prevalent CBG	CBG>0.3% d.w. CBD<0.5% d.w.	CBG only
V	Zero cannabinoids	total cannabinoids content<0.2% d.w.	0 cannabinoids

*d.w. = dry weight

Since the 1970s, breeding for high THC content for recreational purposes has occurred very aggressively in North America. Nearly all drug-type cannabis currently cultivated in the USA, Canada, and Europe are hybridized, resulting in thousands of strains[67]. Strain names are arbitrary, and some strains may simply be clones of other strains[68]. Cannabis breeders and users have adopted a vernacular classification of "Sativa" and "Indica" to describe plants' pedigrees as a percentage of *sativa* and *indica*[67]. Hybrid strains are described as "Sativa-dominant", "Indica-dominant", or as a percentage between the two. This vernacular separation may also be arbitrary and inaccurate due to untracked selection and breeding[69].

Using chemical[70–72] and genetic[43,73–77] tools, researchers have tried to discriminate modern sinsemilla (seedless) strains labeled as "Sativa" (or NLD) and "Indica" (or WLD) and in medicinal applications. However, the category of "Sativa" and "Indica" may be flawed and not adequate for medical purposes. One study found that the genetic structure for 81 drug stains is only moderately correlated with reported ancestry percentage of "Sativa" vs. "Indica"[75]. In addition, the genetic, chemical, and the morphological traits of the CBD dominant strains and intermediate strains (or balanced strains with THC and CBD both as dominant strains), which have been getting increased attention due to CBD's use as a therapeutic, have not been studied or compared with THC strains in the current literature.

The current body of knowledge for cannabis classification contains other shortcomings. Firstly, samples in most classification studies were collected from disparate sources[70,72] and are subject to inconsistent environmental factors, which may affect classification results. For example, cannabinoid and terpenoid content, among other metabolites, change during the growth and flowering stages[78], and cannabinoid acids decarboxylate into their neutral forms via heat or light exposure[79]. In addition, researchers have suggested exploring a wider variety of pharmacologically-interesting compounds other than THC and CBD[80–82]. Recently, a metabolic approach was proposed to profile strains using cannabinoids and terpenoids[70,71,83,84]. Flavonoids, in inflorescences and leaves[85], and triterpenoids & sterols, in stem bark and roots[80], are all therapeutically interesting. Classification information for the types and contents of these metabolites will facilitate further clinical research. Secondly, discrimination studies using genetic tools mostly utilized partial genome information with few or no overlap sequences between datasets[86]. Whole genome sequencing is recommended despite its higher cost because it enables comparison of datasets from different sources[86]. It also provides comprehensive genetic information[86] – studies have shown that the differences between fibre- and drug-type cannabis are at a genome-wide level and not limited to genes involved in THC production[75]. Finally, although morphological characteristics are essential for botanical differentiation, restricted access to live plants limits the morphological study of modern cannabis. Currently, collected samples are visually described by breeders and growers as NLD and WLD based on their leaf shapes, with no quantitative data provided[27]. These descriptors may be unsuitable to describe leaf shapes of hybrid strains beyond a "Sativa" or "Indica" classification. One solution is to first develop categories for modern cannabis strains using reliable classification criteria, such as genomic variation. Then,

quantitative data on plants' representative traits (including plant growth rate, flower length/branch length, leaf length/width ratio, etc.[27]) within each category can be calculated and summarized for determining discrimination criteria.

1.4 Objectives, hypotheses, and methods

Cannabis classification is a fundamental requirement for future medical research and applications because it provides an overview of types and contents of therapeutic secondary metabolites in each part of the cannabis plant. The objective of this project is to develop a novel, integrated, and systematic classification model for modern cannabis strains (including CBD dominant and intermediate strains) using genetic variation at whole genome level, and then to integrate the model with morphological and chemical characteristics. The hypothesis is that chemotypic and phenotypic treatment will align with the pattern of genetic variation, given that genetic differences contribute most to these variations when environmental factors are controlled. This system creates genotypes, chemotypes and phenotypes, by identifying respective markers that can be used individually or in tandem for strain selection and screening of germplasm, live plants, or marketed products.

The common-garden experiment will address the research shortcomings previously identified. In a common-garden experiment, plants will grown in a single location, under identical environmental conditions, and uniformly processed[69], thus enabling more accurate classification results. In our project, we collaborated with a licensed grower to grow between 20 to 30 strains in a common-garden experiment. The whole genome of each strain was sequenced. Morphological characteristics were recorded at the end of vegetative growth. Secondary metabolites (cannabinoids, terpenoids, flavonoids, triterpenoids, and sterols) were profiled in inflorescences, leaves, stem bark, and roots when plants were harvested. Morphological and chemical data for each strain were averaged from four to six clones of that strain.

The metabolites of interest included 14 cannabinoids, 45 terpenoids (29 monoterpenoids and 16 sesquiterpenoids), 7 flavonoids, 3 sterols, and 3 triterpenoids. This multipart study included the development of quantitative methods using liquid chromatography coupled with mass spectroscopy (LC-MS) for cannabinoids, liquid chromatography coupled with a standard ultraviolet detector and mass spectroscopy (LC-UV-MS) for flavonoids, and gas chromatography coupled with mass spectroscopy (GC-MS) for terpenoids and sterols in **Chapter 2**. All methods

were validated for linearity, trueness, precision, repeatability, and robustness. The methods were employed for generating the chemical profiles of the inflorescences, leaves, stem bark, and roots of studied cannabis strains. Strains' genomes were sequenced and strains were first classified into genotypes by Discriminant Analysis of Principal Components (DAPC) in **Chapter 3**. The DAPC clustering method was previously implemented to identify clusters of cannabis strains based on genetic information for studying the separation of fiber vs. drug type cannabis[77] and terpenoid dominance in cannabis samples[87]. In this study, DAPC was used to investigate whether these strains can be differentiated into different groups at genome level. Canonical correlation analysis[88] was used to test the goodness of fit of a chemotypic treatment to the pattern of genetic variation in **Chapter 4**. Using groupings preassigned to genotypes, canonical correlation analysis was applied for investigating whether plants could be predicted into preassigned genotypes based on chemical profiles. Chemotype markers, as compounds with the highest discriminating ability, were identified. Similarly, canonical correlation analysis was used to test the goodness of fit of a phenotypic treatment to the pattern of genetic variation in **Chapter 5**. Using groupings preassigned to genotypes, canonical correlation analysis was applied for investigating whether plants could be predicted into preassigned genotypes based on morphological characteristics. Phenotype markers, as the morphological characteristics with the highest discriminating ability, were identified.

1.5 Significance of the research

The research aims to provide links between genetics, morphological properties, and chemical components, thus providing a foundation for classification. Chemotypic and phenotypic markers are identified for this classification system, facilitating the holistic identification of cannabis strains for medical purposes. Popularization of this system is expected to enhance and accelerate the adoption of cannabis as a standardizable medicine. By creating chemical profiles covering medical-relevant compounds for each plant part, this classification system also provides the novel opportunity to explore the untapped market of cannabis leaves, stems bark, and roots.

Chapter 2 – Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem bark, and Roots for Medicinal Purposes

2.1 Abstract

Cannabis research has historically focused on the most prevalent cannabinoids. However, extracts with a broad spectrum of secondary metabolites may have increased efficacy and decreased adverse effects compared to cannabinoids in isolation. Cannabis's complexity contributes to the length and breadth of its historical usage, including the individual application of the leaves, stem bark, and roots, for which modern research has not fully developed its therapeutic potential. This study is the first attempt to profile secondary metabolites groups in individual plant parts comprehensively. We profiled 14 cannabinoids, 48 terpenoids (29 monoterpenoids and 19 sesquiterpenoids), 7 flavonoids, 3 sterols, and 3 triterpenoids in cannabis flowers, leaves, stem bark, and roots in three chemovars available. Cannabis inflorescence was characterized by cannabinoids (15.77-20.37%), terpenoids (1.28-2.14%), and flavonoids (0.07-0.14%); the leaf by cannabinoids (1.10-2.10%), terpenoids (0.13-0.28%), and flavonoids (0.34-0.44%); stem bark by sterols (0.07-0.08%) and triterpenoids (0.05-0.15%); roots by sterols (0.06-0.09%) and triterpenoids (0.13-0.24%). This comprehensive profile of bioactive compounds can form a baseline of reference values useful for research and clinical studies to understand the "entourage effect" of cannabis as a whole, and also to rediscover therapeutic potential for each part of cannabis from their traditional use by applying modern scientific methodologies.

(Published) Jin, D., Dai, K., Xie, Z. & Chen, J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Scientific Reports* **10**, 3309 (**2020**).

2.2 Introduction

Cannabis is a complex herbal medicine containing several classes of secondary metabolites, including at least 104 cannabinoids, 113 terpenoids (including 61 monoterpenoids, 52 sesquiterpenoids), 26 flavonoids, 11 steroids, and 5 triterpenoids among 545 identified compounds[26,80,89–92]. The postulated biosynthetic pathways for these metabolite groups[85,93] are outlined in **Figure 2.1**. Cannabis has attracted a new wave of interest for its broad medicinal applications as 1) an analgesic, potentially as an adjunct to or substitute for opiates in the treatment of chronic pain[94], and 2) an appetite stimulant and digestive aid[1], among others. Since the 1960s, the research has focussed mainly on cannabinoids, Δ^9-tetrahydrocannabinol (Δ^9-THC), and cannabidiol (CBD) in particular[95–113]. The major psychoactive content expressed as total THC decreases in the order of inflorescences (10-12%), leaves (1-2%), stems (0.1-0.3%), roots (<0.03%), and seeds (generally absent)[79]. As such, female flower tops are harvested while other parts are often discarded by growers[79]. This is a potentially unnecessary waste. As an ancient medicine in various cultures, each part of the cannabis plant has been historically indicated with a wide range of applications relating mostly to painkilling, inflammation releasing, and mental illness treatment [114–117].

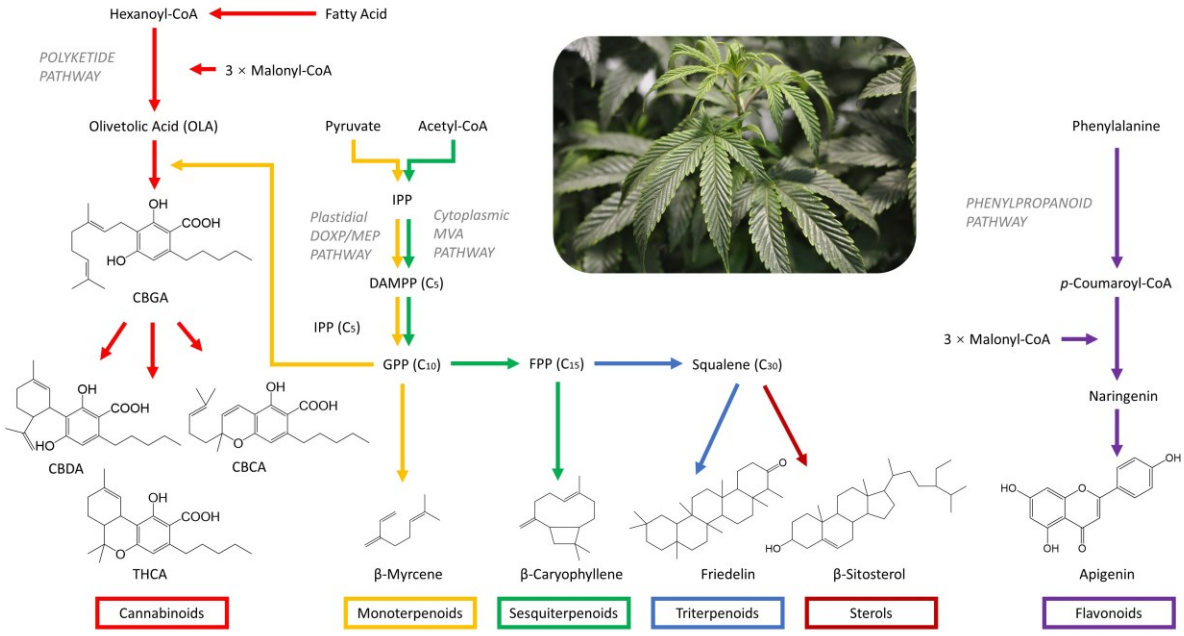


Figure 2.1 Biosynthesis pathways of cannabinoid, terpenoids, sterols, and flavonoids.
Cannabinoids and terpenoids are produced and stored in the secretory cells of glandular trichomes, which are found in the aerial parts of cannabis plants and are especially dense on the top surfaces of seedless female flowers[81]. Two precursors for cannabinoids are olivetolic acid (OLA), derived from the polyketide pathway, and geranyl diphosphate (GPP), derived from the plastidial deoxyxylulose phosphate/methyl-erythritol phosphate (DOXP/MEP pathway)[118–120]. Cannabigerolic acid (CBGA) is formed by the condensation of OLA and GPP and is further converted to cannabidiolic acid (CBDA), Δ^9-tetrahydrocannabinolic acid (Δ^9-THCA), and cannabichromenic acid (CBCA) by CBDA synthase[121], Δ^9-THCA synthase[122], and CBC synthase[123], respectively. If divarinic acid is condensed with GPP instead of OLA, the propyl (C_3 side-chain) instead of pentyl (C_5 side-chain) cannabigerovarinic acid (CBGVA) is produced, which can be further converted to cannabidivarinic acid (CBDVA), tetrahydrocannabivarinic acid (THCVA), and cannabichromevarinic acid (CBCVA) following similar pathways[124]. Terpenoids are derived from the mevalonate (MVA) pathway or from the DOXP/MEP pathway. Both pathways produce isopentenyl diphosphate (IPP), which is further isomerized to dimethylallyl diphosphate (DMAPP), at their endpoints[124]. The DOXP/MEP pathway provides GPP to form monoterpenoids (C_{10}) while MVA pathway provides farnesyl diphosphate (FPP) for sesquiterpenoids (C_{15}) and squalene as precursors for triterpenoids (C_{30}) and sterols[124]. Flavonoids in cannabis, mainly flavones (luteolin, apigenin, orientin, vitexin, and isovitexin) and flavonols (quercetin and kaempferol), exist as free aglycones or as conjugated *O*-glycosides or *C*-glycosides[85,125–127]. The phenylpropanoid pathway produces *p*-coumaroyl-CoA from phenylalanine. In conjunction with three molecules of malonyl-CoA, *p*-coumaroyl-CoA produces naringenin, which is the substrate for flavone and flavonol biosynthesis[93,124]

Compounds other than Δ^9-THC and CBD may contribute to the therapeutic effects of each plant part in their traditional uses. Minor cannabinoids, such as cannabinol (CBN), cannabigerol (CBG), cannabichromene (CBC), also have broad therapeutic potential[95,128–131]. Terpenoids may directly elicit physiological effects or modulate cannabinoid responses[81]. Flavonoids share a wide range of biological effects with cannabinoids and terpenoids that include anti-inflammatory, anti-cancer, and neuroprotective properties[132]. One of the triterpenoids identified in cannabis root, friedelin, contains anti-inflammatory, antioxidant, estrogenic, anti-cancer, and liver protectant properties[115]. Plant sterols may reduce plasma cholesterol levels[133–137]. The combination of different secondary metabolites of varying concentrations is believed to increase the range of therapeutic properties – known as the "entourage effect"[81,138,139]. One recent study showed that whole plant extracts are more beneficial than pure CBD for the treatment of inflammatory conditions in mice[140]. Another preclinical study has shown that a botanical cannabis preparation was more effective than pure THC in producing antitumor responses in vitro[138]. However, the increased potency was attributable to compounds other than the five most abundant terpenoids in the preparation[138]. The literature suggests that a wider range of bioactive compounds should be included when examining the beneficial medicinal properties of botanical cannabis preparations.

The aim of this study is to leverage a comprehensive investigation of chemical profiles in each plant part. The metabolites of the study included 14 cannabinoids, 48 terpenoids (29 monoterpenoids and 19 sesquiterpenoids), 7 flavonoids, 3 sterols, and 3 triterpenoids. This multipart study included the development of quantitative methods using liquid chromatography

coupled with mass spectroscopy (LC-MS) for cannabinoids, liquid chromatography coupled with a standard ultraviolet detector and mass spectroscopy (LC-UV-MS) for flavonoids, and gas chromatography coupled with mass spectroscopy (GC-MS) for terpenoids, sterols, and triterpenoids. Relevant compounds were selected based on their pharmacological activities,[80,93] or use in other cannabis classification studies[71,72,83,84]. The methods were then employed for generating the chemical profiles of the inflorescences, leaves, stem bark, and roots of three selected cannabis chemovars **(Figure 2.2**). The results can form a baseline of reference values useful for future research and clinical studies on these compounds' pharmacological activity.

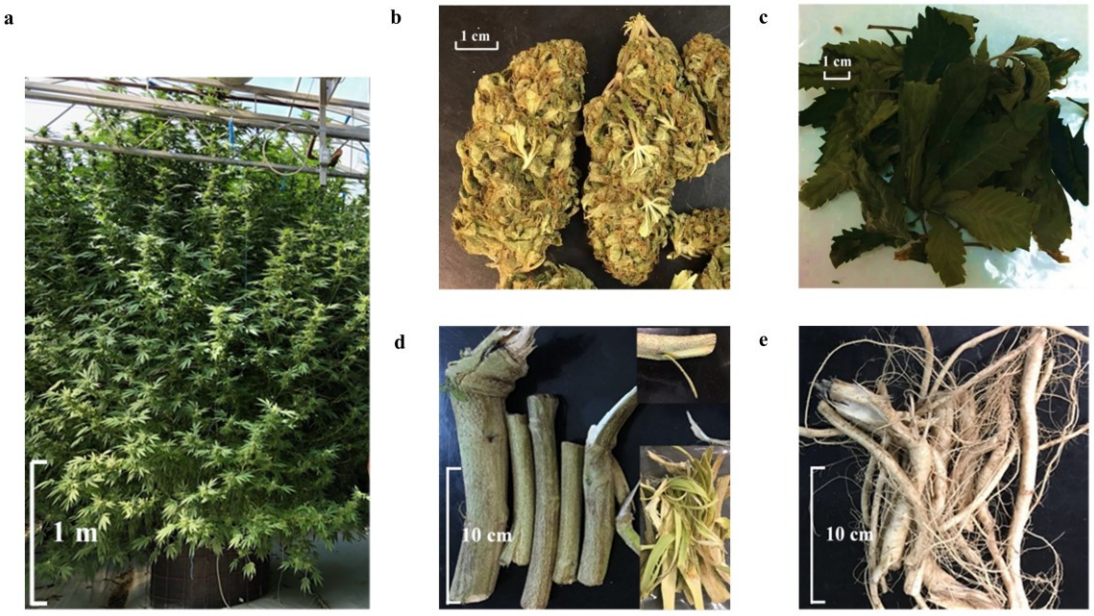


Figure 2.2 Cannabis CBD Mango Haze plant, inflorescences, leaves, root, stem bark, and roots.
(a) CBD Mango Haze plant that has been kept in vegetative stage for six months and initiated flowering for two months in a greenhouse. (b) Dried cannabis inflorescences. (c) Dried cannabis leaves. (d) Fresh cannabis stems with barks and later peeled (right corner). (e) Fresh root material.

2.3 Materials and methods

2.3.1 Solvents and chemicals

The 14 cannabinoid standards and Δ^9-THC-d_3, which was used as an internal standard (IS), were purchased from Sigma-Aldrich Company (Oakville, ON, Canada). All cannabinoid standards were analytical grade 1 mg/mL solution in methanol or acetonitrile. Standards for monoterpenoids (α-

pinene, camphene, β-pinene, myrcene, Δ^3-carene, α-terpinene, p-cymene, limonene, β-ocimene, γ-terpinene, terpinolene, linalool, 1,8-cineole (eucalyptol), (-)-isopulegol, geraniol) and sesquiterpenoids (β-caryophyllene, α-humulene, trans-nerolidol, (-)-guaiol, α-bisabolol, and caryophyllene oxide) were purchased from RESTEK (Bellefonte, PA, US). These 21 terpenoids were certified reference materials provided as mixed standards at approximately 2500 µg/mL in isopropanol. Standards for monoterpenoids (α-phellandrene, sabinene hydrate, camphor, fenchol, borneol, α-terpineol, sabinene, (+)-carvone, (+)-dihydrocarvone, pulegone, terpineol-4-ol, fenchone, and geranyl acetate), sesquiterpenoids (aromadendrene, (+)-cedrol, globulol, ledene, viridiflorol), triterpenoid (friedelin), sterol (stigmasterol), flavonoids (orientin, vitexin, isovitexin), tridecane (used as an IS) for quantification of mono- and sesquiterpenoids, and cholesterol (used as an IS) for quantification of triterpenoids and sterols were analytical standards purchased from Sigma-Aldrich Company (Oakville, Ontario, Canada). Sesquiterpenoids (trans-β-farnesene and valencene) were certified reference materials purchased from ChromaDex (Irvine, CA, US). Standards for nerol, sesquiterpenoids (β-eudesmol and β-elemene), triterpenenoids (β-amyrin and epifriedelanol), sterols (campesterol and β-sitosterol), and flavonoids (quercetin, luteolin, kaempferol, and apigenin) were certified reference materials purchased from Chengdu Push Bio-Technology Co., Ltd. (Chengdu, Sichuan, China). Methanol, ethanol, hexane, and hydrochloric acid were purchased from Fisher Scientific (Ottawa, ON, Canada). Ethyl acetate and formic acid were purchased from Caledon Laboratory Chemicals (Halton Hills, Ontario, Canada). C7-C40 saturated alkanes standard was purchased from Sigma-Aldrich Company (Oakville, Ontario, Canada). Water was produced in-house using a Millipore filtration system, which purified water to 18mΩ resistivity.

2.3.2 Sample collection and preparation

Dried cannabis inflorescence and fresh leaves, stems, and roots from three cannabis chemovars were provided by a licensed producer in Canada (**Figure 2.3**). The plants were kept vegetative for six months and flowered for two months in a greenhouse. Stems were taken near the root tissue. Stems from the upper parts of the plants were not available. Two THC dominant chemovars (Chemovar I - Grand Doggy Purps and Chemovars II - Granddaddy Purple) were alleged to be "Indica" varieties. The intermediate type chemovar (Chemovar III - CBD Mango Haze), characterized by having a total THC to total CBD ratio of 1:2, was purported to be a "Sativa" variety. Five to eight inflorescences (2g to 4g) of each chemovar were pulverized with a manual

grinder. Stem bark and roots, which were cut into 2 cm pieces, were air-dried together with leaves at room temperature for 24 hours. Dried fan leaf material was crushed using a mortar and pestle and sifted through a 1.18mm sieve. Dried stem bark and root samples were ground into a fine powder using an electric blender. All air-dried material was stored under refrigeration for two weeks until analysis.

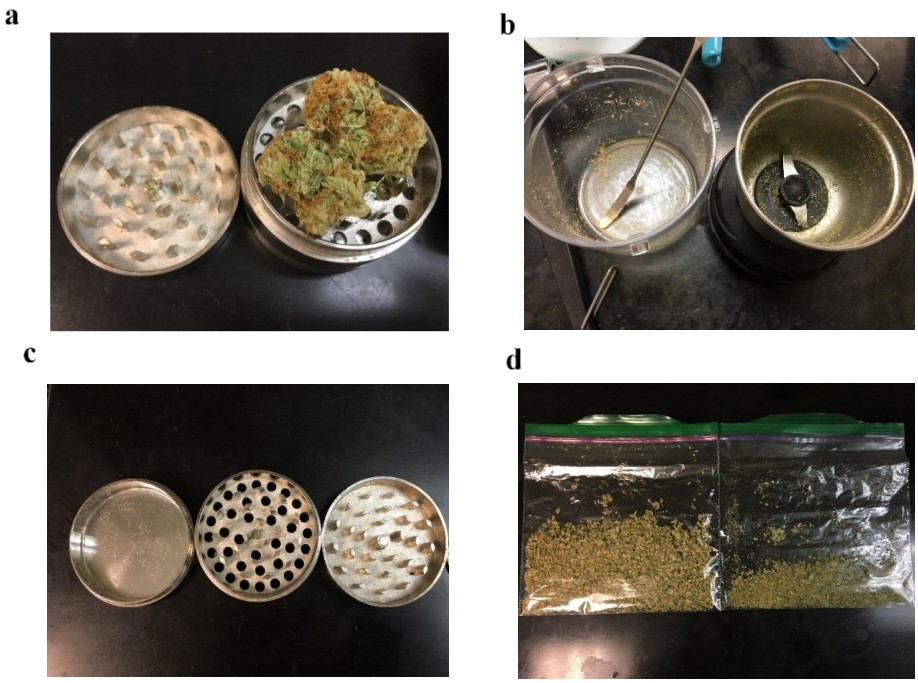


Figure 2.3 Pulverization of cannabis raw samples using manual grinder and electric blender. (a) manual grinder and (b) electric blender. (c) The inside of manual grinder is relatively clean without much resin left behind whereas there is a layer of resin stuck to the inside of the cup of the electric blender. (d) The resultant particle size is larger using the manual grinder (left) than the electric blender (right).

2.3.3 Methanol extraction for cannabinoids, monoterpenoids, and sesquiterpenoids

In brief, 400mg material of each plant part was extracted with 20.0mL methanol (with 100 µg/mL tridecane as IS for mono-and sesquiterpenoids) by sonication for 20 minutes at room temperature. The extract was then filtered through a 0.45µm membrane filter disk. An aliquot of the extract was used to quantify mono- and sesquiterpenoids using GC-MS. For cannabinoids, the prepared solutions were spiked with Δ^9-THC-d_3 (0.5 µg/mL) as IS prior to LC-MS analysis. Dilutions were applied as necessary.

2.3.4 Ethyl acetate extraction for triterpenoids and sterols

One gram of each plant part was extracted with 20.0mL ethyl acetate by sonication for one hour, followed by maceration for one day at room temperature. The extract was filtered through a 0.45μm membrane filter disk and spiked with cholesterol (50 μg/mL) as IS prior to GC-MS analysis.

2.3.5 Acid-hydrolyzation for flavonoids

The method for acid hydrolysis extraction of flavonoids was adapted from the monograph for ginkgo in the latest version of the United States Pharmacopoeia (USP)[141]. In brief, 250mg of the sample was extracted with 5mL extraction solvent (ethanol, water, and hydrochloric acid at a 50:20:8 volume ratio). The air in the tube was displaced with nitrogen. The solution was then vortexed for 10 seconds and sonicated for 10 minutes, followed by hydrolysis in a 100°C water bath for 135 minutes. The tube was left to cool to room temperature. Then the contents were transferred to a 50mL volumetric flask. The tube was then repeatedly rinsed with methanol, and the rinses were combined with the extract. The flask was filled to volume with methanol, then sonicated again for 5 minutes. The solution was filtered through a 0.45μm membrane filter disk, an aliquot of which was used for quantification.

2.3.6 LC-ESI-MS setup for cannabinoids assay

The LC-ESI-MS system used in this study was a modular Agilent 1260 Infinity II LC system comprised of the following components: a vacuum degasser, a quaternary pump (G7111B), an autosampler (G7129A), an integrated column compartment (G7130-60030), and a single quadrupole liquid chromatography/mass selective detector (LC/MSD 6125B) with electrospray ionization (ESI) (C1960-64217). The chromatographic separation of cannabinoids was performed on an Agilent Zorbax RX-C18 column (4.6 mm × 150 mm, 3.5 μm). The mobile phase was composed of 0.2% aqueous formic acid (A) and methanol (B). Gradient elution was as follows: 75%-90% B in 0-13 minutes and 90% B in 13-26 minutes. The post-run time was 4 minutes. The flow rate was 0.6 mL/min. The column temperature was set at 30°C. The injection volume was 5μL. The ESI-MS system was operated in positive ionization mode. Mass to charge ratios (M/z) of fragment ions for each compound were listed in **Table 2.1**. The instrument settings were set as follows: the capillary voltage was 3kV, the nebulizer (N_2) pressure was 50 psi, the drying gas temperature was 350°C, the drying gas flow was 12 L/min, and the fragmentor voltage was 70 V.

Table 2.1 SIM method parameters for cannabinoids

Name	Retention Time (min)	POS/NEG	Quantifier
IS: Δ^9-THC-d_3	4.652	POS	287.2
1. CBDV	8.102	POS	287.2
2. CBDVA	9.256	POS	331.2
3. CBG	11.263	POS	317.3
4. CBD	11.410	POS	315.3
5. CBDA	12.181	POS	359.2
6. THCV	12.393	POS	287.2
7. CBGA	13.539	POS	343.3
8. CBN	15.048	POS	311.2
9. Δ^9-THC	16.471	POS	315.3
10. Δ^8-THC	17.156	POS	315.3
11. THCVA	17.282	POS	331.2
12. CBC	18.398	POS	315.2
13.THCA	22.350	POS	359.3
14. CBCA	23.722	POS	359.3

2.3.7 HPLC-UV-MS setup for flavonoid identification and quantification

The HPLC-UV-MS system used in this study was the same LC-MS system described above, with an Agilent 1260 variable wavelength detector (G7114A) in series. The chromatographic separation of flavonoids was performed on Phenomenex Synergi polar-RP 80 Å LC column (4.6 mm x 150 mm, 4 μm). The mobile phase was composed of 0.2% aqueous formic acid (A) and methanol (B). Gradient elution was as follows: 30% B in 0-3minutes and 30-60% in 3-50minutes. The post-run time was 5 minutes. The flow rate was 1.0 mL/min. The column temperature was set at 30°C. The injection volume was 5μL. The ESI-MS system was operated in negative ionization mode. M/z for of fragment ions for each compound were listed in **Table 2.2**. The instrument settings were set as follows: the capillary voltage was 3kV, the nebulizer (N_2) pressure was 50 psi, the drying gas temperature was 350°C, the drying gas flow was 10 L/min, and the fragmentor voltage was 70 V. The UV detector was monitored at 350nm for quantification of seven flavonoids.

Table 2.2 SIM method parameters for flavonoids

Name	Retention Time (min)	POS/NEG	Quantifier
1. Orientin	15.393	NEG	447
2. Vitexin	19.444	NEG	431
3. Isovitexin	21.643	NEG	431
4. Quercetin	36.288	NEG	301
5. Luteolin	40.932	NEG	285

6. Kaempferol	44.875	NEG	285
7. Apigenin	48.708	NEG	269

2.3.8 GC-MS setup for terpenoids and sterols assay

The GC-MS system used in this study was an Agilent 7890A GC system comprised of the following components: an Agilent 7890A Gas Chromatograph (G3440A), an Agilent 5975C inert MSD with triple-axis detector, a K`(Prime) GC sample injector (MXY 02-01B), and a Phenomenex ZB-5MSi column (30m × 0.25 mm, 0.25 µm). A temperature gradient program was used for the separation of mono- and sesquiterpenoids: 40°C for 2 minutes, ramp of 20°C/min up to 100°C, ramp of 5°C/min up to 160°C, and ramp of 20°C/min up to 280°C. Run time was 20 minutes. The injector temperature was 280°C. Injection volume was 1 µL. Split ratio was 10:1. The carrier gas (helium) flow rate was 1.2 mL/min. The MS source was set to 230°C, the single quad temperature was 150 °C, and the transfer line temperature was set to 280°C. The mass spectrometer was operated in selected ion monitoring (SIM) mode. Quantifier and qualifier ions for each compound were listed in **Table 2.3**. A second temperature gradient program was used for the quantification of triterpenoids and sterols: 80°C for 1 minute, the ramp of 20°C/min up to 250°C, and ramp of 10°C/min up to 300°C. Run time was 34.5 minutes. The injector temperature was 300°C. The injection volume was 1µL and splitless. The carrier gas (helium) flow rate was 1.5 mL/min. The MS source was set to 230°C, the single quad temperature was 150 °C, and the transfer line temperature was set to 280°C. The mass spectrometer was operated in SIM mode. Quantifier and qualifier ions for each compound were listed in **Table 2.4.**

Table 2.3 SIM method parameters for monoterpenoids and sesquiterpenoids

	Retention Time (min)	Quantifier	Qualifier 1	Qualifier 2	Qualifier 3
IS: tridecane	11.497	71	85	57	
1. α-Pinene	5.597	77	91	121	
2. Camphene	5.787	93	121	136	
3. Sabinene	6.073	93	77	79	136
4. β-Pinene	6.132	93	69	77	136
5. β-Myrcene	6.242	69	79	136	
6. α-Phellandrene	6.462	93	77	91	136
7. Δ^3-Carene	6.542	77	91	121	
8. α-Terpinene	6.621	121	91	93	136
9. *p*-Cymene	6.732	119	91	134	
10. Limonene	6.790	93	121	136	
11. 1,8-Cineole (Eucalyptol)	6.835	154	93	139	
12. Ocimene	7.004	93	79	121	

13. γ-Terpinene	7.207	93	77	121	136
14. Sabinene Hydrate	7.373	71	93	121	154
15. Terpinolene	7.662	93	121	136	
16. Fenchone	7.685	81	69	152	
17. Linalool	7.816	93	69	121	
18. Fenchol	8.145	93	111	121	
19. (-)-Isopulegol	8.656	121	136	154	
20. Camphor	8.646	95	108	152	
21. Borneol	9.064	110	67	139	
22. Terpinen-4-ol	9.238	71	93	111	154
23. α-Terpineol	9.492	121	59	136	
24. (+)-Dihydrocarvone	9.569	95	109	137	152
25. Nerol	10.120	69	93	121	
26. Pulegone	10.398	109	67	152	
27. (+)-Carvone	10.485	82	93	108	
28. Geraniol	10.613	69	93	123	
29. Geranyl Acetate	13.309	69	93	121	
30. (-)-β-Elemene	13.665	147	161	189	
31. β-Caryophyllene	14.339	105	91	79	
32. Aromadendrene	14.775	161	189	204	
33. trans-β-Farnesene	14.988	69	93	133	
34. α-Humulene	15.100	93	79	91	
35. Valencene	15.974	161	189	204	
36. Ledene	16.000	161	189	204	
37. trans-Nerolidol	17.406	93	107	136	
38. Caryophyllene Oxide	17.851	95	105	107	
39. Globulol	17.889	161	189	204	
40. Viridiflorol	18.021	161	189	204	
41. (-)-Guaiol	18.067	161	105	107	
42. (+)-Cedrol	18.192	150	95	151	
43. β-Eudesmol	18.810	59	149	164	
44. α-Bisabolol	19.135	204	189	161	

Table 2.4 SIM method parameters for triterpenoids and sterols

Name	Retention Time (min)	Quantifier	Qualifier 1	Qualifier 2
IS: Cholesterol	16.093	105	133	386
1. Campesterol	17.044	105	133	400
2. Stigmasterol	17.350	105	133	412
3. β-Sitosterol	17.955	105	133	414
4. β-Amyrin	18.466	218	189	203
5. Epifriedelanol	20.519	109	123	413
6. Friedelin	20.900	109	123	426

2.3.9 GC-FID setup for terpenoids identification and semi-quantification

Terpenoids without available standards were identified and semi-quantified using GC-MS for identification and GC-FID for semi-quantification, respectively. A Hewlett Packard 5890 Series II GC equipped with a 7673A automatic injector and a flame ionization detector (FID) was used

for the analysis of the available terpenoids standards and samples. The instrument was equipped with a ZB-5HT capillary column (30m × 0.25 mm ID, 0.25 μm film thickness) with the injector temperature at 280 °C, an injection volume of 2μL, a split ratio of 10:1, and carrier gas (helium) flow rate of 1.2mL/min. The temperature gradient started at 60°C and increased at a rate of 3°C/min until 280°C. The total run time was 75 minutes. The mass scan range was from 30amu to 550amu. One 1000μg/mL saturated alkanes standard (C7-C40), one 100μg/mL mixed terpenoid standard, and the samples were injected using the temperature gradient program above. The linear retention index (LRI) was calculated by comparing the retention time of one terpenoid compound ($t_{R,i}$) with those of n-alkanes with n carbons eluted before the compound ($t_{R,n}$) and with n+1 carbons eluted after the compound ($t_{R,n+1}$)[142]:

$$\text{LRI (target compound)} = 100 \times \left(\frac{t_{R,i} - t_{R,n}}{t_{R,n+1} - t_{R,n}} + n\right)$$

Each LRI was compared to the data listed in the NIST Chemistry WebBook[143]. The mass spectrum of the target compound was compared to data in the NIST mass-spectra database embedded in the GC-MS system. If both the LRI and the mass spectrum confirm the identity of the target compound, then the compound can be semi-quantified by comparing the response area of the target compound and a closely eluted compound with known concentration while assuming that the relative response factor is one[127,144].

2.3.10 Sample preparation optimization

Sample preparation procedures were compared and optimized step by step. Two pulverization methods, which were manual grinding and electric blending, were compared for preparing cannabis inflorescence material. The extraction efficiency of solvents (methanol vs. methanol/chloroform (9/1, v/v)) [79,145] for cannabinoids were compared. Extraction durations (sonication for 10 vs. 20 vs. 30 minutes vs. maceration for one day) were studied using the yield of total extracted cannabinoids and total extracted mono- and sesquiterpenoids. The effect of sonication temperature (20°C vs. 30°C vs. 50°C) and the number of extractions (once vs. twice vs. thrice) were also compared for cannabinoids. For triterpenoids and sterols, the compared extraction methods were sonication for 1 hour and maceration for 1, 2, 3, 4, or 5 days in terms of total triterpenoids and total sterols extracted. Five duplicate samples were tested for each scenario. To

investigate potential interference from cannabinoids during flavonoids testing, the effects of the hexane wash before acid hydrolysis were examined by comparing flavonoids yields.

2.3.11 Method validation

Developed methods were validated for selectivity, linearity, trueness, precision (repeatability and intermediate precision), the limit of detections (LOD), the limit of detection (LOQ), and robustness (using a different column, instrument, and analysts). Measurement uncertainty (accuracy) was determined using the total error concept[146]. Matrix effects and extraction efficiency were also determined for cannabinoids.

2.3.12 Selectivity

Selectivity was determined by injecting a solvent blank to confirm that there were no false signal peaks at the targeted retention time. Each compound standard was individually injected to determine retention times for GC and LC analysis. Representative chromatograms were used to demonstrate selectivity. Each compound was labelled correspondingly.

2.3.13 Calibration curve and linearity

A 100 µg/mL mixed standard solution of 14 cannabinoids was further diluted to 1, 0.5, 0.1, 0.05, and 0.01 µg/mL to construct a linear regression curve. The linearity of the responses was confirmed visually by plotting residuals against concentrations. Similarly, calibration curves for mono- and sesquiterpenoids were constructed at concentrations of 1, 5, 25, 50, 100, and 250 µg/mL. Calibration curves for triterpenoids and sterols were constructed at concentrations of 1, 5, 25, 50, and 100 µg/mL. For flavonoids, a 100 µg/mL mixed standard was further diluted into 25, 10, 5, 2, and 1 µg/mL to construct calibration curves.

2.3.14 LOD and LOQ

The instrument LOD is the minimum concentration that distinguishable from background noise within 99% confidence. Multiple duplicate standards that had concentrations near the expected LOD were measured[147]. The relative standard deviations (RSD) of integrated areas were used for instrument LOD determination as follows:

$$\text{LOD (µg/mL)} = t_{\alpha} \times \text{RSD} \times \text{the concentration of injected standard (µg/mL)}$$

where t_α is the confidence factor from the student's t-distribution table (one-sided), and α is the significance level (α = 0.01). For ten injections, t_α = 2.821. LOQ is three times LOD. The repeatedly injected concentration was 0.01 µg/mL for cannabinoids, 0.5 µg/mL for terpenoids and sterols, and 1 µg/mL for flavonoids.

2.3.15 Trueness, precision, and accuracy

Accuracy, expressed as the total error of a method, is affected by systematic error (trueness) and random error (precision). This study determined accuracy using the total error concept[148]. Low, medium and high concentrations of mixed standards were spiked into blank matrices and tested for trueness and precision. Mint leaves were prepared as the cannabinoid blank matrix. For terpenoids and sterols, cannabis plant material was repeatedly washed using organic solvent until the measured terpenoids and sterols levels were below LOD. Approximately 200 mg of the blank matrices were spiked with three levels (0.1, 0.5, and 1.0 for cannabinoids, 50, 100, and 200 µg/mL for mono- and sesquiterpenoids, and 5, 10, and 25 µg/mL for triterpenoids and sterols) of mixed standards with three replicates for each concentration level analyzed for each of three consecutive days. The trueness was calculated using the following equation:

$$\text{Relative bias (\%)} = \frac{\text{Measured spiked value} - \text{Nominal spiked value}}{\text{Nominal spiked value}} \times 100\%$$

$$= \text{Recovery (\%)} - 100\%$$

The measured values were averaged to calculate relative bias for each level (n=9). Cannabis leaves were used to spike three levels of flavonoid standards (20, 50, and 80 µg for orientin and isovitexin; 20, 30, and 80 for vitexin; 50, 100, and 150 µg for quercetin, luteolin, kaempferol, and apigenin) for three replicates before hydrolysis.

2.3.16 Repeatability and intermediate precision

The intraday repeatability was determined as the RSD from assaying blank matrix samples spiked with three levels of standards. Repeatability was calculated as the pooled RSD over the three intraday RSD values:

$$RSD_{pooled} = \sqrt{\frac{RSD_1^2 + RSD_2^2 + RSD_3^2}{3}}$$

Intermediate precision for all analytes except flavonoids was evaluated as the RSD at each concentration level for the average measured value of nine replicates over three days. Intermediate precision for flavonoids was calculated as RSD for twelve replicates by testing six samples for two consecutive days by two analysts.

2.3.17 Accuracy

The total uncertainty in measurement combines bias and error in intermediate precision, and is calculated as follows:

$$\text{Measurement uncertainty (total error)} = \sqrt{(\text{relative bias})^2 + (\text{RSD(intermediate precision)})^2}$$

2.3.18 Matrix effect and extraction efficiency

The matrix effect for cannabinoids was measured by comparing the response of the blank matrix extract spiked with three levels of standards (low, medium, high) to the response of standards of the same concentration in the solvent, calculated as:

Matrix effect =

$$\frac{\text{Analyte response (post-extraction spiked blank matrix)}}{\text{Analyte response (solvent)}} \times 100\%$$

Extraction recovery is measured by comparing the response of the blank matrix before and after extraction, calculated as:

Extraction efficiency =

$$\frac{\text{Analyte response (pre-extraction spiked blank matrix)}}{\text{Analyte response (post-extraction spiked blank matrix)}} \times 100\%$$

2.3.19 Robustness

The robustness of the cannabinoids LC-MS method was verified using an alternate analytical column and a different LC-MS instrument of the same model. The robustness of the GC-MS method for mono- and sesquiterpenoids was verified by an alternate analyst who quantified the same sample batch on an alternate day. Six duplicate samples were tested for each scenario.

2.3.20 Statistical analysis

Each experiment was independently repeated five times for sample preparation optimization. Each compound was measured three times for each plant part. Data is expressed as mean ± standard deviation (SD). Data sets were compared using the two-sided student's *t*-tests at the 0.05 significance level. For multiple groups, one-way ANOVA followed by Tukey honestly significant difference (HSD) post hoc test at the 0.05 significance level were used.

2.4 Results

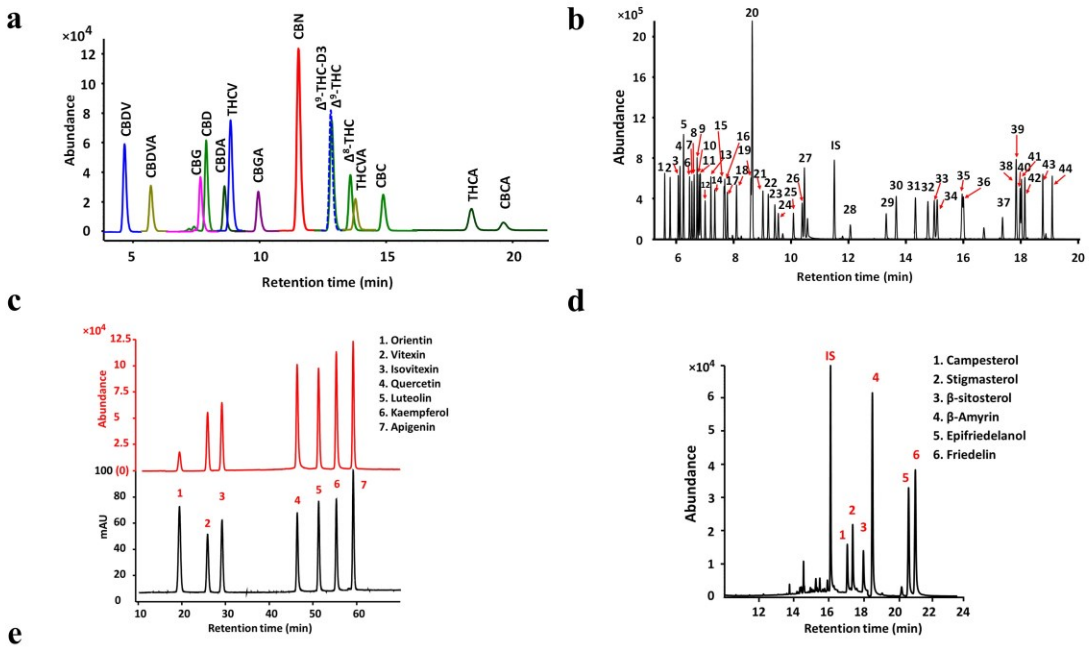
2.4.1 Sample preparation optimization

The yield of total cannabinoids averaged 17.5 ± 0.5% (n=5) using manual grinding with a handheld herb grinder, which was significantly higher (n=5, $p<0.0001$) than using an electric blender, a yield of which averaged 12.0 ± 0.3%. The minimization of analyte loss using the manual grinder is attributed to the fact that resin adheres to the blades and plastic housing surface of a plastic blender during high-speed pulverization (**Figure 2.3**). There were no significant differences in extraction efficiency for cannabinoids between two solvents, methanol and a 9:1 methanol/chloroform mixture (n=5, $p = 0.6379$). Because methanol is less toxic than methanol/chloroform, methanol was used as the solvent in the following tests. The duration of sonication (10, 20, and 30 minutes) had no significant differences in cannabinoid extraction (n=5, $p = 0.3351$). However, yield after sonication was found to be slightly lower than maceration for one day (n=5, $p = 0.0248$). Four extraction methods were tested for terpenoids (sonication at 10, 20, and 30 minutes and maceration for one day after sonication for 20 minutes) and found to have no significant differences in total mono- and sesquiterpenoids yield (n=5, $p = 0.9904$). Sonication at room temperature (20°C) extracted higher total cannabinoids compared to 30°C and 50°C (n=5, $p = 0.018$). Whether extraction was performed once, twice, or thrice did not have significant effects on total cannabinoid yield (n=5, $p = 0.3995$). For all the following experiments, cannabinoids and terpenoids were extracted once using methanol by sonication at room temperature for 20 minutes. For extraction

of total sterols in stem bark, sonication for one hour, maceration for one, two, three, four, and five days were significantly different (n=5, $p < 0.0001$) and the main differences were between sonication and maceration. The differences between sonication and maceration for the extraction of total sterols in root material were not significant (n=5, $p= 0.0661$). For extraction of total triterpenoids in stem bark material, sonication for one hour, maceration for one, two, three, four, and five days were not significantly different (n=5, $p = 0.8001$). The comparison between sonication and maceration for the extraction of total triterpenoids in root material achieved similar results (n=5, $p = 0.1221$). Despite a previous study's concern that large amounts of cannabinoids may interfere with flavonoid quantification[127], the three situations compared in this study (no hexane wash, one hexane wash, and three hexane washes before acid hydrolysis) had no significant difference in leaf material (n=3, $p = 0.8701$) and a reduction in flavonoid yield in inflorescence material (n=3, $p < 0.0001$).

2.4.2 Method validation results for cannabinoids

The chromatogram for a standard solution of 14 mixed cannabinoids by LC-MS is shown in **Figure 2.4(a)**. The slope of the regression curve and the coefficient of determination for each compound were calculated (**Table 2.5**). The correlation coefficients for all 14 cannabinoids were above 0.9998. The intercept for each compound was set to zero because the p-value > 0.05 by analysis of variance (ANOVA) indicates insufficient evidence to reject the null hypothesis that the intercept is 0. LOD was between 0.0004 and 0.004 µg/mL, and LOQ was between 0.001 and 0.01 µg/mL. Repeatability was between 0.4% and 9.2% for all compounds (**Table 2.6**). Intermediate precision was between 1.5% and 12.3%. All relative biases were between -6.4% and 6.9% and all measurement uncertainties were between 1.5% and 12.3%. The matrix effect and extraction efficiency are listed in **Table 2.7**. The matrix effect for all three levels was between 93.03% - 101.65%. The extraction recoveries for all three levels were between 80% - 120%, except for CBGA at 1.0 µg/mL (77.21%) and THCVA at 1.0 µg/mL (79.03%). Compared to their neutral forms, cannabinoid acids had higher degradation during sonication. Method robustness was verified using an alternate chromatographic column and a second LC-MS instrument. Neither the columns (n=5, $p = 0.2914$) nor the machines (n=5, $p = 0.9580$) showed significant differences in extracted cannabinoids.



Monoterpenoids			Sesquiterpenoids	
1. α-Pinene	11. Eucalyptol	21. Borneol	30. (-)-β-Elemene	40. Viridiflorol
2. Camphene	12. Ocimene	22. Terpinen-4-ol	31. β-Caryophyllene	41. (-)-Guaiol
3. Sabinene	13. γ-Terpinene	23. α-Terpineol	32. Aromadendrene	42. (+)-Cedrol
4. (-)-β-Pinene	14. Sabinene Hydrate	24. (+)-Dihydrocarvone	33. trans-β-Farnesene	43. β-Eudesmol
5. β-Myrcene	15. Terpinolene	25. Nerol	34. α-Humulene	44. α-Bisabolol
6. α-Phellandrene	16. Fenchone	26. Pulegone	35. Valencene	
7. Δ3-Carene	17. Linalool	27. Carvone	36. Ledene	
8. α-Terpinene	18. Fenchol	28. Geraniol	37. trans-Nerolidol	
9. *p*-Cymene	19. (-)-Isopulegol	29. Geranyl Acetate	38. Caryophyllene Oxide	
10. Limonene	20. Camphor		39. Globulol	

Figure 2.4 Chromatograms for cannabinoids, mono- and sesquiterpenoids, flavonoids, sterols, and triterpenoids.

(a) Chromatogram for a standard solution of 14 mixed cannabinoids by LC-ESI-MS. (b) Chromatogram for 44 mono- and sesquiterpenoids by GC-MS. Terpenoids corresponded to the labeled number are listed in (e). (c) Chromatogram for 7 flavonoids by HPLC-UV-MS. Mass spectrometry was used for flavonoid identification and UV detector was used for flavonoid quantification. (d) Chromatogram for 3 sterols and 3 triterpenoids by GC-MS. (e) Compound names for 44 mono- and sesquiterpenoids.

Table 2.5 ANOVA regression statistics for cannabinoids.

	Quantification Range (μg/mL)	Correlation Coefficient R^2	Slope ± SD	LOD* (μg/mL)	LOQ (μg/mL)
1. CBDV	0.01-1.00	1.0000	1.3246±0.0031	0.001	0.003
2. CBDVA	0.01-1.00	1.0000	0.7432±0.0008	0.002	0.005
3. CBG	0.01-1.00	0.9998	0.8367±0.0053	0.002	0.005
4. CBD	0.01-1.00	1.0000	1.3841±0.0040	0.0005	0.001
5. CBDA	0.01-1.00	1.0000	0.7174±0.0004	0.001	0.003

6. THCV	0.01-1.00	1.0000	1.8763±0.0022	0.0004	0.001
7. CBGA	0.01-1.00	1.0000	0.6690±0.0015	0.003	0.008
8. CBN	0.01-1.00	0.9999	2.9331±0.0127	0.001	0.002
9. Δ^9-THC	0.01-1.00	1.0000	1.8042±0.0024	0.001	0.003
10. Δ^8-THC	0.01-1.00	1.0000	0.9145±0.0009	0.002	0.005
11. THCVA	0.01-1.00	0.9999	0.5474±0.0028	0.003	0.010
12. CBC	0.01-1.00	1.0000	0.6329±0.0018	0.002	0.007
13.THCA	0.01-1.00	1.0000	0.5700±0.0012	0.003	0.009
14. CBCA	0.01-1.00	0.9998	0.2333±0.0017	0.004	0.012

*Instrument LOD determined by multiple injections of standards with concentrations near the expected LOD.

Table 2.6 Trueness, precision, and accuracy for cannabinoids

Analyte	Spike 0.1 µg/mL				Spike 0.5 µg/mL				Spike 1.0 µg/mL			
	Repeatability - Pooled RSD% (n=3)	**Intermediate precision - RSD% (n=9)**	**Trueness - Relative bias% (n=9)**	**Accuracy - Uncertainty of measurement% (n=9)**	**Repeatability - Pooled RSD% (n=3)**	**Intermediate precision - RSD% (n=9)**	**Trueness - Relative bias% (n=9)**	**Accuracy - Uncertainty of measurement% (n=9)**	**Repeatability - Pooled RSD% (n=3)**	**Intermediate precision - RSD% (n=9)**	**Trueness - Relative bias% (n=9)**	**Accuracy - Uncertainty of measurement% (n=9)**
1. CBDV	1.0	1.7	-4.0	4.3	1.0	4.8	-0.1	4.8	0.9	3.6	-2.6	4.4
2. CBDVA	1.7	1.5	0.2	1.5	0.8	4.6	-0.3	4.6	1.2	3.8	-2.7	4.7
3. CBG	3.0	12.3	0.9	12.3	6.5	6.1	2.1	6.4	4.0	7.7	-5.2	9.3
4. CBD	1.7	2.0	-4.8	5.2	0.7	4.9	-1.1	5.1	1.0	3.9	-3.1	5.0
5. CBDA	1.7	3.7	-4.2	5.6	1.3	4.7	-1.2	4.9	1.3	4.2	-3.4	5.4
6. THCV	1.5	2.3	-4.1	4.7	0.6	4.8	-1.0	4.9	1.0	4.0	-3.0	5.0
7. CBGA	3.8	3.5	-1.8	3.9	0.9	4.7	-1.2	4.9	0.9	3.7	-3.5	5.1
8. CBN	0.9	2.6	-6.4	6.9	0.6	5.5	-2.5	6.1	1.1	4.6	-4.2	6.2
9. Δ9-THC	1.9	2.6	-3.9	4.7	0.5	5.3	-1.3	5.5	1.3	4.2	-3.3	5.3
10. Δ8-THC	1.7	3.7	-4.8	6.0	0.4	5.4	-1.4	5.6	1.0	4.4	-3.9	5.9
11. THCVA	3.2	4.0	-4.8	6.2	0.7	4.7	-1.4	4.9	1.4	4.4	-4.0	5.9
12. CBC	1.9	3.7	-6.3	7.3	0.7	5.0	-1.2	5.2	1.1	4.7	-3.5	5.9
13.THCA	9.2	12.0	0.1	12.0	6.4	5.5	-1.2	5.6	1.2	5.0	-4.2	6.6
14. CBCA	8.5	8.8	6.9	11.2	1.8	4.7	0.0	4.7	0.9	3.7	-3.2	4.8

Table 2.7 Matrix effect and extraction efficiency (recovery).

Analyte	Spike 0.1µg/mL		Spike 0.5µg/mL		Spike 1.0 µg/mL	
	Matrix effect ± SD	Extraction efficiency (recovery± SD)	Matrix effect ± SD	Extraction efficiency (recovery± SD)	Matrix effect ± SD	Extraction efficiency (recovery± SD)
1. CBDV	97.96 ± 0.38	94.41 ± 1.36	96.63 ± 1.40	94.52 ± 0.79	96.78 ± 0.45	99.62 ± 4.23
2. CBDVA	101.65 ± 0.18	87.96 ± 5.07	97.63 ± 1.32	87.29 ± 1.09	97.76 ± 0.32	90.05 ± 3.30
3. CBG	86.39 ± 0.73	98.97 ± 1.07	97.98 ± 0.66	94.42 ± 3.89	94.49 ± 0.10	80.04 ± 2.19
4. CBD	96.72 ± 1.88	95.68 ± 1.20	96.11 ± 0.23	94.04 ± 0.50	96.07 ± 0.51	94.26 ± 5.40
5. CBDA	98.47 ± 0.84	85.05 ± 3.56	96.40 ± 0.82	87.91 ± 1.20	96.58 ± 0.16	89.70 ± 3.26
6. THCV	96.89 ± 0.74	96.55 ± 1.14	96.07 ± 0.63	94.45 ± 0.17	96.04 ± 0.41	80.73 ± 2.19
7. CBGA	94.14 ± 5.08	83.96 ± 3.02	96.45 ± 0.74	86.44 ± 1.17	96.30 ± 0.34	89.99 ± 2.99
8. CBN	94.82 ± 1.78	95.71 ± 1.19	93.24 ± 0.42	93.73 ± 0.51	93.90 ± 0.39	77.21 ± 2.18
9. Δ^9-THC	97.11 ± 1.19	100.44 ± 1.13	95.43 ± 0.57	95.45 ± 0.53	94.60 ± 0.58	89.28 ± 2.29
10. Δ^8-THC	98.13 ± 0.66	97.50 ± 1.00	95.16 ± 0.31	94.73 ± 0.58	94.97 ± 0.65	89.17 ± 2.61
11. THCVA	98.55 ± 1.22	82.83 ± 3.95	96.75 ± 1.05	86.49 ± 0.79	95.28 ± 0.57	89.15 ± 3.04
12. CBC	96.94 ± 1.54	98.26 ± 1.32	95.42 ± 0.65	94.46 ± 0.80	94.73 ± 0.19	79.03 ± 1.54
13.THCA	100.02 ± 16.37	98.28 ± 9.88	96.53 ± 2.21	90.27 ± 1.27	93.03 ± 3.22	88.17 ± 3.00
14. CBCA	99.53 ± 0.33	94.30 ± 0.81	96.63 ± 1.40	89.41 ± 0.72	93.72 ± 0.60	82.94 ± 1.41

2.4.3 Method validation for mono- and sesquiterpenoids

The correlation coefficients for all 44 terpenoids were above 0.9989 (**Table 2.8**). LOD were between 0.009 and 0.167 µg/mL, and LOQ were between 0.026 and 0.500 µg/mL. Repeatability was between 0.4% and 6.4% for all compounds (**Table 2.9**). Intermediate precision was between 0.6% and 8.8%. All relative biases were between -6.3% and 8.7%, and all measurement uncertainties were between 1.5% and 9.1%. Robustness was evaluated by two analysts operating on the same machine by testing twelve replicate cannabis samples. Results were not significantly different in terms of total mono- and sesquiterpenoid yield (n=5, p=0.9588).

Table 2.8 ANOVA regression statistics and LOD and LOQ for mono- and sesquiterpenoids.

	Quantification Range (µg/mL)	R^2	Slope ± SD	LOD* (µg/mL)	LOQ (µg/mL)
1. α-Pinene	1 - 250	0.9993	0.0043 ± 0.0001	0.022	0.065
2. Camphene	1 - 250	0.9999	0.0093 ± 0.0000	0.020	0.059
3. Sabinene	1 - 250	0.9996	0.0154 ± 0.0001	0.015	0.045
4. β-Pinene	1 - 250	0.9994	0.0172 ± 0.0002	0.018	0.054
5. β-Myrcene	1 - 250	0.9998	0.0080 ± 0.0000	0.014	0.042
6. α-Phellandrene	1 - 250	0.9998	0.0161 ± 0.0001	0.014	0.043
7. Δ^3-Carene	1 - 250	0.9996	0.0038 ± 0.0000	0.015	0.044
8. α-Terpinene	1 - 250	0.9997	0.0094 ± 0.0001	0.013	0.038
9. P-Cymene	1 - 250	0.9997	0.0249 ± 0.0002	0.012	0.037
10. Limonene	1 - 250	0.9999	0.0069 ± 0.0000	0.029	0.088
11. 1,8-Cineole (Eucalyptol)	1 - 250	0.9991	0.0020 ± 0.0000	0.009	0.026
12. Ocimene	1 - 250	0.9998	0.0073 ± 0.0000	0.015	0.044
13. γ-Terpinene	1 - 250	1.0000	0.0146 ± 0.0000	0.012	0.035
14. Sabinene Hydrate	1 - 250	0.9997	0.0054 ± 0.0000	0.016	0.049
15. Terpinolene	1 - 250	0.9996	0.0081 ± 0.0001	0.011	0.033

	Quantification Range (µg/mL)	R^2	Slope ± SD	LOD* (µg/mL)	LOQ (µg/mL)
16. Fenchone	1 - 250	0.9995	0.0194 ± 0.0002	0.012	0.035
17. Linalool	1 - 250	0.9996	0.0062 ± 0.0001	0.046	0.137
18. Fenchol	1 - 250	0.9999	0.0015 ± 0.0000	0.012	0.035
19. (-)-Isopulegol	1 - 250	1.0000	0.0023 ± 0.0000	0.012	0.037
20. Camphor	1 - 250	0.9996	0.0116 ± 0.0001	0.010	0.029
21. Borneol	1 - 250	0.9998	0.0036 ± 0.0000	0.026	0.077
22. Terpinen-4-ol	1 - 250	0.9994	0.0112 ± 0.0001	0.016	0.047
23. α-Terpineol	1 - 250	1.0000	0.0038 ± 0.0000	0.012	0.035
24. (+)-Dihydrocarvone	1 - 250	0.9999	0.0033 ± 0.0000	0.012	0.037
25. Nerol	1 - 250	0.9995	0.0111 ± 0.0001	0.167	0.500
26. Pulegone	1 - 250	0.9998	0.0037 ± 0.0000	0.014	0.041
27. (+)-Carvone	1 - 250	0.9999	0.0106 ± 0.0000	0.024	0.072
28. Geraniol	1 - 250	0.9991	0.0145 ± 0.0002	0.125	0.375
29. Geranyl Acetate	1 - 250	0.9996	0.0094 ± 0.0001	0.022	0.065
30. (-)-β-Elemene	1 - 250	0.9999	0.0016 ± 0.0000	0.017	0.051
31. β-Caryophyllene	1 - 250	0.9997	0.0021 ± 0.0000	0.015	0.045
32. Aromadendrene	1 - 250	0.9996	0.0020 ± 0.0000	0.011	0.033
33. trans-β-Farnesene	1 - 250	0.9992	0.0105 ± 0.0001	0.019	0.056
34. α-Humulene	1 - 250	0.9995	0.0104 ± 0.0001	0.013	0.040
35. Valencene	1 - 250	0.9996	0.0034 ± 0.0000	0.014	0.041
36. Ledene	1 - 250	0.9997	0.0014 ± 0.0000	0.010	0.029
37. Trans-Nerolidol	1 - 250	0.9995	0.0024 ± 0.0000	0.017	0.051
38. Caryophyllene Oxide	1 - 250	0.9997	0.0034± 0.0000	0.031	0.093
39. Globulol	1 - 250	0.9989	0.0017± 0.0000	0.019	0.057
40. Viridiflorol	1 - 250	0.9996	0.0020± 0.0000	0.020	0.060
41. (-)-Guaiol	1 - 250	0.9998	0.0043± 0.0000	0.018	0.054
42. (+)-Cedrol	1 - 250	0.9997	0.0029± 0.0000	0.017	0.051
43. β-Eudesmol	1 - 250	0.9997	0.0033± 0.0000	0.028	0.083
44. α-Bisabolol	1 - 250	0.9995	0.0007± 0.0000	0.011	0.033

*Instrument LOD determined by multiple injections of standards with concentrations near the expected LOD.

Table 2.9 Trueness, precision, and accuracy for mono- and sesquiterpenoids.

	Analyte	Spike 10 µg/mL				Spike 20 µg/mL				Spike 50 µg/mL			
		Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)
1.	**α-Pinene**	1.5	4.1	-3.9	5.7	1.6	2.8	0.9	2.9	1.8	4.4	1.3	4.6
2.	**Camphene**	1.3	2.0	-0.3	2.0	1.5	2.4	3.1	3.9	1.2	4.0	2.8	4.9
3.	**Sabinene**	0.8	8.8	2.4	9.1	1.4	3.8	4.2	5.7	0.7	4.0	4.0	5.6
4.	**β-Pinene**	0.9	7.2	0.7	7.2	1.4	3.0	4.0	5.0	1.0	4.2	3.8	5.7
5.	**β-Myrcene**	0.7	6.5	4.2	7.7	1.2	3.3	6.9	7.6	0.8	2.8	5.7	6.3
6.	**α-Phellandrene**	0.7	5.1	0.5	5.1	1.4	2.8	3.4	4.4	1.1	3.5	4.0	5.4
7.	**Δ^3-Carene**	0.8	3.4	-0.2	3.4	1.5	3.2	4.6	5.6	1.2	3.6	5.2	6.3
8.	**α-Terpinene**	0.7	6.2	0.9	6.3	1.4	2.4	3.3	4.1	1.2	3.3	4.5	5.5
9.	**P-Cymene**	0.7	2.0	-1.8	2.7	1.6	2.1	4.4	4.8	0.6	2.1	6.7	7.0
10.	**Limonene**	3.7	4.1	-2.4	4.7	4.5	4.0	1.8	4.4	1.9	4.2	2.3	4.8
11.	**1,8-Cineole(Eucalyptol)**	1.2	7.7	-1.8	7.9	1.2	2.3	3.4	4.1	0.7	1.3	6.0	6.1
12.	**Ocimene**	0.5	2.4	2.0	3.1	1.3	1.3	2.8	3.1	0.5	2.1	3.0	3.6
13.	**γ-Terpinene**	0.5	3.4	1.0	3.6	1.2	2.3	2.5	3.4	0.7	4.2	2.4	4.9
14.	**Sabinene Hydrate**	1.9	4.4	-0.1	4.4	3.0	2.7	6.1	6.7	4.8	5.9	3.8	7.0
15.	**Terpinolene**	0.5	5.0	0.3	5.1	1.3	2.0	3.7	4.2	0.6	1.3	5.1	5.2
16.	**Fenchone**	1.9	2.8	-3.9	4.8	1.6	2.3	4.6	5.2	0.6	0.6	7.9	7.9
17.	**Linalool**	4.9	6.2	3.0	6.9	3.7	3.4	4.2	5.4	1.6	3.3	5.7	6.6
18.	**Fenchol**	1.9	2.8	4.7	5.4	1.2	1.1	7.2	7.3	0.8	1.0	8.7	8.8
19.	**(-)-Isopulegol**	3.8	3.7	1.5	4.0	1.3	1.8	6.6	6.8	1.2	1.2	7.9	8.0
20.	**Camphor**	0.8	4.1	1.7	4.5	1.2	2.1	5.7	6.1	1.3	3.8	6.2	7.2
21.	**Borneol**	1.3	2.7	1.2	3.0	1.7	2.3	2.2	3.2	1.3	3.4	4.5	5.6
22.	**Terpinen-4-ol**	3.9	5.8	2.9	6.5	1.2	4.1	5.8	7.2	1.5	2.9	4.7	5.5
23.	**α-Terpineol**	0.7	3.2	0.5	3.2	1.1	4.1	5.3	6.7	1.7	2.4	6.3	6.7
24.	**(+)-Dihydrocarvone**	1.3	3.8	0.1	3.8	1.1	2.7	3.8	4.7	1.1	2.2	6.0	6.4
25.	**Nerol**	2.6	3.6	-1.5	3.9	6.4	6.0	-1.3	6.2	3.6	3.6	4.5	5.7
26.	**Pulegone**	3.6	5.2	3.8	6.4	3.2	3.4	5.9	6.9	1.6	2.7	5.0	5.7
27.	**(+)-Carvone**	3.2	4.9	0.2	4.9	2.9	3.0	2.3	3.8	1.9	2.4	5.7	6.1
28.	**Geraniol**	2.0	5.9	2.3	6.4	3.5	5.3	3.1	6.2	1.2	2.0	6.2	6.5
29.	**Geranyl Acetate**	2.7	3.3	3.6	4.9	1.5	2.6	6.8	7.3	1.0	1.9	6.3	6.5
30.	**(-)-β-Elemene**	0.4	1.5	-0.2	1.5	1.0	1.1	3.9	4.0	1.3	3.5	6.7	7.5
31.	**β-Caryophyllene**	0.6	3.1	-4.9	5.8	0.9	3.4	3.3	4.7	0.9	1.1	6.9	7.0
32.	**Aromadendrene**	1.4	1.7	-6.3	6.6	1.1	5.1	1.0	5.2	1.2	1.3	5.9	6.0
33.	**trans-β-Farnesene**	0.8	5.5	4.9	7.4	1.1	1.4	6.4	6.6	0.7	1.2	7.7	7.8
34.	**α-Humulene**	2.2	3.8	-4.6	5.9	3.1	3.0	1.5	3.3	1.2	2.2	5.6	6.1
35.	**Valencene**	1.2	4.0	2.3	4.6	2.4	2.4	5.0	5.5	1.5	1.4	6.2	6.4
36.	**Ledene**	3.9	5.4	-1.2	5.5	2.6	2.5	1.0	2.7	2.7	3.2	4.2	5.3
37.	**trans-Nerolidol**	2.3	3.9	-1.3	4.1	3.5	5.5	1.6	5.7	2.4	2.4	5.1	5.7
38.	**Caryophyllene Oxide**	2.9	2.9	2.7	4.0	2.6	2.8	2.9	4.0	1.4	2.0	4.8	5.2
39.	**Globulol**	2.9	5.8	1.3	5.9	3.0	4.1	2.1	4.6	2.0	2.0	4.6	5.0
40.	**Viridiflorol**	2.9	2.7	5.8	6.4	2.4	2.8	5.9	6.6	2.2	3.3	5.0	5.9
41.	**(-)-Guaiol**	4.0	4.4	1.8	4.7	3.4	3.6	1.9	4.0	3.5	3.5	4.2	5.4

Analyte	Spike 10 µg/mL				Spike 20 µg/mL				Spike 50 µg/mL			
	Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability – Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)
42. (+)-Cedrol	2.9	4.0	-0.2	4.0	2.2	2.9	3.2	4.3	2.2	2.0	6.6	6.9
43. β-Eudesmol	2.5	3.0	4.4	5.3	2.6	4.0	4.1	5.7	1.7	1.7	6.5	6.7
44. α-Bisabolol	1.3	2.3	6.6	7.0	2.8	3.6	4.4	5.7	1.7	1.8	6.2	6.5

2.4.4 Method validation for flavonoids

The correlation coefficients for all seven compounds were greater than 0.9997 (**Table 2.10**). Trueness, determined by recovery, for seven flavonoids by acid hydrolysis were between 71.5±1.3% and 106.6±4.0% for level 1, between 70.5±0.9% and 95.8±0.8% for level 2, and between 75.1±0.7% and 94.7±1.7% for level 3 (**Table 2.11**). Recovery for luteolin (84.1±3.5% for level 1, 80.0±2.6% for level 2, and 80.8±1.5% for level 3) and apigenin (80.5±0.9% for level 1, 78.7±1.9% for level 2, and 81.1±0.6%% for level 3) were comparable with a previous study's recovery results of 82% for luteolin and 81% for apigenin[85]. The method is repeatable with intraday RSD% (n=3) ranging between 1.20% and 4.10% for level 1, between 0.9% and 3.2% for level 2, and between 1.0% and 3.0% for level 3. The intermediate precision calculated from twelve replicates of leaf samples ranged between 1.70% and 3.3% and ranged between 2.1% and 5.6% for the cannabis inflorescence sample.

Table 2.10 ANOVA regression statistics for flavonoids.

	Quantification Range (µg/mL)	Correlation Coefficient R^2	Slope ± SD	LOD* (µg/mL)	LOQ (µg/mL)
1. Orientin	1-25	1.0000	21.73 ± 0.06	0.04	0.12
2. Vitexin	1-25	1.0000	12.40 ± 0.03	0.04	0.11
3. Isovitexin	1-25	1.0000	14.99 ± 0.02	0.04	0.13
4. Quercetin	1-25	1.0000	14.02 ± 0.31	0.08	0.23
5. Luteolin	1-25	1.0000	15.76 ± 0.09	0.04	0.13
6. Kaempferol	1-25	0.9997	15.84 ± 0.21	0.07	0.20
7. Apigenin	1-25	1.0000	19.48 ± 0.07	0.06	0.17

*Instrument LOD determined by multiple injections of standards with concentrations near the expected LOD.

Table 2.11 Accuracy (recovery) and repeatability for flavonoids in cannabis leaf over three spiked levels.

Spike level	Compound	Nominal spiked mass (µg)	Mass in sample (µg) (n=3) (mean±SD)	Measured spiked mass (µg) (n=3) (mean±SD)	Recovery% (n=3) (mean±SD)	RSD% (n=3)
Level 1	1. Orientin	20	34.27±0.04	17.77±0.52	88.8±2.6%	2.9%
	2. Vitexin	20	75.24±0.08	21.33±0.80	106.6±4.0%	3.8%
	3. Isovitexin	20	6.23±0.01	15.87±0.36	79.3±1.8%	2.3%
	4. Quercetin	50	ND	35.73±0.67	71.5±1.3%	1.9%
	5. Luteolin	50	58.10±0.07	42.03±1.74	84.1±3.5%	4.1%
	6. Kaempferol	50	ND	36.23±0.86	72.5±1.7%	2.4%
	7. Apigenin	50	29.50±0.03	40.23±0.46	80.5±0.9%	1.2%
Level 2	1. Orientin	50	34.27±0.04	46.39±1.49	92.8±3.0%	3.2%
	2. Vitexin	30	79.05±0.27	28.75±0.25	95.8±0.8%	0.9%
	3. Isovitexin	50	6.23±0.01	36.25±0.76	72.5±1.5%	2.1%
	4. Quercetin	100	ND	72.07±0.64	72.1±0.6%	0.9%
	5. Luteolin	100	58.10±0.07	80.02±2.57	80.0±2.6%	3.2%

Spike level	Compound	Nominal spiked mass (µg)	Mass in sample (µg) (n=3) (mean±SD)	Measured spiked mass (µg) (n=3) (mean±SD)	Recovery% (n=3) (mean±SD)	RSD% (n=3)
	6. Kaempferol	100	ND	70.50±0.87	70.5±0.9%	1.2%
	7. Apigenin	100	29.50±0.03	78.74±1.85	78.7±1.9%	2.4%
Level 3	1. Orientin	80	34.27±0.04	75.78±1.34	94.7±1.7%	1.8%
	2. Vitexin	80	75.24±0.08	75.60±1.64	94.5±2.0%	2.2%
	3. Isovitexin	80	6.23±0.01	63.52±1.05	79.4±1.3%	1.7%
	4. Quercetin	150	ND	115.07±3.51	76.7±2.3%	3.0%
	5. Luteolin	150	58.10±0.07	121.22±2.28	80.8±1.5%	1.9%
	6. Kaempferol	150	ND	112.67±1.10	75.1±0.7%	1.0%
	7. Apigenin	150	29.50±0.03	121.69±0.85	81.1±0.6%	0.7%

2.4.5 Method validation for sterols and triterpenoids

The correlation coefficients for all 6 compounds were between 0.9989 and 0.9999 (**Table 2.12**). LOD were between 0.17 and 0.26 µg/mL, and LOQ were between 0.50 and 0.79 µg/mL. Repeatability was between 0.4% and 9.2% for all compounds (**Table 2.13**). Intermediate precision for 9 replicates was between 1.1% and 4.7%. All relative biases were between -4.0% and 1.4%, and all measurement uncertainties were between 1.4% and 5.8%.

Table 2.12 ANOVA regression statistics for sterols and triterpenoids.

	Quantification Range (µg/mL)	Correlation Coefficient R^2	Slope ± SD	LOD* (µg/mL)	LOQ (µg/mL)
1. Campesterol	1.00-100	0.9996	0.0134±0.0001	0.20	0.61
2. Stigmasterol	1.00-100	0.9999	0.0129±0.0001	0.21	0.63
3. β-Sitosterol	1.00-100	0.9989	0.0123±0.0002	0.26	0.79
4. β-Amyrin	1.00-100	0.9992	0.0559±0.0007	0.17	0.50
5. Epifriedelanol	1.00-100	0.9989	0.0280±0.0004	0.25	0.74
6. Friedelin	1.00-100	0.9998	0.0284±0.0002	0.26	0.78

*Instrument LOD determined by multiple injections of standards with concentrations near the expected LOD.

Table 2.13 Trueness, precision, and accuracy for compounds for sterols and triterpenoids.

Analyte	Spike 5 µg/mL				Spike 10 µg/mL				Spike 25 µg/mL			
	Repeatability - Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability - Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)	Repeatability - Pooled RSD% (n=3)	Intermediate precision - RSD% (n=9)	Trueness - Relative bias% (n=9)	Accuracy - Uncertainty of measurement% (n=9)
1. Campesterol	4.9%	4.7%	-2.6%	5.5%	2.3%	2.1%	-1.1%	2.5%	2.3%	2.1%	0.5%	2.3%
2. Stigmasterol	4.2%	4.7%	-4.0%	5.8%	4.0%	3.9%	0.2%	4.0%	1.6%	1.4%	0.8%	1.8%
3. β-Sitosterol	2.5%	2.9%	0.1%	2.5%	2.7%	3.4%	0.7%	2.8%	1.5%	1.3%	0.8%	1.7%
4. β-Amyrin	4.7%	4.3%	-2.8%	5.5%	1.8%	2.8%	-0.8%	2.0%	2.4%	2.2%	0.1%	2.4%
5. Epifriedelanol	3.1%	4.5%	-2.9%	4.3%	1.5%	2.2%	0.0%	1.5%	1.1%	1.1%	0.8%	1.4%
6. Friedelin	3.2%	3.0%	0.4%	3.3%	2.2%	2.9%	1.4%	2.7%	3.84%	3.6%	0.1%	3.8%

2.4.6 Cannabinoids profile in inflorescences, leaves, stem bark, and roots

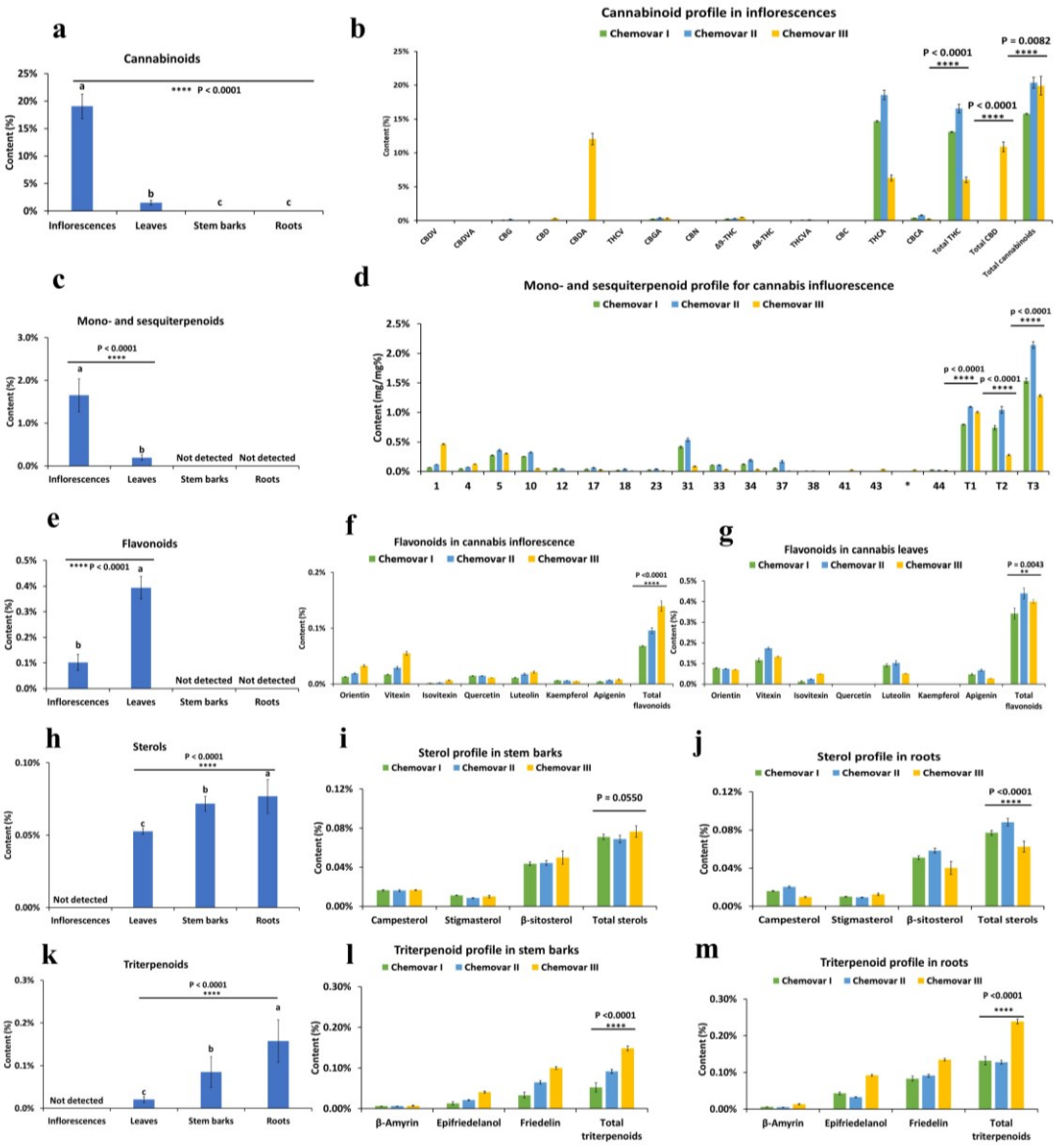


Figure 2.5 Secondary metabolites profiling in cannabis roots, stem bark, leaves, and inflorescences.

(a) Total cannabinoid content (mg/mg%) in each part of cannabis plant averaged from three cannabis chemovars (n=9, mean ± SD %). (b) Individual and total cannabinoid content (mg/mg%) in cannabis inflorescences of three chemovars (n=3, mean ± SD %). Asterisks indicate statistically significant differences (one-way ANOVA, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). (c) Total mono- and sesquiterpenoid content (mg/mg%) in each part of cannabis plant averaged from three cannabis chemovars (n=9, mean ± SD %). (d) Individual and total mono- and sesquiterpenoids content (mg/mg%) in cannabis inflorescences of three chemovars (n=3, mean ± SD %). Terpenoids labelled by their numbers are listed in **Figure 2.4(e)**. The compound labelled as Asterisks was α-eudesmol, which was semi-quantified by GC-FID. T1 = total monoterpenoids. T2 = total sesquiterpenoids. T3 = total mono- and sesquiterpenoids. (e) Total flavonoid content (mg/mg%) in each part of cannabis plants averaged from three cannabis chemovars (n=9, mean ± SD %). (f) Individual and total flavonoid content (mg/mg%) in cannabis inflorescences of three chemovars (n=3, mean ± SD %). (g) Individual and total flavonoid content (mg/mg%) in cannabis leaves of three chemovars (n=3, mean ±

SD %). (h) Total sterol content (mg/mg%) in each part of cannabis plant averaged from three cannabis chemovars (n=9, mean ± SD %). (i) Individual sterol content (mg/mg%) in cannabis stem bark of three chemovars (n=3, mean ± SD %). (j) Individual sterol content (mg/mg%) in cannabis roots of three chemovars (n=3, mean ± SD %). (k) Total triterpenoid content (mg/mg%) in each part of cannabis plant averaged from three cannabis chemovars (n=9, mean ± SD %). (l) Individual and total triterpenoid content (mg/mg%) in cannabis stem bark of three chemovars (n=3, mean ± SD %). (m) Individual and total triterpenoid content (mg/mg%) in cannabis stem bark of three chemovars (n=3, mean ± SD %). In each figure, the one-way ANOVA followed by correction for multiple comparisons (Tukey honestly significant difference (HSD) post hoc test) at the 0.05 significance level was used (p values indicated above each bar). Asterisks indicate statistically significant differences (one-way ANOVA, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

Cannabinoid content decreased from inflorescences to leaves, stem bark, and roots (**Figure 2.5(a)**). Roots contained between 0.001% and 0.004% cannabinoids in all three chemovars (**Table 2.14**), which agrees with the minuscule amounts reported by other studies (0% and 0.03%)[79,80]. Stem bark contained between 0.005% and 0.008% cannabinoids in all three chemovars and was found to be less than the amounts previously reported (0.02% and 0.1-0.3%)[79,80]. Differences may be caused by variations in chemovar and the position where the sample was taken (next to root). Cannabinoids quantified in cannabis leaf and inflorescence are shown in **Figure 2.5(b)** and listed in **Table 2.15**. Total cannabinoids quantified in leaves were between 1.10% and 2.10%, which agreed with the previously-reported amounts (1-2% and 1.40-1.75%)[79,124] but not others (0.05%)[80]. Total cannabinoids quantified in inflorescence were between 15.77% and 20.37% in all three chemovars, as typical of modern drug-type chemovars[72,76,91,149].

Table 2.14 Cannabinoid profile in root and stem bark for three strains.

Compound	Strain I root	Strain I stem bark	Strain II root	Strain II stem bark	Strain III root	Strain III stem bark
1. CBDV	ND	ND	ND	ND	ND	ND
2. CBDVA	ND	ND	ND	ND	ND	ND
3. CBG	0.0007±0.00003%	0.0002±0.00001%	ND	ND	0.0001±0.00002%	0.0001±0.00009%
4. CBD	ND	ND	ND	ND	ND	ND
5. CBDA	ND	ND	ND	ND	ND	ND
6. THCV	ND	ND	ND	ND	ND	ND
7. CBGA	ND	0.0001±0.00003%	ND	0.0001±0.00001%	0.0002±0.00001%	0.0001±0.00001%
8. CBN	ND	0.0001±0.00001%	ND	ND	ND	0.0001±0.00001%
9. Δ^9-THC	0.0001±0.00008%	0.0006±0.00008%	ND	0.0003±0.000001%	0.0001±0.0001%	0.0010±0.0002%
10. Δ^8-THC	0.0001±0.00016%	0.0001±0.00005%	0.0001±0.00001%	ND	ND	0.0001±0.00004%
11. THCVA	0.0001±0.00004%	0.0002±0.00002%	0.0001±0.00001%	0.0001±0.00002%	0.0001±0.00001%	0.0002±0.00004%
12. CBC	ND	0.0001±0.00001%	ND	ND	ND	0.0001±0.000009%
13.THCA	0.0016±0.0004%	0.0058±0.0005%	0.0008±0.00001%	0.0040±0.0003%	0.0037±0.0002%	0.0062±0.0008%
14. CBCA	0.0001±0.00005%	0.0003±0.00005%	ND	0.0003±0.000005%	0.0001±0.00001%	0.0004±0.00009%
Total THC**	0.0015±0.0005%	0.0056±0.0005%	0.0007±0.00002%	0.0038±0.0003%	0.0034±0.00004%	0.0064±0.00082%
Total CBD**	0.0001±0.00001%	0.0001±0.00001%	ND	ND	ND	0.0001±0.00001%
Total cannabinoids	0.0027±0.0006%	0.0074±0.0008%	0.0011±0.00005%	0.0050±0.00007%	0.0043±0.00007%	0.0083±0.001%

* Content expressed in mean ± SD% (n=3). ND=Not detected.

Table 2.15 Cannabinoid profile in leaves and inflorescences for three strains.

Compound	Strain I leaf	Strain I inflorescence	Strain II leaf	Strain II inflorescence	Strain III leaf	Strain III inflorescence
1. CBDV	ND	ND	ND	ND	ND	ND
2. CBDVA	ND	ND	ND	ND	ND	0.05 ± 0.003%
3. CBG	ND	0.08 ± 0.004%	ND	0.18 ± 0.01%	ND	0.03 ± 0.001%
4. CBD	ND	ND	ND	ND	0.02 ± 0.001%	0.33 ± 0.02%
5. CBDA	ND	0.04 ± 0.001%	ND	0.04 ± 0.002%	1.16 ± 0.02%	12.06 ± 0.84%
6. THCV	ND	ND	ND	ND	ND	ND
7. CBGA	0.02 ± 0.001%*	0.26 ± 0.01%	0.02 ± 0.001%	0.37 ± 0.04%	0.02 ± 0.001%	0.35 ± 0.03%
8. CBN	ND	ND	ND	ND	ND	ND
9. Δ^9-THC	0.05 ± 0.002%	0.24 ± 0.003%	0.06 ± 0.003%	0.30 ± 0.02%	0.03 ± 0.001%	0.47 ± 0.02%
10. Δ^8-THC	ND	ND	ND	ND	ND	ND
11. THCVA	0.02 ± 0.001%	0.09 ± 0.004%	0.01 ± 0.001%	0.13 ± 0.007%	ND	0.02 ± 0.001%
12. CBC	ND	ND	ND	ND	ND	0.04 ± 0.002%
13.THCA	1.01 ± 0.02%	14.68 ± 0.07%	0.68 ± 0.02%	18.55 ± 0.70%	0.62 ± 0.01%	6.32 ± 0.44%
14. CBCA	0.29 ± 0.005%	0.39 ± 0.002%	0.34 ± 0.01%	0.79 ± 0.05%	0.25 ± 0.01%	0.24 ± 0.02%
Total THC**	0.93 ± 0.01%	13.11 ± 0.06%	0.65 ± 0.02%	16.57 ± 0.63%	0.57 ± 0.01%	6.02 ± 0.40%
Total CBD**	ND	0.03 ± 0.001%	ND	0.04 ± 0.002%	1.04 ± 0.02%	10.91 ± 0.75%
Total cannabinoids	1.42 ± 0.023%	15.77 ± 0.81%	1.10 ± 0.04%	20.37 ± 0.80%	2.10 ± 0.04%	19.93 ± 1.36%

* Content expressed in mean (n=3) ± SD%. ND=Not detected.
** [91] Total THC = THC + 0.877× THCA, Total CBD = CBD + 0.877× CBDA

Chemovars I and II displayed THC dominant profiles, with THCA as the dominant compound (14.68% and 18.55%) and other cannabinoids less than 1% in both leaf and inflorescence tissue[127]. Chemovar III showed a total CBD to THCA ratio of 1.8, which matched with its reported profile in its marketing materials. These amounts were representative of modern North American-cultivated seedless chemovars, which contain up to 25% total cannabinoids, with THCA and CBDA as the main constituents[91]. Cannabinoids mainly exist in the plant as carboxylic acids and are decarboxylated into neutral forms over time - heat or light exposure expedites decarboxylation[79]. Due to the convertibility of THCA, total THC dose is calculated as the sum of the amount of THCA multiplied by a correction factor 0.877 plus the amount of THC[79]. Neutral form cannabinoids, including CBDV, CBG, CBD, THCV, Δ^9-THC, and CBC, were either not detected or found at several times less than acid form cannabinoids. CBN was detected at less than 0.01% in the leaf and inflorescence samples of the chemovars – this indicates that there was minimal degradation and that sample preparation was proper[127]. The ratios of total THC to total CBD matched with some of those representatives of the wide-leaflet drug (WLD) ("Indica" in the vernacular) and narrow-leaflet drug (NLD) ("Sativa" in the vernacular) biotypes[67,76] but contradicted others[72,27]. Studies have shown that the concentrations of total THC and total CBD have no discriminatory value for chemovars in the modern vernacular ("Sativa" vs. "Indica") due to the misuse of the botanical nomenclature, extensive cross-breeding, and unreliable labelling during unrecorded hybridization [67,70–73,83,84].

CBDVA was detectable in Chemovar III at 0.05% but was not detected in the other two chemovars. The correlation between CBDVA and CBD is unclear, but elevated levels of CBDV and THCV are more common in *C. indica* drug biotypes (WLD and NLD) than the *C. sativa* hemp biotype[73]. CBDV is reported to rival CBD's therapeutic potential for the treatment of epilepsy, particularly focal seizures[150]. It is also reported to have therapeutic potential for treating nausea and vomiting[151]. Total THC and total CBD ratio in the leaves of the intermediate type chemovar was consistent with that in the inflorescence, which is consistent with conclusions from other studies[78,152,153]. Notably, the ratio of total CBC to total THC was ten times higher in leaves than in inflorescence for all three chemovars.

2.4.7 Mono- and sesquiterpenoid profile in inflorescences, leaves, stem bark, and roots

Mono- and sesquiterpenoids were not detected in stem bark or roots (**Figure 2.5(c)**). Total mono- and sesquiterpenoids ranged from 0.125% to 0.278% in leaf and 1.283% to 2.141% in inflorescence in the three chemovars (**Table 2.16**), which were less than the 4% reported in unfertilized flowers in a previous study[80]. Total sesquiterpenoid content was higher than total monoterpenoids in fan leaves in Chemovar I and Chemovar II but was comparable in Chemovar III. This observation was clearer when contents were expressed as ratios: sesquiterpenoids comprised approximately 90% of total terpenoids in Chemovar I and II and comprised 53% of the total terpenoids in Chemovar III.

Table 2.16 Mono- and sesquiterpenoid profile in cannabis leaf and inflorescence for three cannabis strains

	Compound	LRI Calc.	LRI Lit.[a]	Strain I Leaf	Strain I inflorescence	Strain II Leaf	Strain II inflorescence	Strain III Leaf	Strain III inflorescence
1.	α-Pinene	934	932	0.004±0.0002%*	0.067±0.002%	0.002±0.0001%	0.117±0.006%	0.083±0.003%	0.463±0.006%
2.	Camphene	948	952	ND	0.008±0.0001%	ND	0.011±0.0006%	ND	ND
3.	Sabinene	974	976	0.001±0.0001%	ND	0.001±0.0001%	ND	ND	ND
4.	β-Pinene	978	980	0.002±0.0001%	0.047±0.001%	0.001%±0.000%	0.071±0.005%	0.020±0.0007%	0.124±0.007%
5.	β-Myrcene	992	992	0.001±0.0001%	0.274±0.013%	ND	0.359±0.016%	0.024±0.0007%	0.302±0.011%
6.	α-Phellandrene	1006	1006	ND	ND	ND	ND	ND	ND
7.	Δ^3-Carene	1011	1011	ND	ND	ND	ND	ND	ND
8.	α-Terpinene	1017	1017	ND	ND	ND	ND	ND	ND
9.	P-Cymene	1025	1026	ND	ND	ND	ND	ND	ND
10.	Limonene	1029	1031	ND	0.252±0.004%	0.004±0.0001%	0.322±0.015%	0.005±0.0002%	0.049±0.002%
11.	1,8-Cineole (Eucalyptol)	1031	1032	0.006±0.0016%	ND	0.006±0.001%	0.006±0.0003%	ND	ND
12.	Ocimene	1047	1050	ND	0.047±0.006%	ND	0.045±0.003%	ND	0.002±0.0001%
13.	γ-Terpinene	1058	1059	ND	ND	ND	ND	ND	ND
14.	Sabinene Hydrate	1069	1068	ND	0.003±0.0001%	0.001±0.0002%	0.004±0.0005%	ND	0.003±0.0003%
15.	Terpinolene	1089	1088	ND	0.003±0.0001%	ND	0.003±0.0001%	ND	0.001±0.0001%
16.	Fenchone	1088	1088	ND	0.003±0.0001%	ND	0.003±0.0001%	ND	0.002±0.0002%
17.	Linalool	1103	1100	ND	0.037±0.005%	0.001±0.0001%	0.064±0.006%	ND	0.027±0.002%

	Compound	LRI Calc.	LRI Lit.[a]	Strain I Leaf	Strain I inflorescence	Strain II Leaf	Strain II inflorescence	Strain III Leaf	Strain III inflorescence
18.	Fenchol	1117	1112	ND	0.023±0.001%	ND	0.040±0.004%	ND	0.009±0.0009%
19.	(-)-Isopulegol	1146	1146	ND	0.002±0.0001%	ND	0.004±0.0005%	ND	0.002±0.0002%
20.	Camphor	1143	1143	ND	ND	ND	ND	ND	ND
21.	Borneol	1168	1168	ND	0.004±0.0002%	ND	0.007±0.0007%	ND	0.005±0.0004%
22.	Terpinen-4-ol	1179	1179	ND	0.001±0.0001%	ND	0.001±0.0001%	ND	0.001±0.0001%
23.	α-Terpineol	1194	1190	ND	0.024±0.001%	ND	0.040±0.004%	ND	0.013±0.0005%
24.	(+)-Dihydrocarvone	1197	1200	ND	ND	ND	ND	ND	ND
25.	Nerol	1232	1228	ND	ND	ND	ND	ND	ND
26.	Pulegone	1239	1244	ND	ND	ND	ND	ND	ND
27.	(+)-Carvone	1245	1243	ND	ND	ND	ND	ND	ND
28.	Geraniol	1258	1256	ND	0.001±0.001%	ND	0.001±0.0002%	ND	ND
29.	Geranyl Acetate	1385	-	ND	ND	ND	ND	ND	ND
	Total monoterpenoids	-	-	0.014±0.002%	0.796±0.008%	0.016±0.002%	1.097±0.045%	0.132±0.005%	1.004±0.043%
30.	(-)-β-Elemene	1394	1392	ND	ND	ND	ND	ND	ND
31.	β-Caryophyllene	1420	1420	0.077±0.003%	0.416±0.030%	0.057±0.002%	0.538±0.034%	0.053±0.0024%	0.086±0.002%
32.	Aromadendrene	1440	1440	ND	ND	ND	ND	ND	ND
33.	trans-β-Farnesene	1459	1446	0.032±0.003%	0.106±0.002%	0.021±0.002%	0.106±0.008%	0.034±0.002%	0.035±0.001%
34.	α-Humulene	1455	1455	0.025±0.001%	0.123±0.006%	0.022±0.0007%	0.193±0.019%	0.021±0.001%	0.036±0.002%
β – selinene		1485	1485	ND	0.043±0.001%	ND	0.046±0.001%	ND	0.034±0.001%
α – selinene		1494	1496	ND	0.037±0.001%	ND	0.036±0.001%	ND	0.026±0.001%
35.	Valencene	1494	1491	ND	ND	ND	ND	ND	0.001±0.001%
36.	Ledene	1497	1493	ND	0.011±0.0004%	ND	0.013±0.0009%	ND	0.005±0.0004%
	α - Farnesene	1508	1508	ND	0.017±0.001%	ND	0.025±0.001%	ND	0.024±0.001%
37.	trans-Nerolidol	1568	1565	0.012±0.003%	0.048±0.014%	0.009±0.0005%	0.166±0.014%	0.004±0.0004%	0.009±0.001%
38.	Caryophyllene Oxide	1584	1583	ND	0.006±0.0003%	ND	0.005±0.0008%	ND	0.002±0.0001%
39.	Globulol	1588	1584	ND	0.002±0.0001%	ND	0.001±0.0001%	ND	ND
40.	Viridiflorol	1595	1588	ND	ND	ND	ND	ND	ND
41.	(-)-Guaiol	1602	1602	ND	ND	ND	ND	0.007±0.0003%	0.027±0.0009%
42.	(+)-Cedrol	1605	1601	ND	ND	ND	ND	ND	ND
43.	β-Eudesmol	1651	1650	ND	0.001±0.0002%	ND	0.002±0.0001%	0.004±0.0002%	0.034±0.001%
	α -Eudesmol	1653	1657	ND	0.002±0.0001%	ND	0.002±0.0002%	0.003±0.0002%	0.026±0.001%
44.	α-Bisabolol	1685	1684	0.013±0.002%	0.027±0.005%	ND	0.002±0.0006%	0.020±0.002%	0.018±0.0006%
Total sesquiterpenoids		-	-	0.159±0.005%	0.741±0.035%	0.109±0.004%	1.044±0.043%	0.146±0.008%	0.279±0.011%
Total mono- and sesquiterpeniods		-	-	0.173±0.007%	1.537±0.043%	0.125±0.005%	2.141±0.102%	0.278±0.011%	1.283±0.018%

* Terpenoid content expressed in mean ± SD% (n=3). Bolded terpenoid was semi-quantified by GC-FID. ND=Not detected

[a] LRI from the literature[143,154,155].

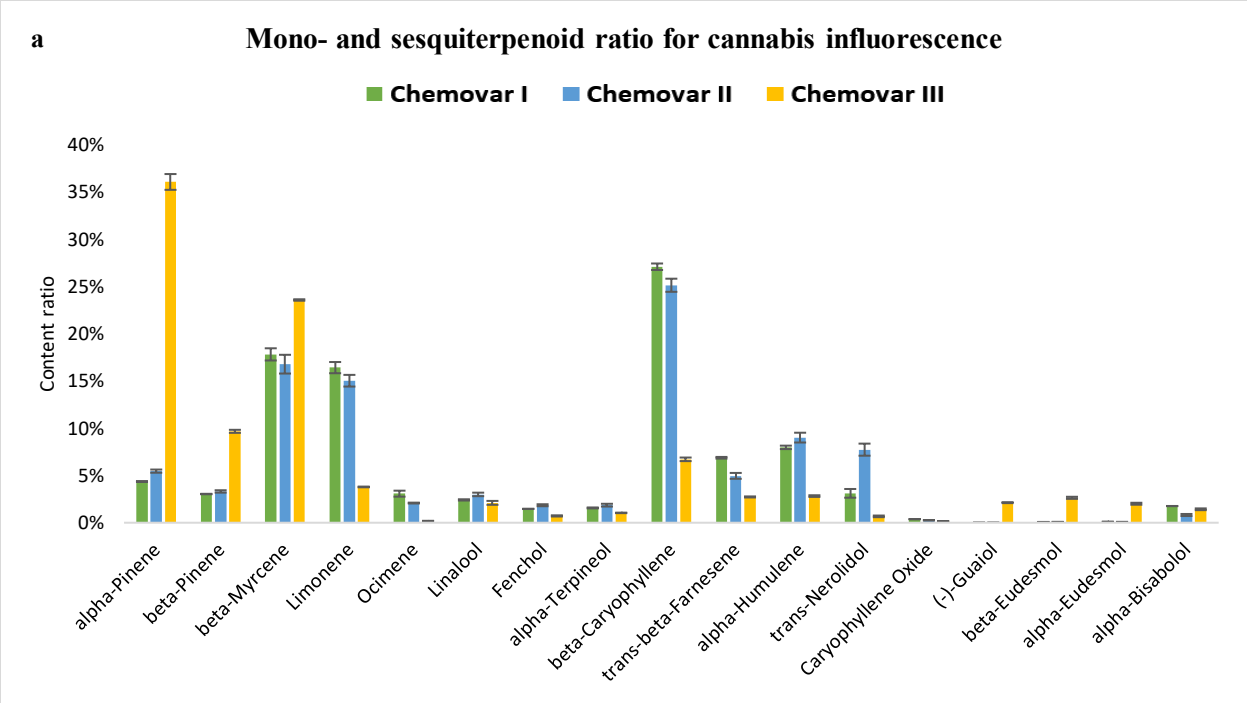
The ratios of major terpenoids relative to total terpenoids in the inflorescence (**Table 2.17**), agreed with values reported in a compiled study[79]. β-myrcene was the most abundant monoterpenoid at concentrations ranging from 16.78% to 23.57%. α-Pinene ranged between 4.26% and 36.07%. β-Pinene ranged between 3.04% and 7.12%. Limonene ranged between 3.79% and 16.42%. Linalool ranged between 2.10% and 2.99%. β-Caryophyllene was the most abundant sesquiterpenoid and ranged between 6.71% and 45.25%. α-Humulene ranged between 2.82% and 7.97%. β-Eudesmol ranged between 0.07% and 2.64%. All mono- and sesquiterpenoid ratios were consistent with previously reported essential oil contents in fresh plant material (between 47.9% - 92.48% and 6.84% - 47.5%, respectively)[79]. The ratios of individual terpenoids in the leaf were comparable to those in inflorescence for all three chemovars (**Figure 2.6**).

Table 2.17 Mono- and sesquiterpenoid ratios relative to total terpenoids in leaves and inflorescences

Compound	LRI Calc.	LRI Lit.[a]	Strain I Leaf	Strain I inflorescence	Strain II Leaf	Strain II inflorescence	Strain III Leaf	Strain III inflorescence
1. α-Pinene	934	932	2.37±0.22%*	4.36±0.04%	1.82±0.07%	5.46±0.17%	29.81±0.75%	36.07±0.84%
2. Camphene	948	952	0.18±0.015%	0.51±0.01%	0.17±0.01%	0.51±0.02%	ND	ND
3. Sabinene	974	976	0.30±0.05%	ND	0.56±0.05%	ND	ND	ND
4. β-Pinene	978	980	1.08±0.06%	3.04±0.02%	0.89±0.02%	3.31±0.13%	7.12±0.16%	9.67±0.16%
5. β-Myrcene	992	992	0.52±0.07%	17.82±0.65%	ND	16.78±0.99%	8.74±0.16%	23.57±0.08%
6. α-Phellandrene	1006	1006	ND	ND	ND	ND	ND	ND
7. Δ^3-Carene	1011	1011	ND	ND	0.21±0.05%	ND	ND	ND
8. α-Terpinene	1017	1017	0.11±0.02%	ND	ND	ND	ND	ND
9. P-Cymene	1025	1026	ND	ND	0.23±0.03%	ND	ND	ND
10. Limonene	1029	1031	ND	16.42±0.59%	0.74±0.14%	15.03±0.62%	1.77±0.06%	3.79±0.04%
11. 1,8-Cineole (Eucalyptol)	1031	1032	3.19±0.80%	ND	4.56±0.73%	0.28±0.04%	ND	0.03±0.01%
12. Ocimene	1047	1050	ND	3.07±0.32%	ND	2.08±0.04%	ND	0.19±0.01%
13. γ-Terpinene	1058	1059	0.14±0.01%	ND	0.23±0.03%	ND	ND	ND
14. Sabinene Hydrate	1069	1068	ND	0.22±0.01%	0.74±0.14%	0.02±10.57%	ND	0.27±0.02%
15. Terpinolene	1089	1088	ND	0.16±0.01%	ND	0.15±0.01%	ND	0.07±0.001%
16. Fenchone	1088	1088	ND	0.22±0.001%	ND	0.16±0.001%	ND	0.13±0.01%
17. Linalool	1103	1100	ND	2.41±0.07%	ND	2.99±0.19%	ND	2.10±0.21%
18. Fenchol	1117	1112	ND	1.46±0.01%	ND	1.85±0.11%	ND	0.72±0.06%
19. (-)-Isopulegol	1146	1146	ND	0.15±0.001%	ND	0.17±0.01%	ND	0.12±0.01%
20. Camphor	1143	1143	ND	ND	ND	ND	ND	ND
21. Borneol	1168	1168	ND	0.29±0.01%	ND	0.32±0.02%	ND	0.40±0.02%
22. Terpinen-4-ol	1179	1179	ND	0.06±0.001%	ND	0.05±0.001%	ND	0.06±0.01%
23. α-Terpineol	1194	1190	ND	1.56±0.05%	ND	1.86±0.15%	ND	1.04±0.03%
24. (+)-Dihydrocarvone	1197	1200	ND	ND	ND	ND	ND	ND
25. Nerol	1232	1228	ND	ND	ND	ND	ND	ND
26. Pulegone	1239	1244	ND	ND	ND	ND	ND	ND
27. (+)-Carvone	1245	1243	ND	ND	ND	ND	ND	0.02±0.001%
28. Geraniol	1258	1256	ND	0.04±0.001%	ND	0.05±0.001%	ND	ND
29. Geranyl Acetate	1385	-	ND	ND	ND	ND	ND	ND
Total monoterpenoids	-	-	7.91±0.72%	51.8±0.94%	12.93±0.82%	51.25±1.55%	47.45±1.12%	78.25±0.65%
30. (-)-β-Elemene	1394	1392	ND	ND	ND	ND	ND	ND
31. β-Caryophyllene	1420	1420	44.61±0.74%	27.09±0.35%	45.25±0.78%	25.14±0.70%	19.25±0.34%	6.71±0.19%
32. Aromadendrene	1440	1440	ND	ND	ND	ND	ND	ND
33. trans-β-Farnesene	1459	1446	18.75±2.23%	6.87±0.09%	17.12±1.3%	4.96±0.31%	12.17±0.06%	2.73±0.05%
34. α-Humulene	1455	1455	14.22±0.19%	7.97±0.18%	17.34±0.85%	9.01±0.52%	7.68±0.22%	2.82±0.09%
β - selinene	1485	1485	ND	0.04±0.001%	ND	0.05±0.001%	ND	0.03±0.001%
α - selinene	1494	1496	ND	0.04±0.001%	ND	0.04±0.001%	ND	0.03±0.001%
35. Valencene	1494	1491	ND	ND	ND	ND	ND	0.06±0.001%
36. Ledene	1497	1493	ND	0.74±0.01%	ND	0.63±0.03%	ND	0.39±0.03%
α - Farnesene	1508	1508	ND	0.02±0.01%	ND	0.03±0.001%	ND	0.02±0.001%
37. trans-Nerolidol	1568	1565	7.07±1.51%	3.11±0.46%	7.35±0.38%	7.73±0.64%	1.49±0.14%	0.67±0.09%
38. Caryophyllene Oxide	1584	1583	ND	0.38±0.01%	ND	0.25±0.02%	ND	0.16±0.01%
39. Globulol	1588	1584	ND	0.10±0.001%	ND	0.06±0.001%	ND	ND
40. Viridiflorol	1595	1588	ND	ND	ND	ND	ND	ND
41. (-)-Guaiol	1602	1602	ND	ND	ND	ND	2.36±0.05%	2.12±0.05%
42. (+)-Cedrol	1605	1601	ND	ND	ND	ND	ND	0.02±0.001%
43. β-Eudesmol	1651	1650	ND	0.07±0.01%	ND	0.07±0.02%	1.50±0.04%	2.64±0.11%
α -Eudesmol	1653	1657	ND	0.10±0.02%	ND	0.08±0.01%	0.90±0.09%	2.01±0.10%
44. α-Bisabolol	1685	1684	7.42±0.65%	1.76±0.01%	ND	0.82±0.11%	7.20±0.63%	1.42±0.09%
Total sesquiterpenoids	-	-	92.08±0.72%	48.20±0.94%	87.07±0.82%	48.75±1.55%	52.55±1.12%	21.75±0.65%
Total mono- and sesquiterpenoids	-	-	100%	100%	100%	100%	100%	100%

* Terpenoid ratio (relative to total terpenoid) is expressed as mean value ± SD% (n=3). ND= Not detected. Bolded entries were semi-quantified by relative area by GC-FID.

[a] LRI from the literature[143,154,155].



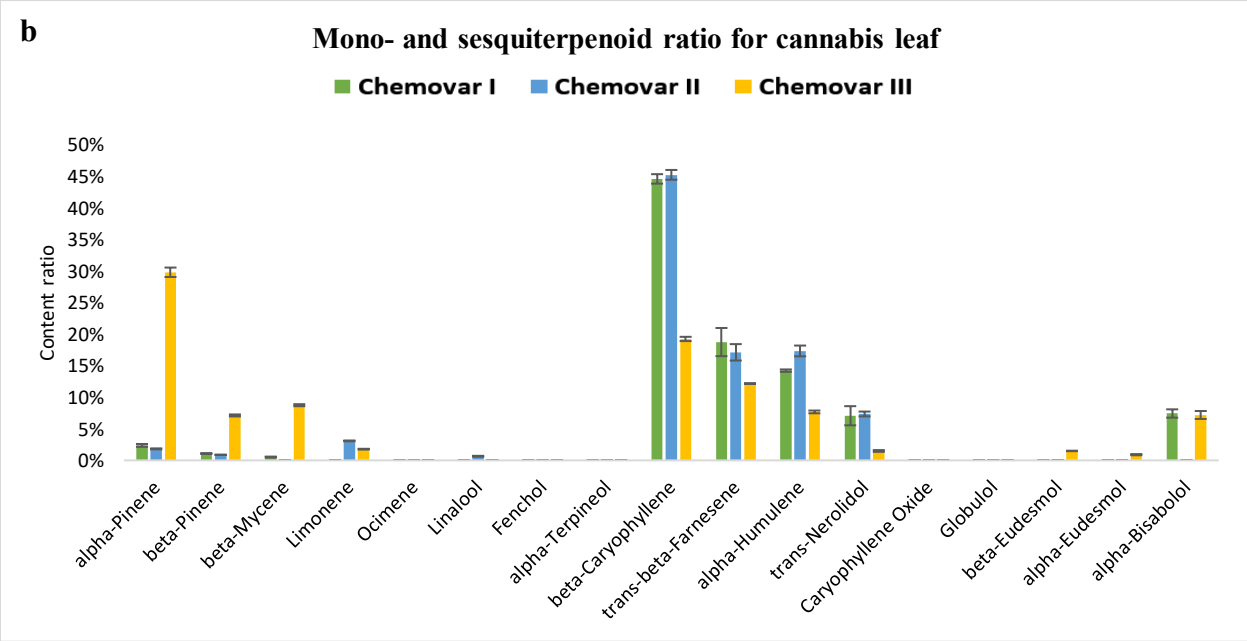


Figure 2.6 Mono- and sesquiterpenoid ratios in inflorescences and leaves in three strains.
(a) Mono- and sesquiterpenoid ratios in inflorescence. (b) Mono- and sesquiterpenoid ratios in leaves. The absolute values and ratios of individual terpenoids were consistent when compared among three strains, although total terpenoid content was significantly different.

For terpenoids whose analytical standards were unavailable for sourcing, identification was performed using its mass spectrum, and semi-quantification was performed using individual response area relative to the total response area of all terpenoid peaks using GC-FID, where the response factor was taken as one[127,144,155,156]. Several chemotaxonomic studies utilized this method to discriminate "Sativa" and "Indica" varieties and found that terpenoid profiles are uniquely retained from their respective landrace ancestors[67,70–73,83]. The presence of more hydroxylated terpenoids in Chemovar III does not fit its reported classification as *C. indica* ssp. *indica* (NLD, vernacular "Sativa"), but more closely aligns with *C. indica* ssp. *afghanica* (WLD, vernacular "Indica"). Similarly, although the Chemovar I and II were reported as "Indica," their terpenoid profiles were characteristic of "Sativa" chemovars. One study found that the reported ancestry percentages of "Sativa" vs. "Indica" for 81 drug-type chemovars are only moderately correlated with the calculated genetic structure[75], indicating that the vernacular classifications do not reliably communicate genetic identity. For medicinal research and applications, cannabis chemovars should be identified by their chemical fingerprints, which are more reliable than their names[67,71,72].

2.4.8 Flavonoid profile in inflorescences, leaves, stem bark, and roots

A total of twenty-six flavonoids have been identified in cannabis plants, which are methylated and prenylated aglycones or conjugated *O*-glycosides or *C*-glycosides of orientin, vitexin, isovitexin, quercetin, luteolin, kaempferol, and apigenin[85,92,157]. In this study, total flavonoid content was expressed as the sum of these seven flavonoids after acid hydrolysis. Flavonoids were not detected in roots, and stem bark, less detected in the inflorescence (0.07%-0.14%), and were highest in leaves (0.34%-0.44%) (**Figure 2.5(e)**). The total flavonoid in cannabis leaves is estimated to be around 1%[139], which matches with our result considering flavonoids exist as both free flavonoids (aglycones) and conjugated glycosides. Flavonoid content also varied between chemovars. Total flavonoid content in inflorescence was significantly higher in Chemovar III (0.14 ± 0.002%) than Chemovar I (0.07 ± 0.001%) and Chemovar II (0.010 ± 0.005%) ($n=3$, $p< 0.0001$) (**Figure 2.5(f)**). The total flavonoid content in leaves was higher in Chemovar II (0.44 ± 0.02%) and Chemovar III (0.40 ± 0.01%) than in Chemovar I (0.34 ± 0.02%) ($n=3$, $p=0.0043$). Vitexin was found to be the most abundant flavonoid, ranging from 0.12% to 0.17% in leaves and 0.02% to 0.06% in the inflorescence (**Figure 2.5(f-g)**, **Table 2.18**), consistent with the previous studies[85,158]. Orientin content ranged from 0.07% to 0.08% in leaves and 0.01% to 0.03% in

inflorescence in our samples, which are similar to results reported by Vanhoenacker[158], but are lower than results reported by Flores-Sanchez and Verpoorte[85]. The analyzed isovitexin and luteolin contents were lower than other studies[85,158]. Apigenin content ranged from 0.03% to 0.07% in leaves and 0.004% to 0.01% in inflorescence in our samples, which are similar to results reported by Vanhoenacker[158] but are lower than results reported by Flores-Sanchez and Verpoorte[85]. Neither quercetin nor kaempferol was found in leaf samples – these results are different from a previous study[85] that reported 0.2% quercetin in leaves. The inconsistency of reported values may be caused by differences in plant age and chemovar varieties. Unlike cannabinoid accumulation, individual and total flavonoid content decreases as the plants age[85]. Orientin, vitexin, and their glucosides were reported to have value in discriminating cannabis subspecies[157]. Cannflavin A and B are also notable flavonoids with medicinal potential identified in cannabis[159]. However, due to the unavailability of reference standards at the time, they were not included in this study.

Table 2.18 Flavonoid profile in cannabis

Compound	Orientin	Vitexin	Isovitexin	Quercetin	Luteolin	Kaempferol	Apigenin	Total flavonoids
Strain I leaf	0.08± 0.003%	0.12± 0.008%	0.01± 0.004%	ND	0.09± 0.008%	ND	0.05± 0.004%	0.34± 0.02%
Strain II leaf	0.07± 0.002%	0.17± 0.005%	0.02± 0.002%	ND	0.10± 0.01%	ND	0.07± 0.006%	0.44± 0.02%
Strain III leaf	0.07± 0.002%	0.13± 0.004%	0.05± 0.001%	ND	0.05± 0.002%	ND	0.03± 0.001%	0.40± 0.009%
Strain I inflorescence	0.01± 0.0003%	0.02± 0.0006%	0.002± 0.00005%	0.01± 0.0003%	0.01± 0.0004%	0.01± 0.0002%	0.004± 0.0002%	0.07± 0.001%
Strain II inflorescence	0.02± 0.001%	0.03± 0.002%	0.002± 0.0001%	0.01± 0.0004%	0.02± 0.001%	0.01± 0.0002%	0.01± 0.0004%	0.10± 0.005%
Strain III inflorescence	0.03± 0.0001%	0.06± 0.0008%	0.01± 0.0001%	0.01± 0.0008%	0.02± 0.0001%	0.005± 0.0003%	0.01± 0.0003%	0.14± 0.002%

* Content expressed in mean ± SD% (n=3). ND = Not detected.

2.4.9 Sterol profile in inflorescences, leaves, stem bark, and roots

Total sterol content was expressed as the sum of campesterol, stigmasterol, and β-sitosterol, increased from inflorescences, leaves, roots, to stem bark (**Figure 2.5(h)**). The ratio of three sterols was consistent with a previous study on cannabis roots[160]. β-sitosterol was the most abundant sterol in roots and stem bark for all three chemovars, ranging from 0.04 to 0.06% (**Figure 2.5(i-j), Table 2.19**). Campesterol content ranged between 0.01% to 0.02% in roots and stem bark and was not detected in leaf. Stigmasterol had the lowest concentration in roots and stem bark at 0.01% and

was most concentrated in leaves at 0.03%. Total sterols in stem bark were comparable between three chemovars (n=3, p = 0.0550) while they were significantly different in root material (n=3, p < 0.0001). Campesterol was not significantly different in the stem bark of three chemovars (n=3, p = 0.3523) but was significantly different (n=3, p< 0.0001) in root material. Stigmasterol was significantly different in stem bark in three chemovars (n=3, p = 0.0012) and in root material (n=3, p < 0.0001). β-sitosterol was significantly different in the roots of three chemovars (n=3, p<0.0001) but less variable in the stem bark of three chemovars (n=3, p = 0.1216).

Table 2.19 Sterol profile in cannabis

Compound	Campesterol	Stigmasterol	β-sitosterol	Total sterols
Strain I root	0.016 ± 0.001%	0.010 ± 0.001%	0.051 ± 0.004%	0.077 ± 0.005%
Strain II root	0.020 ± 0.001%	0.009 ± 0.0005%	0.058 ± 0.003%	0.088 ± 0.004%
Strain III root	0.010 ± 0.0004%	0.013 ± 0.001%	0.040 ± 0.001%	0.063 ± 0.002%
Strain I stem bark	0.016 ± 0.001%	0.011 ± 0.0003%	0.044 ± 0.002%	0.071 ± 0.003%
Strain II stem bark	0.016 ± 0.001%	0.008 ± 0.0004%	0.044 ± 0.003%	0.069 ± 0.004%
Strain III stem bark	0.017 ± 0.0005%	0.010 ± 0.001%	0.050 ± 0.007%	0.076 ± 0.006%
Strain I leaf	ND	0.026 ± 0.003%	0.027 ± 0.001%	0.053 ± 0.003%
Strain II leaf	ND	0.030 ± 0.002%	0.022 ± 0.0004%	0.052 ± 0.002%
Strain III leaf	ND	0.030 ± 0.001%	0.024 ± 0.0002%	0.053 ± 0.001%

* Content expressed in mean ± SD% (n=3). ND = Not detected.

2.4.10 Triterpenoids profile inflorescences, leaves, stem bark, and roots

Total triterpenoid content was expressed as the sum of β-amyrin, epifriedelanol, and friedelin. It increased from inflorescences (not detected), to leaves (<0.05%), stem bark (0.05-0.15%), and roots (0.1-0.3%) (**Figure 2.5(k)**). Total triterpenoid in both the roots and stem bark in Chemovar III was significantly higher than in Chemovar I and II (n=3, p<0.0001). Friedelin is the most prominent triterpenoid in cannabis and is concentrated in the stem bark and roots[80] (**Figure 2.5(l-m)**). It ranged from 0.083% to 0.135% in roots and 0.033% to 0.100% in stem bark (**Table 2.20**). The results were significantly higher than the 0.00128% (12.8mg/kg) reported in a previous study[160]. Epifriedelanol was found to range from 0.033% to 0.092% in roots and 0.013% to 0.041% in stem bark, which was higher than the 0.00213% (21.3mg/kg) previously reported[160]. Chemovar III had significantly higher friedelin, epifriedelanol, and β-amyrin in stem bark and roots than the other chemovars (n=3, p<0.0001). Neither friedelin nor epifriedelanol was found in leaf samples. Conversely, β-amyrin was found to be higher in leaf (0.012% to 0.026%) than in stem bark (0.006% to 0.007%) or root (0.005% to 0.013%).

Table 2.20 Triterpenoid profile in cannabis.

Compound	β-Amyrin	Epifriedelanol	Friedelin	Total triterpenoids
Strain I root	0.006 ± 0.0003%	0.043 ± 0.004%	0.083 ± 0.007%	0.132 ± 0.011%
Strain II root	0.005 ± 0.0001%	0.033 ± 0.001%	0.091 ± 0.004%	0.128 ± 0.005%
Strain III root	0.013 ± 0.001%	0.092 ± 0.003%	0.135 ± 0.003%	0.239 ± 0.006%
Strain I stem bark	0.006 ± 0.001%	0.013 ± 0.001%	0.033 ± 0.004%	0.052 ± 0.006%
Strain II stem bark	0.006 ± 0.001%	0.021 ± 0.002%	0.065 ± 0.007%	0.092 ± 0.010%
Strain III stem bark	0.007 ± 0.0004%	0.041 ± 0.002%	0.100 ± 0.004%	0.149 ± 0.007%
Strain I leaf	0.026 ± 0.002%	ND	ND	0.026 ± 0.002%
Strain II leaf	0.024 ± 0.001%	ND	ND	0.024 ± 0.001%
Strain III leaf	0.012 ± 0.001%	ND	ND	0.012 ± 0.001%

* Content expressed in mean ± SD% (n=3). ND = Not detected.

2.5 Discussion

To bridge traditional medicine and modern evidence-based medicine, biochemically active compounds must be identified, and their molecular mechanisms determined through preclinical and clinical studies. The secondary metabolites quantified in each part of cannabis are summarized in **Table 2.21**. Since concentrations above 0.05% are pharmacologically interesting[81], cannabis inflorescence and leaf material may contain sufficient cannabinoids, mono- and sesquiterpenoids, and flavonoids for therapeutic applications. For example, the leaves of Chemovar III contain 0.34% flavonoids in terms of total aglycones. In comparison, ginkgo leaves, which are used for ginkgo extract and are among the best sources of flavonoids, contains 0.4% total flavonoids in terms of total aglycones[161]. The stem bark and root are sources of triterpenoids and sterols. For example, friedelin is found in the leaves of *Azima tetracantha* Lam. (bee sting bush), containing a relatively high amount at 0.36%[162]. In comparison, dried cannabis roots and stem bark contain between 0.1% to 0.15%. Friedelin is also isolated from the dried leaves of *shorea robusta* (shala tree), which has been commonly used in traditional Indian medicine[163]. Cannabis contains more than ten times more friedelin than shala tree. The potential therapeutic properties of the identified compounds have been comprehensively reviewed[80,164]. For example, terpenoids and flavonoids identified in inflorescences and leaves have anti-inflammatory, anti-rheumatic, analgesic, anticonvulsant, antioxidant and neuroprotective, larvicidal, gastroprotective properties, and beneficial effects on the respiratory system[165,165–168]. Triterpenoids and sterols identified in stem bark and roots have anti-inflammatory, analgesic, antimicrobial, antioxidant, neuroprotective, angiogenic, anti-osteoarthritic, and estrogenic properties[169]. These secondary metabolites possess pharmaceutical values and may contribute to the overall therapeutic benefits of cannabis; however, these synergies

require further investigation to provide direct evidence. Such evidence should be derived from studies using cannabis material instead of proxy studies of extracts from other plants.

Table 2.21 Chemical profile of cannabis plant parts.

	Root	Stem bark	Leaf	Inflorescence
Total cannabinoids	-*	-	1.10% - 2.10%	15.77% - 20.37%
Total mono- and sesquiterpenoids	-	-	0.13% - 0.28%	1.28% - 2.14%
Total triterpenoids	0.13% - 0.24%	0.05% - 0.15%	-	-
Total sterols	0.06% - 0.09%	0.07 - 0.08%	0.05% - 0.05%	-
Total flavonoids	-	-	0.34% - 0.44%	0.07% - 0.14%

* Less than 0.05%.

In history, cannabis has been indicated for a wide range of conditions relating to pain, inflammation, and mental illness. For example, the inflorescences were used in traditional Chinese medicine for conditions including acute pain, mania, insomnia, coughing/panting, and wounds. The leaves were indicated for malaria, panting, roundworm, scorpion stings, hair loss, greying of hair. The stem bark were used for strangury and physical injury. The roots were used for strangury, spotting, vaginal discharge, difficult births, retention of the placenta, and physical injury[116,117]. Although the terminology in historical texts may be different from modern science and the nuances lost in the translation between Chinese and English, the uses of cannabis inflorescence indicated in ancient Chinese literature are comparable to those found in modern preclinical and clinical studies for cannabinoids[164,170–174]. But modern medicine has not fully developed the medical potential of cannabis leaf, stem bark, and root. Their traditional use may be used as a point of reference for clinical research. Similarly, the study of biomechanisms and the clinical effects of individual compounds may be consolidated for the development of applications using each plant part. For example, Δ^9-THC has antiemetic, and appetite stimulant properties and have been used to treat nausea or vomiting associated with chemotherapy and anorexia associated with AIDS-related weight loss by two approved medicine Marinol (dronabinol, synthetic Δ^9-THC) and Cesamet (nabilone, a THC-derivative)[174]. In addition, other previously unapplied cannabinoids have also been proven to have antiemetic and appetite stimulant properties, including CBDV[175], CBD[176], CBDA[177–181], and THCA[182], and may improve the therapeutic potential and reduce undesired side effects when used synergistically with other compounds in the plant material. Modern research also indicates that cannabis and cannabinoids have therapeutic potential for multiple sclerosis, Huntington's disease, Parkinson's disease, glaucoma, hypertension, stress and psychiatric disorders, Alzheimer's disease and dementia, and anti-neoplasia, many of which have

not been described in traditional use. Identification of bioactive compounds followed by well-designed clinical studies can convert each part of cannabis plant into evidence-based medicine.

2.6 Conclusions

Secondary metabolites, including cannabinoids, terpenoids, sterols, and flavonoids, were individually profiled in cannabis inflorescences, leaves, stem bark, and roots for three chemovars. Inflorescences and leaves are relatively abundant in cannabinoids, monoterpenoids, sesquiterpenoids, and flavonoids. Stem bark and roots contain triterpenoids and sterols. These bioactive compounds may underlie the traditional medicinal applications of each cannabis plant part in various cultures over thousands of years of cultivation. A comprehensive profile of bioactive compounds and thorough investigations of their synergistic interactions enables the correlation between plant compositions and therapeutic effects, ultimately bridging traditional herbal medicine with modern science. This approach enables the development of new cannabis-based medicine using all or subsets of plant parts, as opposed to the inflorescence only. One future trend for the cannabis industry is to fully utilize each part of cannabis by applying modern scientific methodologies for validating its traditional use.

Chapter 3 – Classification of Cannabis Strains in the Canadian market with Discriminant Analysis of Principal Components Using Genome-Wide Single Nucleotide Polymorphisms

3.1 Abstract

The cannabis community typically uses the terms "Sativa" and "Indica" to characterize drug strains with high tetrahydrocannabinol (THC) levels. Due to large scale, extensive, and unrecorded hybridization in the past 40 years, this vernacular naming convention has become unreliable and inadequate for identifying or selecting strains for clinical research and medicinal production. Additionally, cannabidiol (CBD) dominant strains and balanced strains (or intermediate strains, which have intermediate levels of THC and CBD), are not included in the current classification studies despite the increasing research interest in the therapeutic potential of CBD. This paper is the first in a series of studies proposing that a new classification system be established based on genome-wide variation and supplemented by data on secondary metabolites and morphological characteristics. This study performed a whole-genome sequencing of 23 cannabis strains marketed in Canada, aligned sequences to a reference genome, and, after filtering for minor allele frequency of 10%, identified 137,858 single nucleotide polymorphisms (SNPs). Discriminant analysis of principal components (DAPC) was applied to these SNPs and further identified 344 structural SNPs, which classified individual strains into five chemotype-aligned groups: one CBD dominant, one balanced, and three THC dominant clusters. These structural SNPs were all multiallelic and were predominantly tri-allelic (339/344). The largest portion of these SNPs (37%) occurred on the same chromosome containing genes for CBD acid synthases (CBDAS) and THC acid synthases (THCAS). The remainder (63%) were located on the other nine chromosomes. These results showed that the genetic differences between modern cannabis strains were at a whole-genome level and not limited to THC or CBD production. These SNPs contained enough genetic variation for classifying individual strains into corresponding chemotypes. In an effort to elucidate the confused genetic backgrounds of commercially available cannabis strains, this classification attempt investigated the utility of DAPC for classifying modern cannabis strains and for identifying structural SNPs.

(**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Classification of Cannabis strains in the Canadian market with Discriminant Analysis of Principal Components Using Genome-wide Single Nucleotide Polymorphisms. *PLOS ONE* 16, e0253387 (**2021**).

3.2 Introduction

Cannabis has a complex breeding history. Whether its botanical classification is monotypic (*sativa*) or polytypic (*sativa* and *indica*) remains controversial [27]. Since the 1980s, breeding for high psychoactive THC content has occurred very aggressively in North America [69]. Nearly all drug-type cannabis currently cultivated in the USA, Canada, and Europe are hybridized, resulting in thousands of strains [67]. Recent genetic studies focused on validating the vernacular classification of "Sativa" and "Indica" [74–77]. However, this terminology is inadequate for identifying or selecting strains for clinical research and medicinal production due to the misuse of the botanical nomenclature, extensive cross-breeding, and unreliable labelling during unrecorded hybridization [69]. One genetic study found that the reported ancestry percentage of "Sativa" vs. "Indica" for 81 drug stains is only moderately correlated with the calculated genetic structure ($r^2 = 0.36$) [75]. In addition, CBD dominant strains and balanced strains (THC ≈ CBD), which have gained increasing attention due to CBD's use as a therapeutic [79,183–186], have been omitted in recent classification studies.

Cannabis has a diploid genome (2n=20) with nine autosomal chromosomes and one pair of sex chromosomes [187]. The length of the haploid genome size is 818 Mbp for females and 843 Mbp for males [188]. An SNP is a variation of a single nucleotide at a specific position in the genome, and it is useful for understanding the genetic basis of diversity among populations [189]. SNPs are usually bi-allelic, with two alleles observed in the population [190]. Multiallelic SNPs have more than one alternative allele for that locus. Tri-allelic SNPs, which have three nucleotide substitution-based alleles at the same position, are relatively rare but are being considered of great relevance in epidemiological studies [191], in disaster victim identification using mixed and/or degraded DNA samples [192], and in animals pedigree accuracy studies [193]. Tri-allelic SNPs are reported to have a higher power of discrimination than bi-allelic SNPs requiring fewer markers and lowering costs [192,194]. However, tri-allelic SNPs have been excluded in cannabis population structural analysis in the current literature [76,195].

Cannabis classification studies that employ SNPs generally used partial genome information with few or no overlap sequences between datasets [86]. Whole-genome sequencing is used less often in the literature, but is preferable despite its higher cost because it enables comparison of genome datasets from different sources [86]. It also provides comprehensive genetic information [86], as studies

showed that differences between fiber- and drug-type cannabis are at a genome-wide level and not necessarily limited to genes involved in THC production [75]. The recent release of the 10-chromosome map of the cannabis genome [196–200] may improve the understanding of the genetic architecture, identify a superior set of SNPs associated with interesting traits, and reduce future targeted genotyping costs by using fewer but more accurate SNPs [201].

Several approaches are now available for the analysis of population genetic structure. One of these approaches is the DAPC, which is a multivariate clustering method that combines the merits of both principal component analysis (PCA) and discriminant analysis (DA) [77,87,202,203]. PCA is a multivariate analysis that can be applied to large datasets to reduce dimensions, but does not provide a group assessment, which is essential for investigating genetic structures of biological populations [84]. DA achieves the best classification of individuals into pre-defined groups by maximizing between-group variation and minimizing within-group variation, but the number of variables (alleles) needs to be fewer than the number of observations (individuals), which is generally not the case for SNP data [202]. DAPC first uses PCA to transform raw data (genome-wide identified SNPs) into principal components (PC), which are mutually orthogonal linear combinations of the original variables. This ensures that variables submitted to DA are perfectly uncorrelated and that there are fewer variables than number of individuals. Then, linear discriminant functions, which are synthetic variables of linear combinations of these SNPs, are constructed to maximize inter-cluster differences and minimize intra-cluster variation [202]. By combining the advantages of PCA and DA, DAPC can identify groups, assign individuals to groups, visualize between-population differentiation, and identify individual alleles that have contributed to population structuring.

The objectives of this study are to:

1. investigate whether modern cannabis strains can be classified and differentiated at the whole-genome level, and
2. investigate the chromosomal location and putative functions of identified structural SNPs.

This study is a part of an integrated cannabis strain classification project utilizing genetic, chemical, and morphological profiles, wherein plants were grown in a commercial greenhouse under the same condition.

3.3 Materials and methods

3.3.1 DNA extraction and whole genome sequencing

This study included 23 commercially available cannabis strains, and the research was carried out under a cannabis research license issued by Health Canada. Where possible, the reported ancestry ("Sativa", "Indica", or "Sativa-dominant" and "Indica-dominant") was obtained from the licensed producer providing the strain or from an online strain database (https://www.leafly.ca) (**Table 3.1**). Each strain was analyzed for chemical composition using methods established in our previous study [204] and labelled as "THC dominant", "balanced", or "CBD dominant". DNA was extracted from 100 mg of fresh leaves for each strain using a Qiagen DNeasy Plant Mini Kit (QIAGEN, Canada). DNA concentrations were determined using a Qubit Fluorometer (Thermo Fisher Scientific, US). DNA integrity was tested by agarose gel electrophoresis. Library construction and sequencing were performed by BGI (USA) using DNBseq™ sequencing technology to a depth of 30x. DNBseq™ is a high-throughput sequencing solution, where DNA is fragmented into 100-300 bp and made into DNA nanoballs (DNB™), which are continuous DNA molecule with multiple head-to-tail copies of the same DNA fragment by linear isothermal rolling-circle replication. They are loaded onto high-density sequencing templates and sequenced by combinatorial probe-anchor synthesis (cPAS), where fluorescently tagged nucleotides complete for addition to the growing chain. After the addition of each nucleotide, high-resolution digital imaging is carried out where the DNB clusters are excited by a light source and a characteristic fluorescent signal is emitted. Hundreds of and thousands of clusters are sequenced in a massively parallel process. The emission wavelength, along with the signal intensity, determines the base call and the number of the cycles determines the length of the read. Sequence reads were then aligned to the reference genome assembly ASM23057v4 of a drug type strain Purple Kush (PK) in the NCBI BioProject database under accession number PRJNA73819 [205] using Burrows-Wheeler Alignment (BWA) tool [206]. New assignments of chromosomes numbers (1-10) were used as in ASM23057v5 [207]. The first step of SNP calling is marking duplications in BAM format files, and selected duplications are included in SNP calling by GATK (Genome Analysis Toolkit) (https://www.broadinstitute.org/gatk/). Local realignment around inDels is performed to avoid the bias of SNP calling, and the variation sites around inDel are identified as SNPs. A total of 235,334 SNPs was identified, including 225,046 bi-allelic and 10,288 multiallelic SNPs. After filtering for

SNPs with no missingness by locus and a minor allele frequency less than 10% using VCFtools, 137,858 SNPs, including 128,810 bi-allelic and 9,048 multiallelic SNPs, remained for analysis.

Table 3.1 Strain information of 23 strains and preassigned clusters by DAPC

Strain number	Strain name	Chemotypes	Clusters (W-SNPs)	Clusters (I-SNPs)	"Sativa" or "Indica"
1	Lemon Garlic OG	1-Balanced	C1	C4	"Indica" dominant
2	Royal Medic	2-Balanced	C3	C2	"Sativa" dominant
3	Blue Hawaiian	3-CBD	C3	C1	"Sativa" dominant
4	Kandy Kush	4-CBD	C3	C1	"Sativa" dominant
5	Special	5-CBD	C3	C1	Not provided
6	NN	6-CBD	C3	C1	Not provided
7	Dance World	7-Balanced	C3	C2	"Sativa" dominant
8	Treat	8-CBD	C3	C1	Not provided
9	High	9-Balanced	C3	C2	Not provided
10	CB7	10-CBD	C3	C1	Not provided
11	33°	11-THC	C1	C4	Not provided
12	Banana Cake	12-THC	C2	C5	"Indica" dominant
13	Bananium	13-THC	C3	C3	"Indica" dominant
14	Burmese Blueberry	14-THC	C2	C5	"Indica" dominant
15	Divine Banana	15-THC	C2	C4	"Indica" dominant
16	Granddaddy Purple	16-THC	C2	C5	"Indica" dominant
17	Lemon Love	17-THC	C1	C5	"Indica" dominant
18	Lemon Sorbet	18-THC	C1	C4	"Indica" dominant
19	MeatHead	19-THC	C2	C5	"Indica" dominant
20	Nanitro	20-THC	C1	C4	"Indica" dominant
21	Platinum Jelly Punch	21-THC	C1	C4	"Indica" dominant
22	SBSK2 (Lemon Thai)	22-THC	C3	C3	50/50 hybrid
23	Super sherbet	23-THC	C1	C4	"Indica" dominant

*The column of clusters W-SNPs was obtained using the whole set of 137,858 filtered SNPs. The column of clusters I-SNPs was obtained using 344 structural SNPs.

3.3.2 Analysis of population structure and identification of structural SNPs

The population structure in this work was analyzed by DAPC using the *adegenet* package [208] in R software [209]. First, the *find.clusters* function ran successive K-means [210] for a range of *k* values (where the number of clusters *k* = K), and identified the optimal number of clusters by comparing the Bayesian Information Criterion (BIC) [211] of the corresponding models. After groups were assigned, a cross-validation function (*xvalDapc*) was used to determine the optimal number of PCs to avoid over-sacrificing information or over-fitting in the subsequent DAPC. In cross-validation, the data were divided into a training set (90% of the data) and a validation set (10% of the data) by default. DAPC was carried out on the training set and the accuracy of predicting the membership of individuals in the validation set was used to identify the number of PCs. The

sampling and DAPC were repeated 30 times by default at each level of PC retention. After assigning individuals to clusters, DA was carried out on the retained PCs and contributions of the alleles to each discriminant function were stored. An SNPZIP analysis (*snpzip*) in R was then used to provide objective delineation between structural and non-structural SNPs, as identified by DAPC, to determine which SNPs contribute significantly to the between-population structure [212].

First, the whole set of 137,858 SNPs were applied to DAPC to identify SNPs that contributed most to the identified clusters. DAPC was carried out again using the identified SNPs to validate their differentiation efficiency by confirming the separation of the 23 strains into their preassigned clusters. A short sequence (about 600 nt) around each one of these identified SNP was searched using the BLAST software (https://blast.ncbi.nlm.nih.gov) against *Cannabis sativa* Annotation Release 100 [213]. In addition to DAPC, other clustering methods, including PCA, neighbor-joining (NJ) tree [214], and hierarchical dendrogram using Ward's minimum variance method [215], were also employed to assess the robustness of the final inferred clusters. PCA and NJ tree were plotted using R. The hierarchical dendrogram was plotted using JMP 14.0.0.

3.4 Results and discussions

3.4.1 Discriminant analysis of principal components using 137,858 SNPs

As indicated by the elbow in the curve of BIC values as a function of *k* in **Figure 3.1(a)**, the optimal number of identified clusters was three, corresponding to the lowest BIC values. The number of PCs retained for DAPC analysis was four, as calculated by cross-validation in **Figure 3.1(b)**, where it had 100% predictive success, and 0% associated root mean squared error (RMSE). In this study, the number of PCs associated with the highest mean success was also associated with the lowest MSE, which made it easier to choose the number of PCs to retain. For the subsequent DAPC analysis, four PCs and two discriminant functions were retained. The DAPC plot of 23 cannabis genotypes is shown in **Figure 3.1(c)**. The grouping assignment for individual strains by DAPC is listed in **Table 3.1 (**as W-SNPs). C1 is a THC dominant cluster and includes six THC dominant strains (11, 17, 18, 20, 21, and 23-THC) and one balanced strain (1-balanced). C2 is another THC dominant cluster and includes five THC dominant strains (12, 14, 15, 16, and 19-THC). C3 is a cluster dominated by CBD dominant and the balanced strains which includes six CBD dominant strains (3, 4, 5, 6, 8, and 10-CBD), three balanced strains (2, 7, and 9-balanced), and two THC dominant strains (13 and 22-THC). While C2 is closer to C3 and is more distant to

C1, C1and C3 are clearly separated along linear discriminant 1 (LD1). While C1 and C3 are roughly at the same level with respect to linear discriminant 2 (LD2), C2 is separated from both. PCA was also carried out on the same set of SNPs and results are shown in **Figure 3.2**. Twenty-three cannabis strains are plotted along pair-wise PCs of the first 4 PCs, which account for 18.4%, 11.5%, 9.5%, and 8.7% of the total variance, respectively. Similarly, the first PC suggests the existence of a relatively compact CBD & balanced clade on the left side of the plot and a more dispersed THC dominant clade on the right side of the plot. Balanced strains share a closer gene pool with CBD dominant strains, while the THC gene pool is more dispersed. Because THC is psychoactive and its potency can be readily assessed through consumption, selection for increasing THC content started early and widely for recreational purposes by traditional breeding [216]. In contrast, CBD is non-psychoactive and must be analyzed in a laboratory for potency, and therefore breeding for high CBD concentrations began later [216]. A complete genome assembly implied that CBD dominant varieties were generated by integrating hemp-type CBD acid synthase gene clusters into a background of drug-type cannabis to elevate CBDA production [197]. These balanced strains may have been created by crossing purebred THC dominant types with CBD dominant types [42]. Therefore, there may be a relatively limited selection of CBD dominant strains for breeding balanced strains.

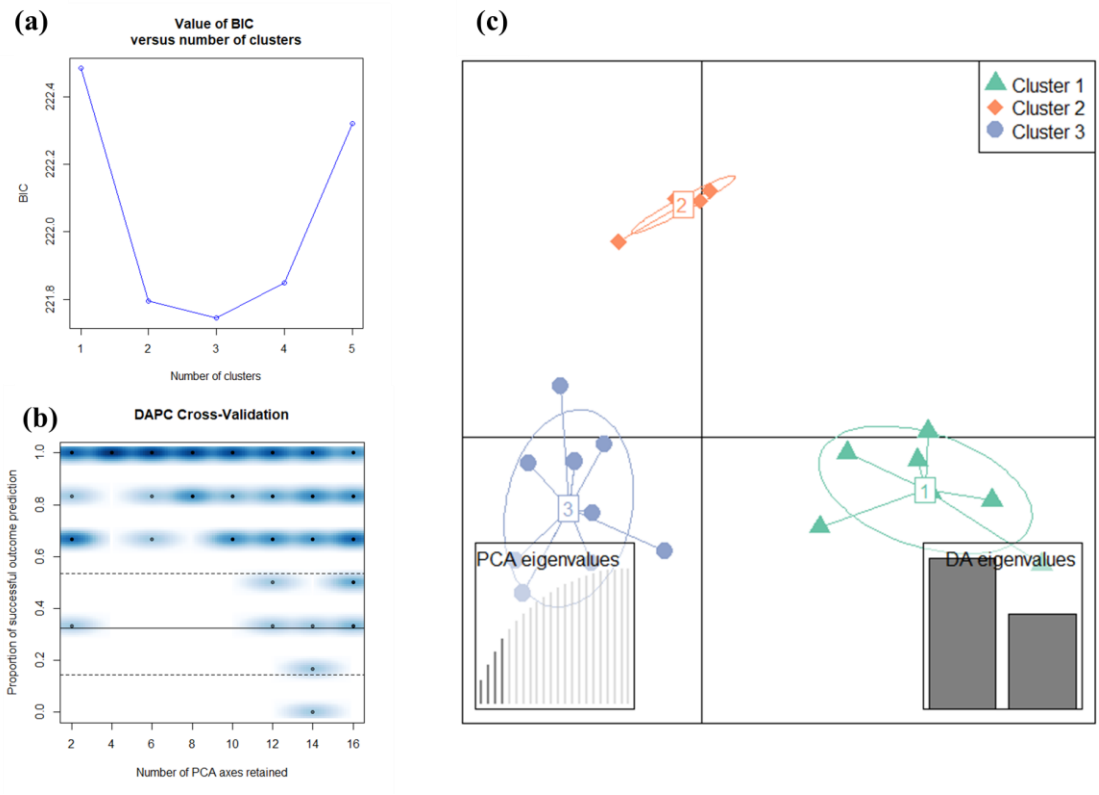
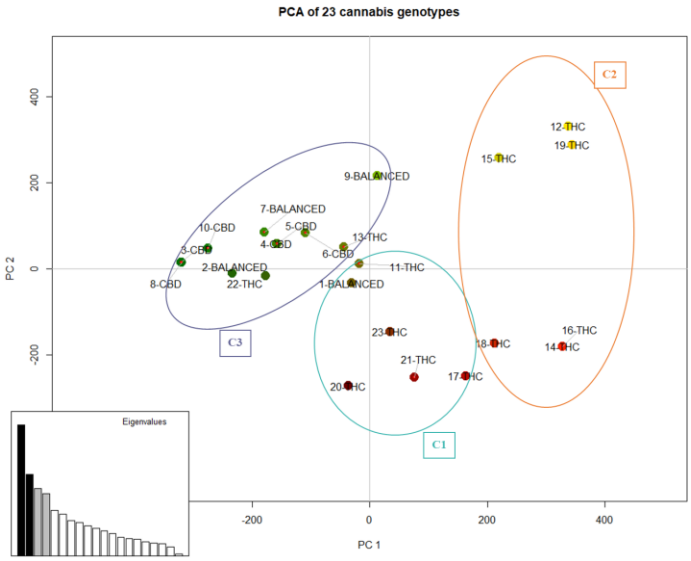
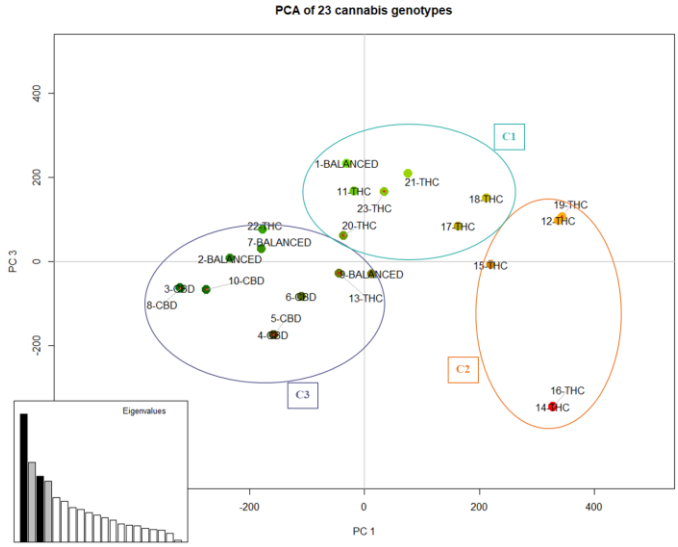


Figure 3.1 DAPC for 23 cannabis genotypes using 137,858 SNPs.
(a) The x-axis is the number of clusters *k* and the y-axis is the corresponding value of BIC. (b) The plot of DAPC cross-validation. The x-axis is the number of PCA axes retained for DAPC, and the y-axis is the proportion of successful outcome prediction. Individual replicates appear as points, and the density of those points in different regions of the plot is displayed in blue. (c) DAPC plot for 23 cannabis genotypes along two linear discriminants (LD 1 and LD 2).

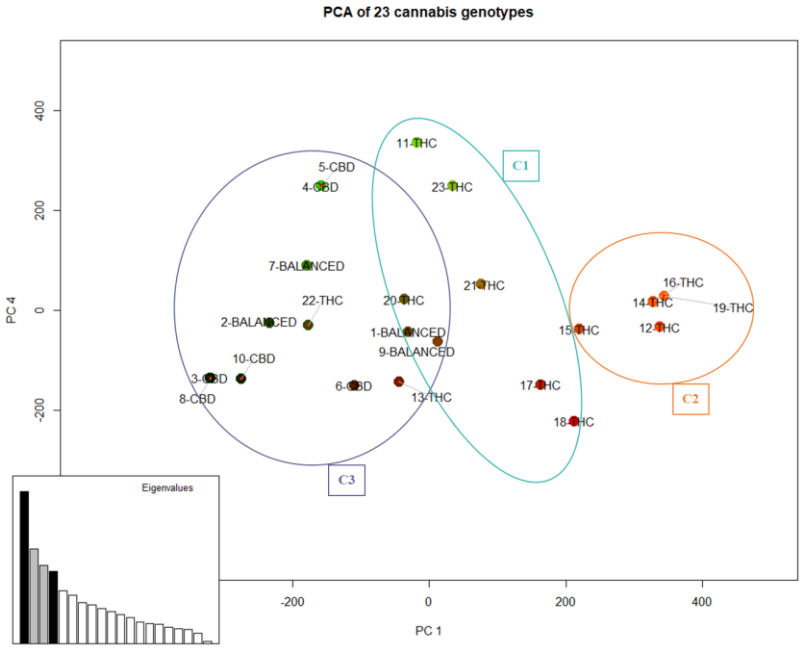
(a) P1&P2


PCA of 23 cannabis genotypes
C2
C3
C1
12-THC
19-THC
15-THC
9-BALANCED
7-BALANCED
10-CBD
5-CBD
13-THC
3-CBD
4-CBD
6-CBD
11-THC
2-BALANCED
8-CBD
22-THC
1-BALANCED
23-THC
16-THC
18-THC
14-THC
21-THC
17-THC
20-THC
Eigenvalues
PC 2
PC 1
400
200
0
-200
-200
0
200
400

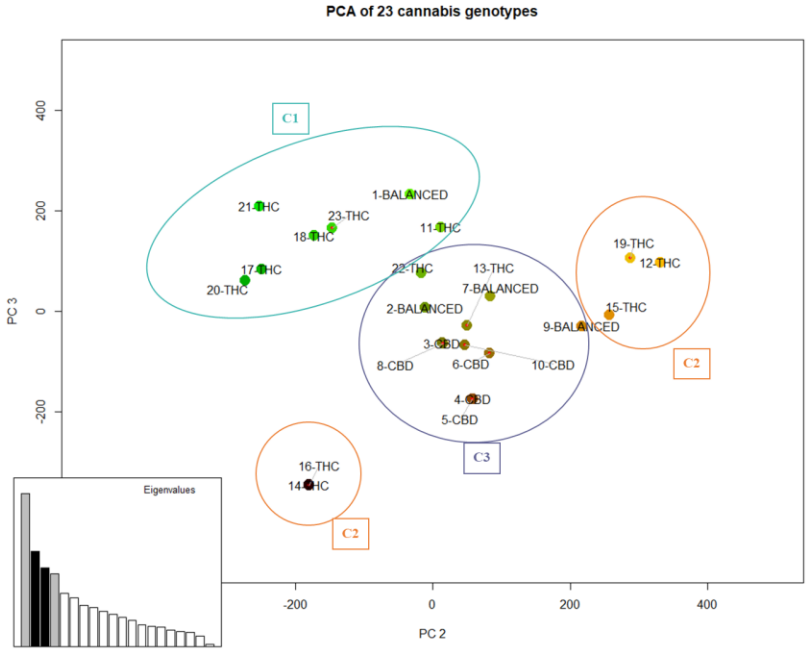
(b) P1&P3


PCA of 23 cannabis genotypes
C1
C2
C3
1-BALANCED
21-THC
11-THC
18-THC
23-THC
19-THC
12-THC
22-THC
20-THC
17-THC
7-BALANCED
2-BALANCED
15-THC
9-BALANCED
10-CBD
3-CBD
6-CBD
13-THC
8-CBD
5-CBD
4-CBD
16-THC
14-THC
Eigenvalues
PC 3
PC 1
400
200
0
-200
-200
0
200
400

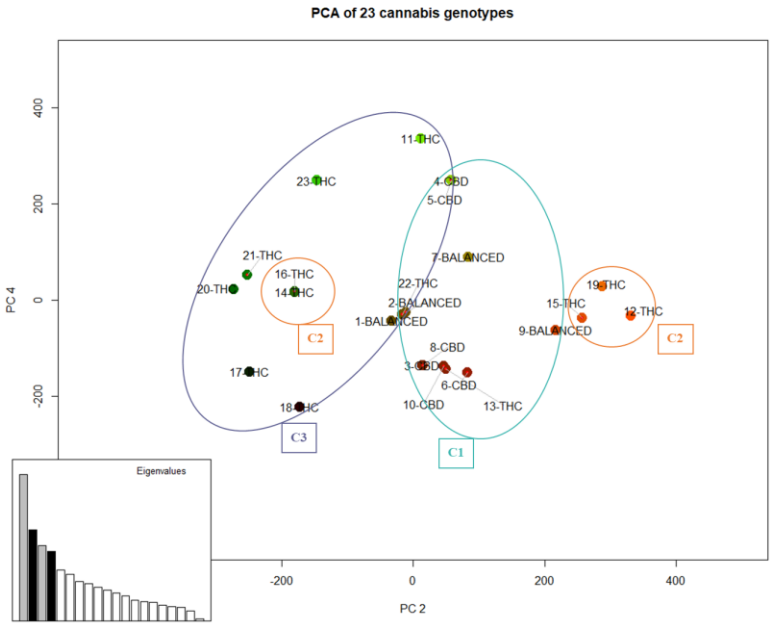
(c) P1&P4


PCA of 23 cannabis genotypes
PC 4
PC 1
400
200
0
-200
-200
0
200
400
11-THC
5-CBD
4-CBD
23-THC
C1
7-BALANCED
21-THC
16-THC
22-THC
20-THC
14-THC
19-THC
2-BALANCED
1-BALANCED
15-THC
12-THC
9-BALANCED
10-CBD
3-CBD
6-CBD
17-THC
8-CBD
13-THC
C2
18-THC
C3
Eigenvalues

(d) P2&P3


PCA of 23 cannabis genotypes
PC 3
PC 2
400
200
0
-200
-200
0
200
400
C1
1-BALANCED
21-THC
23-THC
11-THC
18-THC
19-THC
12-THC
17-THC
22-THC
13-THC
20-THC
7-BALANCED
2-BALANCED
15-THC
9-BALANCED
3-CBD
C2
8-CBD
6-CBD
10-CBD
4-CBD
5-CBD
C3
16-THC
14-THC
C2
Eigenvalues

(e) P2&P4



(f) P3&P4

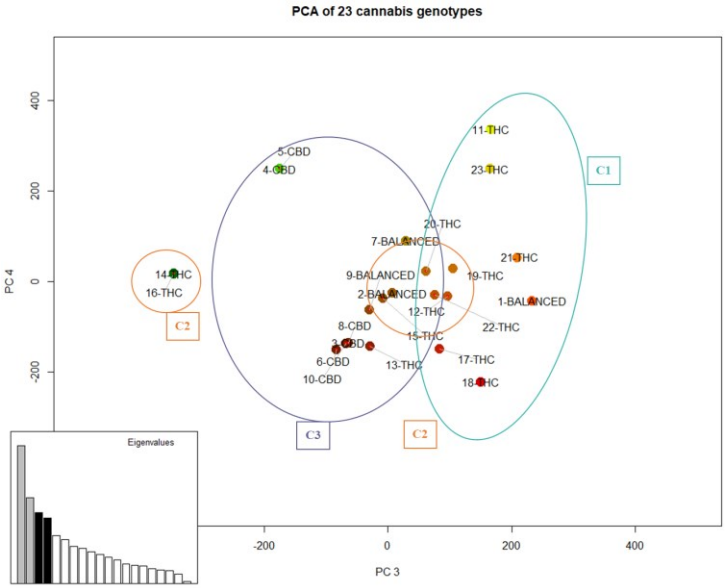


Figure 3.2 PCA of 23 strains using whole set of SNPs
Scatter plot of 23 cannabis strains on (a) PC1 & PC2, (b) PC1 & PC3, (c) PC1 & PC4 (d) PC2 & PC3 (e) PC2 & PC4 (f) PC3 & PC4 using 137,858 SNPs. Clusters indicated as C1, C2, and C3 correspond to W-SNPs in **Table 3.1**.

3.4.2 Discriminant analysis of principal components using 344 structural SNPs

DAPC was repeated using identified 344 structural SNPs. The optimal number of identified clusters was five, corresponding to the lowest BIC values **(Figure 3.3(a))**. Two PCs were retained for the following DAPC analysis in **Figure 3.3(b)**, where it had 98.9% predictive success and 0.04% RMSE. For the subsequent DAPC analysis, two PCs and two discriminant functions were retained. The grouping assignment for individual strains by DAPC is listed in Table 3.1 (as I-SNPs). Within the five clusters (**Figure 3.3(c)**), C1 is a CBD dominant cluster that includes six strains (3, 4, 5, 6, 8, and 10-CBD), C2 includes three balanced strains (2, 7, and 9-balanced), and C3, C4, and C5 are THC dominant clusters that include two (13 and 22-THC), seven (1-balanced, 11, 15, 18, 20, 21, 23-THC), and five (12, 14, 16, 17, and 19-THC) strains, respectively.

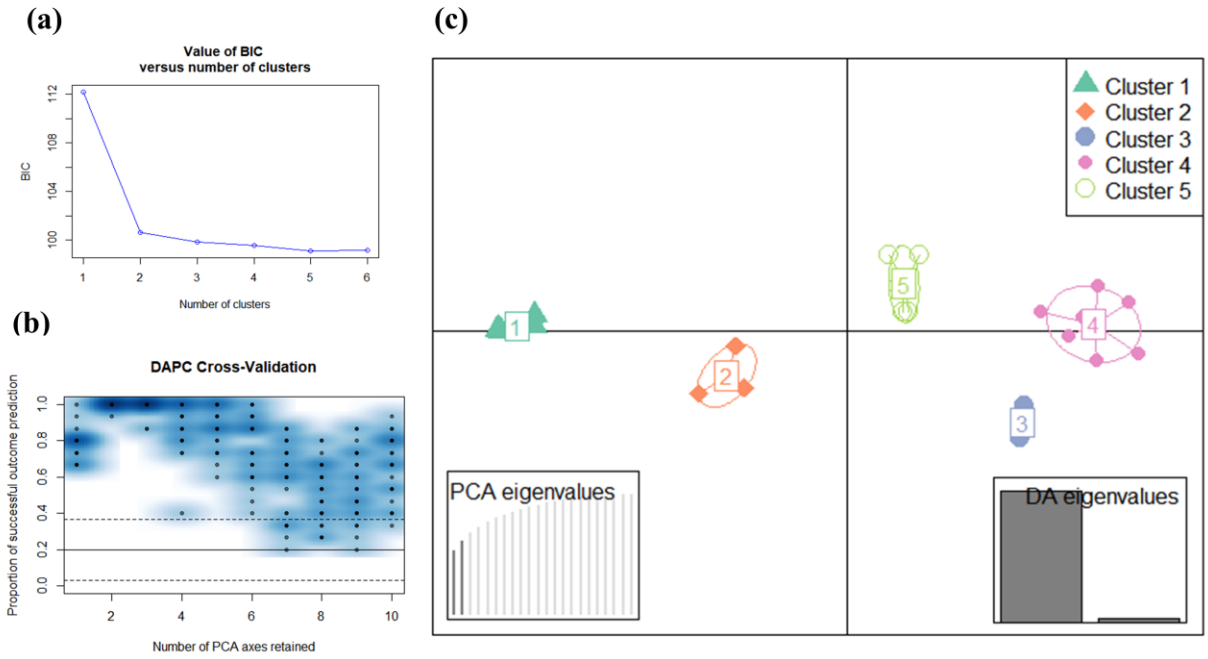


Figure 3.3 DAPC of 23 cannabis genotypes using 344 multiallelic structural SNPs.
Clusters indicated as C1, C2, C3, C4, and C5 corresponds to the I-SNPs in **Table 3.1**.

These multiallelic SNPs were also subjected to PCA, NJ tree, and hierarchical clustering analysis. In **Figure 3.4**, the 23 cannabis strains are plotted along PC1 and PC2, which account for 44.5% and 10.0% of the total variance, respectively. The proportions of explained variance are higher compared to the previous PCA results (18.4% and 11.5%) obtained using the whole set of SNPs. CBD dominant cluster C1 and balanced cluster C2 are on the left side of the scatter plot (PC1<0)

and the THC dominant clusters C3, C4, and C5 are on the right side of the scatter plot (PC1>0). Notably, six CBD dominant strains are separated from three balanced strains, while they were previously combined in the analysis using the whole set of SNPs. In addition, two THC dominant strains 13-THC and 22-THC are separated from the CBD and balanced cluster, and instead placed closer to other THC dominant strains. Strain 1-balanced is closer to THC dominant strain regardless of whether the whole set of SNPs or 344 identified SNPs were used.

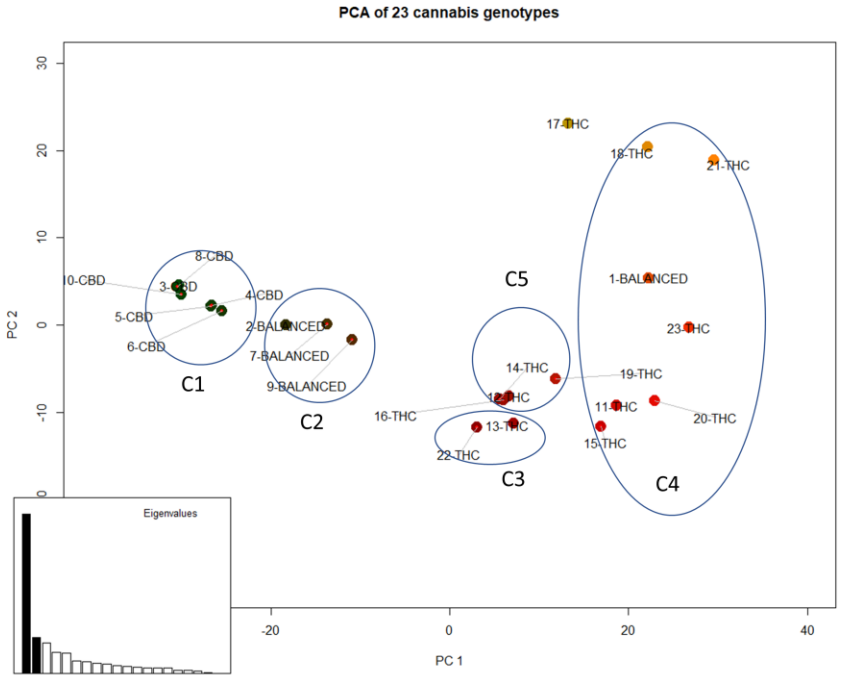


Figure 3.4 Scatter plot of 23 cannabis strains on PC1 & PC2 using 344 structural SNPs.
Clusters indicated as C1, C2, C3, C4 and C5 correspond to I-SNPs in **Table 3.1**.

The genetic structure from NJ-tree and hierarchical clustering using the 344 multiallelic are displayed in **Figure 3.5**, mostly congruent with that of DAPC. In the NJ-tree, all six CBD dominant strains are clustered together, with three balanced strains clustered closer on the same branch (**Figure 3.5 (a)**). Most THC dominant strains are also clustered adjacent to strains within their own clusters. The dendrogram using hierarchical clustering by Ward's method reveals two major groups, where one group is comprised of CBD dominant & balanced strains, and the other of THC dominant strains (**Figure 3.5 (b)**). They are further separated into five subclusters, where CBD dominant and balanced clusters are consistent with the DAPC grouping results, and several THC dominant strains clustered differently. Two strains, 15-THC and 18-THC, were assigned to C4

using DAPC but are assigned closer to C5 in the dendrogram. Two other strains, 14-THC and 16-THC, were assigned to C5 in DAPC but are assigned closer to C3 in the dendrogram. The clustering results are congruent between DAPC and hierarchical clustering with an assignment agreement rate of 83% (19/23).

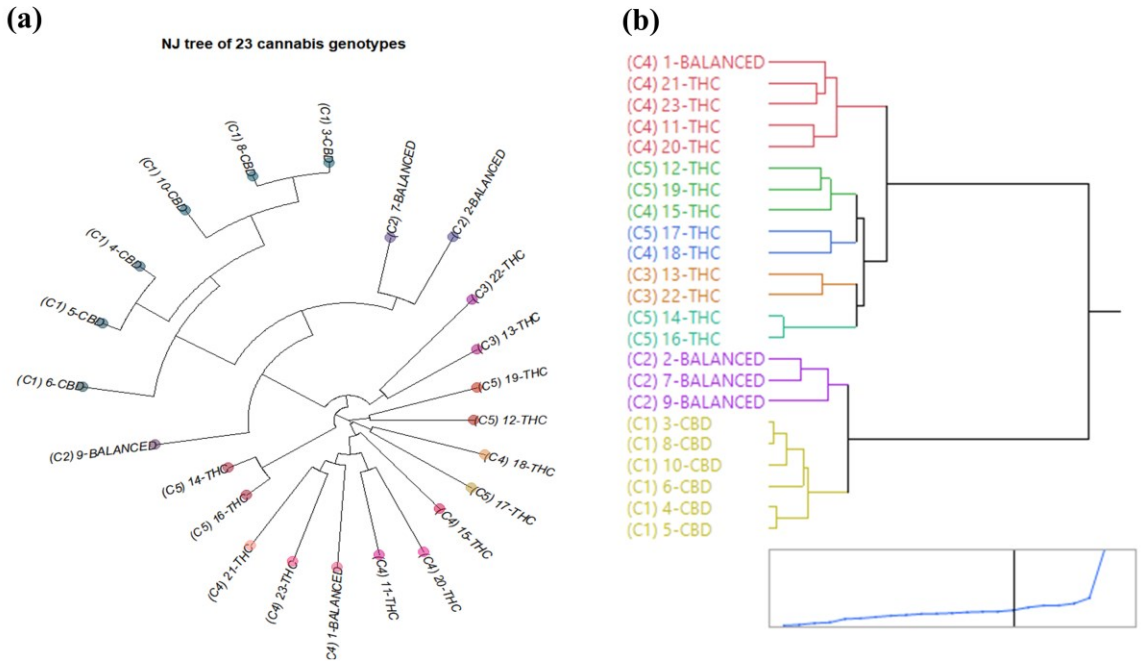


Figure 3.5 NJ-tree and hierarchical clustering using the 344 multiallelic SNPs.
(a) NJ-tree and (b) The dendrogram using hierarchical clustering by Ward's method for 23 cannabis genotypes. Clusters indicated as C1, C2, C3, C4, and C5 corresponds to I-SNPs in **Table 3.1**.

3.4.3 Allele frequencies for 344 multiallelic SNPs in three chemotypes

DAPC identified 344 highly contributing SNPs (**Supplementary Table 3.1**). All the structural SNPs are multiallelic, among which 98.5% (339/344) are tri-allelic and the remainder 1.5% (5/344) are tetra-allelic. The dendrogram of 23 strains using hierarchical clustering based on the allele counts in the 344 structural SNPs (**Supplementary Table 3.2**) separated the strains into CBD dominant, balanced, and THC dominant strains, mostly corresponding to the grouping results of DAPC **(Figure 3.6)**. The allele frequency was calculated by dividing the counts of that allele for all strains within the targeted group by the sum of the counts for each allele for that SNP within the targeted group. Allele frequencies of the structural SNPs were calculated for three major branches, each corresponding one of three chemotypes (**Supplementary Table 3.1**). If 1-balanced

strain was assigned to the THC dominant group as indicated by DAPC for allele frequency calculation, there are 87% (300/344) SNPs in CBD dominant clusters, 46% (157/344) SNPs in balanced clusters, and 11% (39/344) SNPs in THC dominant clusters that have one allele with allele frequencies > 80% (**Supplementary Table 3.1_Sheet1**). Among them, 140 SNPs shared same alleles with allele frequencies > 80% in CBD dominant strains (140/300) and balanced strains (140/157), which further indicated that CBD dominant strains and balanced strains closely share a gene pool. There are 38 SNPs that have one allele present in CBD dominant strains with allele frequencies > 80% and are not detected in THC dominant strains. There are 322 SNPs whose alleles that are present in THC dominant strains but were not detected in CBD dominant strains.

If the 1-balanced strain is assigned to the balanced group for allele frequency calculation, there are 87% (300/344) SNPs in CBD dominant clusters, 10% (36/344) SNPs in balanced clusters, and 13% (44/344) SNPs in THC dominant clusters that have one allele with allele frequencies > 80% (**Supplementary Table 3.1_Sheet2**). Among them, 32 SNPs shared same alleles with allele frequencies > 80% in CBD dominant strains (32/300) and balanced strains (32/36). There are 38 SNPs that have one allele present in CBD dominant strains with allele frequencies > 80% and are not detected in THC dominant strains. There are 321 SNPs whose alleles are present in THC dominant strains but were not detected in CBD dominant strains. Assigning the 1-balanced strain to the balanced group added more genetic diversity to the balanced group, and the effect of adding or deleting this strain for the THC dominant group in terms of allele frequency is small and can be neglected.

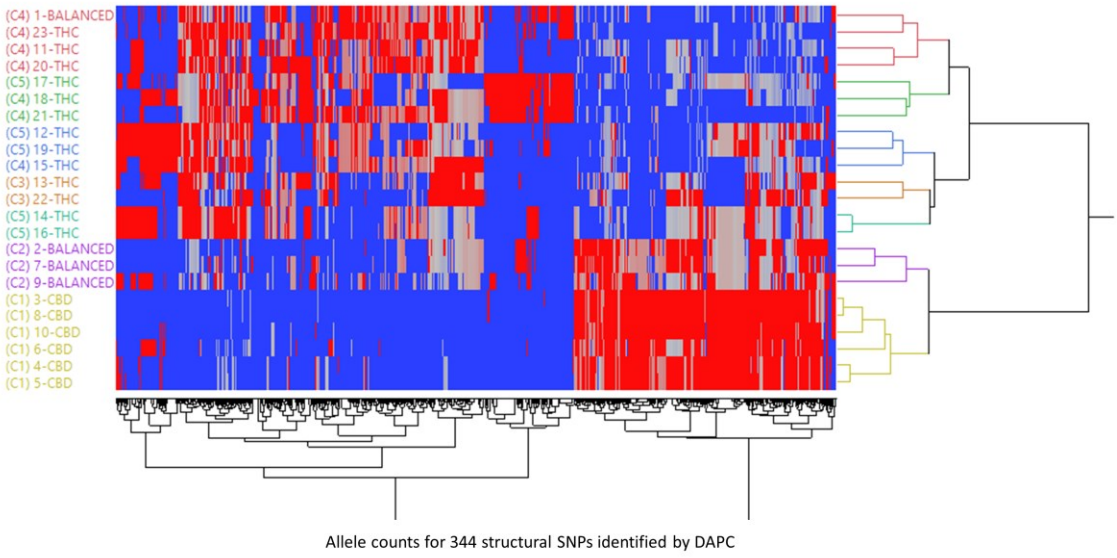


Figure 3.6 Hierarchical clustering of 23 strains based on the allele counts for 344 structural SNPs.

3.4.4 BLAST analysis of 344 multiallelic SNPs

These 344 SNPs were spread across all 10 chromosomes (**Figure 3.7 (a)**), indicating that commercially available cannabis strains in North America are significantly differentiated at a genome-wide level. The number of identified SNPs ranged from 7 to 127 on each genome, with 37% of the genetic variation occurring (127 SNPs) on chromosome 6, where CBDAS and THCAS are located [187]. The rest SNPs were spread over the remaining nine chromosomes. All ten chromosomes have genes related to the biochemical pathways of secondary metabolites, including cannabinoids, monoterpenoids, and sesquiterpenoids [119,120,187,197,217–219]. BLAST results showed that 90% (310/344) of these structural SNPs had no feature, 7% (24/344) are uncharacterized loci with unknown functions, and 3% (10/344) are predicted for certain functions **(Figure 3.7 (b)**).

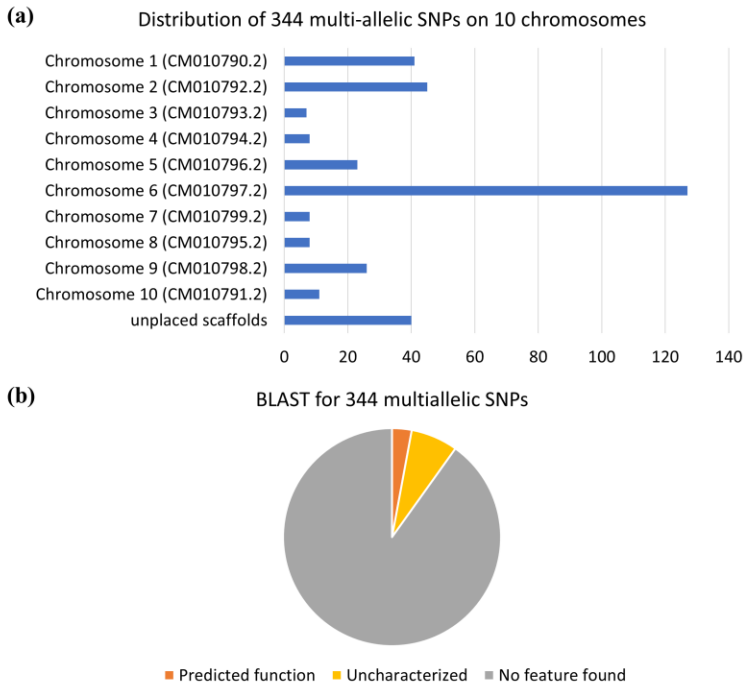


Figure 3.7 Features of 344 multiallelic SNPs.
(a) Distribution of structural SNPs on chromosome 1-10 and unplaced scaffolds. (b) BLAST results for structural SNPs against a fully annotated genome.

3.5 Conclusions

Although the cannabis industry is rapidly advancing after the relaxation of legal restrictions in North America, the increasing number of THC dominant strains, CBD dominant strains, and balanced strains only adds confusion to the currently poorly understood genetic background of the thousands of varieties already in existence. Although there were only 23 strains included in this study, they covered the three typical chemotypes of cannabis strains currently available in the market. Leveraging as much genetic variation as possible using whole-genome sequencing, we identified 344 multiallelic SNPs that were used to investigate the genetic structure of 23 cannabis genotypes using DAPC, PCA, NJ tree, and hierarchical clustering, which provided consistent observations and groupings despite the differences in algorithms. The clustering results revealed that these 23 strains could be separated into five clusters, with one cluster containing six CBD dominant strains, another cluster containing three balanced strains, and the remaining three clusters containing 13 THC dominant strains and one balanced strain. CBD dominant strains and the balanced strains are closer genetically. This may be attributed to how medical interest in breeding

for non-psychoactive, CBD-elevated strains (CBD dominant and balanced strains) has only recently been in vogue, resulting in an overlapping and less diverse gene pool for CBD dominant and balanced strains compared to the longer breeding history for THC strains. Some alleles are only present in CBD dominant strains or in THC dominant strains. More alleles present in balanced strains are shared with CBD dominant strains. One third of these structural SNPs are located on the chromosome containing THCAS and CBDAS. The remaining SNPs are located on the other nine chromosomes. An area of potential investigation is how the identified structural SNPs are associated with the production of other cannabinoids, mono- and sesquiterpenoids, flavonoids, other compounds, or morphological characteristics.

Since the late 20th century, genetic methodologies have been developed for separating industrial hemp from drug-type cannabis for forensic purposes, thus differentiating CBD dominant and THC dominant strains [43,46,52,56,220]. For the past 20 years, with the extensive hybridization of THC dominant strains, many classification studies have focused on separating "Sativa" and "Indica" strains and many have suggested abolishing this vernacular [75–77]. The genotyping results of this study indicate that modern, extensively hybridized strains can still be separated using genome-wide information. As a powerful multivariate approach that investigates population structures based solely on genetic information, DAPC separated strains into clusters aligned with their chemotypes. Additionally, DAPC has the potential to sort the disordered genetic background of thousands of THC dominant strains by identifying the number of genetic clusters within THC dominant strains, describing clusters by interpreting group memberships, and identifying the contributing SNPs that have the potential to be used as genetic markers for strain classification and identification. This would require a concerted effort from the cannabis industry by contributing whole genome sequence data to public databases and by building a common taxonomy based on genomics. Optimally, the identified genetic markers can be used as genomic fingerprints in combination with chemical fingerprints and morphological characteristics for strain identification. These markers can be leveraged for strain selection in clinical trials and for manufacturing cannabis-based products and medicines.

Supplementary Table 3.1 344 multiallelic SNPs identified by DAPC

(Deposited in ERA: https://doi.org/10.7939/r3-eq07-rp47)

Supplementary Table 3.2 Allele counts for 344 structural SNPs identified by DAPC

(Deposited in ERA: https://doi.org/10.7939/r3-eq07-rp47)

Chapter 4 – Identification of Chemotypic Markers in Three Chemotype Categories of Cannabis Using Secondary Metabolites Profiled in Inflorescences, Leaves, Stem Bark, and Roots

4.1 Abstract

Previous chemotaxonomic studies of cannabis only focused on tetrahydrocannabinol (THC) dominant strains while excluded the cannabidiol (CBD) dominant strains and intermediate strains (THC ≈ CBD). This study investigated the utility of the full spectrum of secondary metabolites in different plant parts in three cannabis chemotypes (THC dominant, intermediate, and CBD dominant) for chemotaxonomic discrimination. Hierarchical clustering, principal component analysis (PCA), and canonical correlation analysis assigned 21 cannabis varieties into three chemotypes using the content and ratio of cannabinoids, terpenoids, flavonoids, sterols, and triterpenoids across inflorescences, leaves, stem bark, and roots. The same clustering results were obtained using secondary metabolites, omitting THC and CBD. Significant chemical differences were identified in these three chemotypes. Cannabinoids, terpenoids, flavonoids had differentiation power while sterols and triterpenoids had none. CBD dominant strains had higher amounts of total CBD, cannabidivarin (CBDV), cannabichromene (CBC), α-pinene, β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, α-bisabolol, orientin, vitexin, and isovitexin, while THC dominant strains had higher total THC, total tetrahydrocannabivarin (THCV), total cannabigerol (CBG), camphene, limonene, ocimene, sabinene hydrate, terpinolene, linalool, fenchol, α-terpineol, β-caryophyllene, trans-β-farnesene, α-humulene, trans-nerolidol, quercetin, and kaempferol. Compound levels in intermediate strains were generally equal to or in between those in CBD dominant and THC dominant strains. Overall, with higher amounts of β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, and α-bisabolol, intermediate strains more resemble CBD dominant strains than THC dominant strains. The results of this study provide a comprehensive profile of bioactive compounds in three chemotypes for medical purposes. The simultaneous presence of a predominant number of identified chemotype markers (with or without THC and CBD) could be used as chemical fingerprints for quality standardization or strain identification for research, clinical studies, and cannabis product manufacturing.

(**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Chemotypic Markers in Three Chemotype Categories of Cannabis Using Secondary Metabolites Profiled in Inflorescences, Leaves, Stem Bark, and Roots. *Front. Plant Sci.* 12, (**2021**).

4.2 Introduction

Cannabis is a complex herbal medicine containing several classes of secondary metabolites, including cannabinoids, terpenoids, flavonoids, and steroids among 545 identified compounds [26,80,89–92,204]. For medical applications, researchers widely adopt a chemotaxonomic perspective that describes three chemotypes (chemical phenotypes) based on the content of two major cannabinoids: psychoactive THC and non-psychoactive CBD [59,64,221,222]. THC dominant strains have a ratio of THC/CBD > 1, intermediate strains have THC/CBD ≈ 1, and CBD dominant strains have THC/CBD <1. Although most clinical studies focus on THC and CBD, increasing amounts of evidence show that whole plant extract has additional benefits when compared to single cannabinoids. In one study, whole cannabis extract was more effective in inducing cancer cell death than applying pure THC on cancer cell lines [223]. In addition, individual cannabis extracts with similar amounts of THC produced significantly different effects on the survival of specific cancer cells, and specific cannabis extracts may selectively and differentially affect different cancer cells lines [223]. In another study, extracts from five strains with similar CBD concentrations had different anticonvulsant properties in mice [224]. These studies suggest that there may exist therapeutic-enhancing interactions or synergistic effects amongst cannabinoids as well as between cannabinoids and other secondary metabolites, known as the "entourage effect" [81,138,139]. It is therefore essential to have a comprehensive, full spectrum metabolic fingerprinting of secondary metabolites in cannabis materials for research and clinical studies. Previous research also focused on female inflorescences, however, each part of the plant has a wide range of indications, primarily related with pain and inflammation, as ancient herbal medicines in various cultures [114–117]. Our previous study profiled cannabinoids, terpenoids, flavonoids, sterols, and triterpenoids, not only in cannabis inflorescences, but also in leaves, stem bark, and roots [204]. By profiling these compounds in each cannabis plant part and associating them with therapeutic benefits, cannabis plant material that is currently treated as waste has potential to be developed into natural health products or medications.

Cannabis classification is a fundamental requirement for future medical research and applications, and it is best enabled through an overview of the class and content of potentially therapeutic secondary metabolites in each plant part. Currently, researchers attempted to discriminate and identify the chemical differences between the categories of "Sativa" (narrow-leaflet drug, NLD)

and “Indica” (wide-leaflet drug, WLD) [70–72]. Results of the chemotaxonomic separation of “Sativa” and “Indica” were mixed, and THC and CBD concentrations appeared to have no differentiation value. However, certain terpenoids were more prominent in some strains than others [67,70–73,83,84,149]. The mixed results in the current body of literature may be due to experimental design shortcomings. Firstly, the vernacular terminology (“Sativa” and “Indica”) is inadequate for medical applications due to the misuse of the botanical nomenclature, extensive cross-breeding, and unreliable labelling during unrecorded hybridization [69]. Secondly, samples in most classification studies were collected from disparate sources [70,72] and are subject to inconsistent environmental factors during the growth phases [78] and post-harvest treatment [225]. Additionally, inappropriate sample preparation and extraction procedures during laboratory analysis may affect classification results [204]. All these factors contribute to the variation in chemical profiles of the final products, which in turn leads to inconsistent results and poor classification accuracy. More accurate classification results are obtainable when plants are grown in a single location, under identical environmental conditions, and uniformly processed [69].

The chemical profile of CBD dominant and intermediate strains, which have gained increasing attention due to CBD’s use as a therapeutic [183–186], have not been studied or compared to THC dominant strains in the current literature. In this study, we used unsupervised hierarchical clustering and PCA as well as supervised canonical correlation analysis to test the goodness of fit between chemotype labelling (THC dominant, intermediate, and CBD dominant) and chemotypic variation of the full spectrum of secondary metabolites in various plant parts of 21 strains. This study also identifies chemotypic markers within each chemotype, which will facilitate strain selection for further clinical and research studies.

The objectives of this study were to:

1. Investigate whether modern cannabis strains can be differentiated using a full spectrum of secondary metabolites in three chemotypes, including 14 cannabinoids, 45 terpenoids, 7 flavonoids, 3 sterols, and 3 triterpenoids, in inflorescences, leaves, stem bark, and roots;

2. Investigate whether the secondary metabolites described above can differentiate strains into three chemotypes without leveraging THC and CBD data; and

3. Identify chemotypic markers that can be leveraged to select and distinguish chemotypes.

4.3 Materials and methods

4.3.1 Plant material

In this project, 23 commercially available cannabis strains were grown in a commercial greenhouse (**Figure 4.1**) under a cannabis research license issued by Health Canada. Plants for two strains were not rooted successfully and were excluded in the study. Where possible, the reported ancestry ("Sativa-dominant", "Indica-dominant", or "hybrid") was obtained from the Leafly online database (https://www.leafly.ca/) or from the licensed cultivator providing the strain (**Table 4.1**). Three to five cuttings per strain were rooted for two weeks, followed by vegetative growth under 24 hours photoperiod for two months, and then flowered under 12 hours photoperiod. After two months of flowering, the plants were harvested and hung to dry in a closed environment. Cannabis roots were removed and dried in the same room together with the other plant parts. Horticultural fans were used to maintain air circulation, and the temperature was kept under 35°C. The plants were dried for 7 days until the leaves and stems became brittle. At this time, the plants' moisture content is usually below 10-15% (mg/mg%) [226,227].

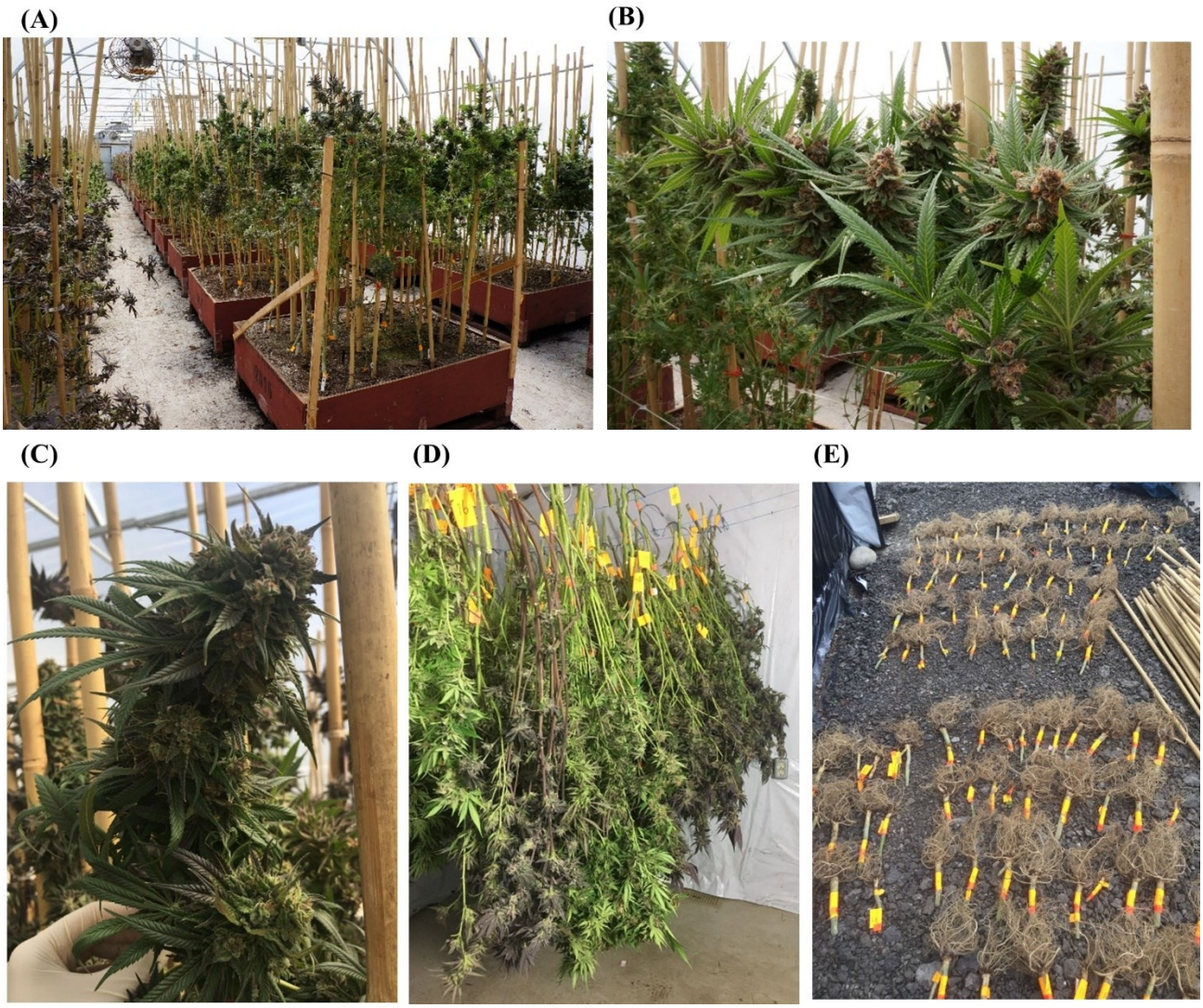


Figure 4.1 Cannabis grown in a commercial greenhouse.
(A)(B)(C) Cannabis plants before harvest. (D) Whole cannabis plants were cut above the ground and hang to dry in a drying room. (E) Cannabis roots were individually labelled and dried in the drying room with the other plant parts.

Table 4.1 Strain information and assignment of 21 strains into three chemotypes.

Variety number	Variety name	Number of plants	Chemotypes	Clusters	"Sativa" or "Indica"	Voucher
1	Lemon Garlic OG	4	1-Intermediate	C2	"Indica" dominant	Teff_Lgog
2	Royal Medic	3	2-Intermediate	C2	"Sativa" dominant	Teff_Rm
3	Blue Hawaiian	4	3-CBD	C1	"Sativa" dominant	Teff_Bh
4	Kandy Kush	5	4-CBD	C1	"Sativa" dominant	Teff_KK
5	Special	3	5-CBD	C1	Not provided	Teff_Sp
6	NN	4	6-CBD	C1	Not provided	Teff_Nn
7*	Dance World	0	7-Intermediate	N/A	"Sativa" dominant	Teff_Dw
8	Treat	4	8-CBD	C1	Not provided	Teff_Tr
9	High	5	9-Intermediate	C2	Not provided	Teff_Hi
10	CB7	4	10-CBD	C1	Not provided	Teff_Cb

11	33°	5	11-THC	C3	Not provided	Teff_33
12	Banana Cake	4	12-THC	C3	"Indica" dominant	Teff_BC
13	Bananium	5	13-THC	C3	"Indica" dominant	Teff_Bc
14	Burmese Blueberry	3	14-THC	C3	"Indica" dominant	Teff_Ba
15	Divine Banana	4	15-THC	C3	"Indica" dominant	Teff_Db
16	Granddaddy Purple	3	16-THC	C3	"Indica" dominant	Teff_Gp
17*	Lemon Love	2	17-THC	N/A	"Indica" dominant	Teff_Ll
18	Lemon Sorbet	4	18-THC	C3	"Indica" dominant	Teff_Ls
19	Meat Head	4	19-THC	C3	"Indica" dominant	Teff_Mh
20	Nanitro	4	20-THC	C3	"Indica" dominant	Teff_Na
21	Platinum Jelly Punch	4	21-THC	C3	"Indica" dominant	Teff_PJP
22	SBSK2 (Lemon Thai)	3	22-THC	C3	50/50 hybrid	Teff_Lt
23	Super Sherbet	3	23-THC	C3	"Indica" dominant	Teff_Ss

* All strains were provided by licensed cultivator The Emerald Flower Farm Inc. (Kelowna, BC, Canada). Specimens of each variety were stored at the research-licenced Labs-Mart Inc. (Edmonton, AB, Canada).
*Strain 7-intermediate was not included in the final analysis due to unsuccessful rooting. Only two plants were available for strain-17 Lemon Love and they were not included in the analysis.

4.3.2 Sample preparation, extraction, and assay

A total of 82 plants representing 21 strains were harvested. Inflorescences, leaves (fan leaves), stem bark, and roots were separately collected for each plant and analyzed for the full spectrum of secondary metabolites. Sugar leaves (small leaves extending from the inflorescences) were treated as a part of the inflorescences. Samples were prepared and analyzed according to previously developed and validated methodologies [204]. Five to eight inflorescences (2g to 4g) of each strain were pulverized with a SPEX Geno/Grinder homogenizer (SPEX SamplePrep, Canada). Dried leaf material was crushed using a mortar and pestle and sifted through a 1.18mm sieve. Dried stem bark and root samples were ground with the SPEX Geno/Grinder homogenizer. For cannabinoids and terpenoids extraction, 400mg of plant material was extracted with 20.0mL methanol (with 100 µg/mL tridecane as an internal standard for mono- and sesquiterpenoids) by sonication for 20 minutes at room temperature. For cannabinoids, the extract was spiked with Δ^9-THC-d_3 (0.5 µg/mL) as an internal standard prior to LC-MS analysis. One aliquot of the extract was used to quantify mono- and sesquiterpenoids using GC-MS. For flavonoids extraction, 250mg of the sample was extracted with 5mL of ethanol, water, and hydrochloric acid at a 25:10:4 volume ratio. The extract was hydrolyzed in a 100°C water bath for 135 minutes. The tube was then repeatedly rinsed with methanol, and the rinses were combined with the extract in a 50 mL volumetric flask, which was filled to volume with methanol. For the flavonoids assay, HPLC was used with an UV detector at 350nm for the quantification of seven flavonoids and MS detector for compound identification.

For triterpenoids and sterols extraction, one gram of dried sample was extracted with 20 mL ethyl acetate by sonication for one hour, followed by maceration for one day at room temperature. The extract was spiked with cholesterol (50 μg/mL) as an internal standard prior to GC-MS analysis.

4.3.3 Statistical analysis

In total, 82 plants representing 21 strains were included in the following analysis. Cannabinoids were calculated as the sum of their neutral forms, metabolites (if applicable), and cannabinoid acids (multiplied by a factor converting acids into their corresponding neutral forms). For example, total THC = Δ^9-THC + Δ^8-THC + CBN (cannabinol, degradation product of THC) + 0.877 × tetrahydrocannabinolic acid (THCA) , total CBD = CBD + 0.877 × cannabidiolic acid (CBDA), total CBG = CBG + 0.878 × cannabigerolic acid (CBGA), total CBC = CBC + 0.877× cannabichromenic acid (CBCA), total THCV = THCV + 0.867 × tetrahydrocannabivarinic acid (THCVA), and total CBDV = CBDV + 0.867 × cannabidivarinic acid (CBDVA) [79,204]. Total cannabinoids was calculated as the sum of 14 cannabinoids. Total monoterpenoids (terpenoids with two isoprene units in the chemical structure) was the sum of the 29 monoterpenoids in **Table 4.7**, and total sesquiterpenoids (terpenoids with three isoprene units) were calculated as the sum of the 16 sesquiterpenoids. Total terpenoids was the sum of total mono- and sesquiterpenoids. Total flavonoids was the sum of seven flavonoids after acid hydrolysis, including orientin, vitexin, isovitexin, quercetin, luteolin, kaempferol, and apigenin. Total sterols was the sum of campesterol, stigmasterol, and β-sitosterol. Total triterpenoids was the sum of β-amyrin, epifriedelanol, and friedelin. Compound ratios were calculated by dividing the content of one compound by the total content of that metabolite group. For example, the ratio of β-pinene was calculated as its absolute value divided by total terpenoids.

Secondary metabolites were quantified in each plant part. The following analyses were carried out only on the metabolites in the plant part where they were of highest levels among all plant parts. This distinction is made for isolating metabolites where they are present in sufficiently high concentrations (above 0.05%) to be of pharmacological interest [81]. First, correlations were calculated between individual cannabinoids, terpenoids, flavonoids, sterols, and triterpenoids. Because absolute values vary with environmental factors and relative proportions are more stable [27], compound ratios were used. Then, unsupervised (no preassigned categories as constraints) hierarchical clustering using Ward's minimum variance method [215] and PCA [228] were used to check

within-strain and between-cluster variation. Finally, the data were subjected to supervised (with preassigned categories as constraints) canonical correlation analysis with preassigned chemotypes in **Table 4.2**. The full spectrum of secondary metabolites, without THC and CBD, were subjected to hierarchical clustering, PCA, and canonical correlation analysis to investigate whether the absence of THC and CBD data would affect differentiating strains into chemotypes.

Table 4.2 Preassigned chemotypes as the working groups for canonical correlation analysis

Clusters	Number of strains	Strain codes as chemotypes
C1 (CBD dominant)	6	3-CBD, 4-CBD, 5-CBD, 6-CBD, 8-CBD, 10-CBD
C2 (Intermediate)	3	1-Intermediate, 2-Intermediate, 9-Intermediate
C3 (THC dominant)	12	11-THC, 12-THC, 13-THC, 14-THC, 15-THC, 16-THC, 18-THC, 19-THC, 20-THC, 21-THC, 22-THC, 23-THC

Canonical correlation analysis is also called canonical variates analysis, and is a multiple discriminant analysis that calculates the correlation between preassigned clusters and the set of covariates (chemical compounds in this study) describing the observations [88]. The first canonical variable is the linear combination of the covariates that maximizes the multiple correlation between the clusters and the covariates. The second canonical variable is a linear combination uncorrelated with the first canonical variable that maximizes the multiple correlation. The analysis outputs a biplot with the first two canonical variables that provide maximum separation among the clusters. To identify marker metabolites that contribute most to the groupings, one-way ANOVA followed by Tukey honestly significant difference (HSD) post hoc test at the 0.05 significance level were used to determine whether significant differences exist between all clusters and each pair of clusters. Statistical analysis was performed with JMP 14.0.0.

4.4 Results

4.4.1 Secondary metabolites profiled in cannabis inflorescences, leaves, stem bark, and roots

Secondary metabolites profiled in inflorescences, leaves, stem bark, and roots are provided in **Supplementary Table 4.1**. Average total cannabinoids content from 82 plants of 21 strains decreased in order of inflorescences, leaves, stem bark, and roots, as shown in **Table 4.3**. Total cannabinoids were between 7.06% and 24.42% with an average of 15.90% ± 4.02% (SD) in inflorescences, between 0.95% and 4.28% with an average of 2.17% ± 0.71% in leaves, between 0.06% and 2.33% with an average of 0.58% ± 0.28% in stem bark, and less than 0.03% in roots.

Total average cannabinoids content in inflorescences were 17.16% ± 4.60%, 14.98% ± 2.63%, and 13.96% ± 2.15% in THC dominant, intermediate, and CBD dominant strains, respectively (**Table 4.4**). These values are typical for modern cannabis strains in North America and mostly agreed with reported values in the literature, which are generally between 5% to 25% [72,76,91,149,229,230]. THC dominant strains had significantly higher concentrations of cannabinoids than the other two chemotypes ($p = 0.0035$). Total cannabinoids content in leaves and stem bark averaged from three chemotypes are summarized in **Table 4.5** and **Table 4.6**.

Table 4.3 Secondary metabolites profiled in inflorescences of 82 plants of 21 strains.

	Inflorescences (mean ± SD)	Leaves (mean ± SD)	Stem bark (mean ± SD)	Roots (mean ± SD)
Total cannabinoids	15.904% ± 4.017%	2.166% ± 0.706%	0.581% ± 0.284%	<0.03%
Total terpenoids	1.509% ± 0.467%	0.110% ± 0.037%	<0.03%	<0.03%
Total flavonoids	0.091% ± 0.050%	0.188% ± 0.098%	<0.03%	<0.03%
Total sterols	<0.03%	<0.03%	0.055% ± 0.013%	0.066% ± 0.009%
Total triterpenoids	<0.03%	<0.03%	0.039% ± 0.023%	0.182% ± 0.043%

Table 4.4 Cannabinoids profiled in inflorescences of 82 plants for three chemotypes.

Inflorescences	**C1 - CBD**	**C2 - Intermediate**	**C3 - THC**
	(N=24)	(N=12)	(N=46)
1. CBDV	0.0078% ± 0.0004%	0.007% ± 0.003%	0.0002% ± 0.0012%
2. CBDVA	0.039% ± 0.010%	0.035% ± 0.013%	0.006% ± 0.006%
3. CBG	0.057% ± 0.040%	0.069% ± 0.031%	0.078% ± 0.042%
4. CBD	0.374% ± 0.097%	0.216% ± 0.051%	0.005% ± 0.004%
5. CBDA	12.020% ± 1.863%	8.931% ± 1.829%	0.061% ± 0.021%
6. THCV	0.007% ± 0.002%	0.008% ± 0.000%	0.010% ± 0.003%
7. CBGA	0.280% ± 0.102%	0.355% ± 0.219%	0.689% ± 0.402%
8. CBN	ND*	ND	ND
9. Δ^9-THC	0.042% ± 0.009%	0.294% ± 0.114%	0.350% ± 0.140%
10. Δ^8-THC	ND	ND	ND
11. THCVA	ND	0.016% ± 0.010%	0.186% ± 0.231%
12. CBC	0.040% ± 0.010%	0.027% ± 0.009%	0.017% ± 0.007%
13. THCA	0.488% ± 0.088%	4.463% ± 0.808%	15.333% ± 4.221%
14. CBCA	0.600% ± 0.140%	0.558% ± 0.143%	0.427% ± 0.224%
Total CBDV*	0.042% ± 0.009%	0.037% ± 0.010%	0.005% ± 0.006%
Total CBG*	0.303% ± 0.100%	0.380% ± 0.212%	0.682% ± 0.374%
Total CBD*	10.915% ± 1.686%	8.049% ± 1.575%	0.059% ± 0.019%
Total THCV*	0.007% ± 0.002%	0.022% ± 0.009%	0.171% ± 0.203%
Total THC*	0.471% ± 0.080%	4.208% ± 0.665%	13.797% ± 3.750%
Total CBC*	0.566% ± 0.127%	0.516% ± 0.120%	0.392% ± 0.200%
Total cannabinoids*	13.956% ± 2.147%	14.979% ± 2.626%	17.162% ± 4.597%

*Total CBDV = CBDV + 0.867 × CBDVA

*Total CBG = CBG + 0.878 × CBGA.
*Total CBD = CBD + 0.877 × CBDA.
*Total THCV = THCV + 0.867 × THCVA.
*Total THC = Δ^9-THC + Δ^8-THC + CBN + 0.877 × THCA.
*Total CBC = CBC + 0.877× CBCA.
*Total cannabinoids = sum of 14 cannabinoids.
*ND = Not detected or below quantification limit (trace amount).

Table 4.5 Cannabinoid profile in leaves of 82 plants for three chemotypes.

Leaves	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=18)	(N=9)	(N=43)
1. CBDV	0.002% ± 0.004%	0.002% ± 0.004%	ND
2. CBDVA	0.021% ± 0.004%	0.0192% ± 0.0004%	0.007% ± 0.009%
3. CBG	0.016% ± 0.005%	0.014% ± 0.005%	0.016% ± 0.006%
4. CBD	0.070% ± 0.048%	0.043% ± 0.020%	0.008% ± 0.004%
5. CBDA	1.438% ± 0.357%	1.012% ± 0.269%	0.056% ± 0.054%
6. THCV	0.006% ± 0.018%	0.003% ± 0.005%	0.004% ± 0.005%
7. CBGA	0.055% ± 0.021%	0.057% ± 0.025%	0.098% ± 0.068%
8. CBN	ND*	ND	ND
9. Δ^9-THC	0.019% ± 0.006%	0.063% ± 0.024%	0.141% ± 0.100%
10. Δ^8-THC	ND	ND	ND
11. THCVA	ND	0.002% ± 0.004%	0.024% ± 0.021%
12. CBC	0.020% ± 0.002%	0.022% ± 0.005%	0.032% ± 0.017%
13. THCA	0.213% ± 0.114%	0.702% ± 0.165%	1.622% ± 0.621%
14. CBCA	0.095% ± 0.019%	0.133% ± 0.037%	0.263% ± 0.180%
Total CBDV	0.021% ± 0.005%	0.019% ± 0.005%	0.006% ± 0.008%
Total CBG	0.064% ± 0.021%	0.064% ± 0.021%	0.102% ± 0.062%
Total CBD	1.332% ± 0.337%	0.931% ± 0.243%	0.057% ± 0.049%
Total THCV	0.006% ± 0.018%	0.005% ± 0.005%	0.026% ± 0.021%
Total THC	0.206% ± 0.104%	0.680% ± 0.159%	1.564% ± 0.580%
Total CBC	0.103% ± 0.016%	0.139% ± 0.032%	0.263% ± 0.166%
Total cannabinoids	1.956% ± 0.451%	2.075% ± 0.481%	2.273% ± 0.812%

Table 4.6 Cannabinoid profile in stem bark of 82 plants for three chemotypes.

Stem bark	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=23)	(N=12)	(N=46)
1. CBDV	0.0004% ± 0.0021%	0.001% ± 0.003%	ND
2. CBDVA	0.007% ± 0.004%	0.007% ± 0.005%	0.002% ± 0.004%
3. CBG	0.009% ± 0.003%	0.0099% ± 0.0001%	0.010% ± 0.001%
4. CBD	0.009% ± 0.006%	0.005% ± 0.005%	0.004% ± 0.005%
5. CBDA	0.286% ± 0.156%	0.209% ± 0.117%	0.055% ± 0.046%
6. THCV	ND*	ND	ND
7. CBGA	0.022% ± 0.008%	0.025% ± 0.012%	0.025% ± 0.011%
8. CBN	ND	ND	ND
9. Δ^9-THC	0.015% ± 0.005%	0.021% ± 0.003%	0.024% ± 0.007%
10. Δ^8-THC	ND	ND	ND

Stem bark	C1 - CBD	C2 - Intermediate	C3 - THC
11. THCVA	0.004% ± 0.005%	0.004% ± 0.005%	0.004% ± 0.007%
12. CBC	0.010% ± 0.008%	0.012% ± 0.008%	0.007% ± 0.007%
13. THCA	0.150% ± 0.104%	0.219% ± 0.094%	0.446% ± 0.295%
14. CBCA	0.032% ± 0.011%	0.030% ± 0.014%	0.030% ± 0.019%
Total CBDV	0.007% ± 0.005%	0.007% ± 0.006%	0.001% ± 0.003%
Total CBG	0.028% ± 0.008%	0.032% ± 0.011%	0.032% ± 0.009%
Total CBD	0.260% ± 0.140%	0.189% ± 0.105%	0.052% ± 0.042%
Total THCV	0.004% ± 0.004%	0.004% ± 0.005%	0.004% ± 0.006%
Total THC	0.146% ± 0.094%	0.212% ± 0.084%	0.416% ± 0.260%
Total CBC	0.038% ± 0.013%	0.037% ± 0.014%	0.034% ± 0.019%
Total cannabinoids	0.545% ± 0.252%	0.542% ± 0.232%	0.609% ± 0.312%

Average total terpenoids as the sum of mono- and sesquiterpenoids in the same population decreased in order of inflorescences, leaves, stem bark, and roots (**Figure 4.2**). Total terpenoids in inflorescences was between 0.753% and 3.305% with an average of 1.509% ± 0.467%, in leaves between 0.035% and 0.197% with an average of 0.103% ± 0.032%, and in stem bark and roots less than 0.03% (**Table 4.3**). Average total terpenoids content in inflorescences and leaves for the three chemotypes are summarized in **Table 4.7** and **Table 4.8.**

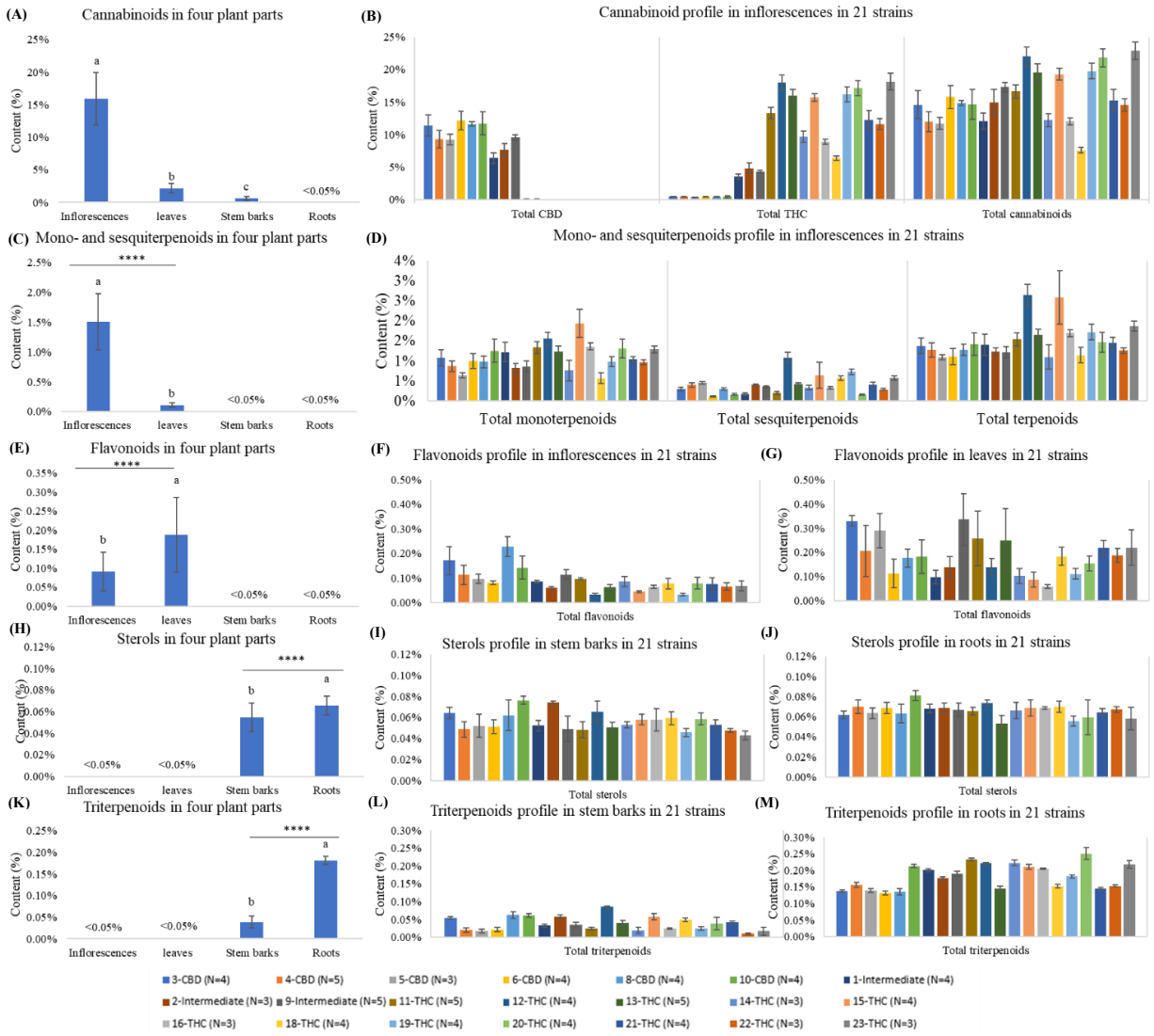


Figure 4.2 Secondary metabolites profiling in cannabis roots, stem bark, leaves, and inflorescences in 82 plants of 21 strains.

(A) Total cannabinoid content (mg/mg%) in each plant part averaged from 82 plants (N = 82, mean ± standard deviation (SD)%). **(B)** Total CBD, total THC, and total cannabinoid content (mg/mg%) in inflorescences of 21 strains. **(C)** Total mono- and sesquiterpenoid content (mg/mg%) in each plant part averaged from 82 plants (N = 82, mean ± SD%). **(D)** Total mono- and sesquiterpenoids content (mg/mg%) in inflorescences of 21 strains. **(E)** Total flavonoid content (mg/mg%) in each plant part averaged from 82 plants (N = 82, mean ± SD%). **(F)** Total flavonoid content (mg/mg%) in inflorescences of 21 strains. **(G)** Total flavonoid content (mg/mg%) in leaves of 21 strains. **(H)** Total sterol content (mg/mg%) in each plant part averaged from 82 plants (N = 82, mean ± SD%). **(I)** Total sterol content (mg/mg%) in stem bark of 21 strains. **(J)** Total sterol content (mg/mg%) in roots of 21 strains. **(K)** Total triterpenoid content (mg/mg%) in each plant part averaged from 82 plants (N = 82, mean ± SD%). **(L)** Total triterpenoid content (mg/mg%) in stem bark of 21 strains. **(M)** Total triterpenoid content (mg/mg%) in roots of 21 strains. One-way ANOVA followed by correction for multiple comparisons (Tukey honestly significant difference (HSD) post hoc test) at the 0.05 significance level was used (p values indicated above each bar). Asterisks indicate statistically significant differences (one-way ANOVA, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

Table 4.7 Mono- and sesquiterpenoids profile in inflorescences of 82 plants for three chemotypes

Inflorescences	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. α-Pinene	0.187% ± 0.067%	0.083% ± 0.085%	0.130% ± 0.085%
2. Camphene	0.006% ± 0.002%	0.007% ± 0.001%	0.015% ± 0.009%
3. Sabinene	ND*	ND	ND
4. β-Pinene	0.077% ± 0.026%	0.056% ± 0.025%	0.104% ± 0.039%
5. β-Myrcene	0.516% ± 0.143%	0.548% ± 0.160%	0.297% ± 0.228%
6. α-Phellandrene	ND	ND	ND
7. Δ^3-Carene	ND	ND	ND
8. α-Terpinene	ND	ND	ND
9. p-Cymene	ND	ND	ND
10. Limonene	0.092% ± 0.019%	0.134% ± 0.023%	0.326% ± 0.223%
11. 1,8-Cineole (Eucalyptol)	0.007% ± 0.002%	0.012% ± 0.007%	0.005% ± 0.005%
12. Ocimene	0.015% ± 0.018%	0.009% ± 0.005%	0.077% ± 0.061%
13. γ-Terpinene	ND	ND	ND
14. Sabinene Hydrate	0.006% ± 0.002%	0.007% ± 0.002%	0.010% ± 0.004%
15. Terpinolene	0.008% ± 0.009%	0.012% ± 0.012%	0.063% ± 0.123%
16. Fenchone	ND	ND	ND
17. Linalool	0.028% ± 0.008%	0.052% ± 0.019%	0.078% ± 0.063%
18. Fenchol	0.015% ± 0.003%	0.021% ± 0.004%	0.041% ± 0.028%
19. (-)-Isopulegol	ND	ND	ND
20. Camphor	ND	ND	ND
21. Borneol	0.006% ± 0.001%	0.007% ± 0.001%	0.010% ± 0.005%
22. Terpinen-4-ol	ND	ND	ND
23. α-Terpineol	0.019% ± 0.004%	0.027% ± 0.004%	0.054% ± 0.027%
24. (+)-Dihydrocarvone	ND	ND	ND
25. Nerol	ND	ND	ND
26. Pulegone	ND	ND	ND
27. Carvone (isomers)	ND	ND	ND
28. Geraniol	ND	ND	ND
29. Geranyl Acetate	ND	ND	ND
30. (-)-β-Elemene	ND	ND	ND
31. β-Caryophyllene	0.045% ± 0.022%	0.081% ± 0.039%	0.223% ± 0.163%
32. Aromadendrene	ND	ND	ND
33. trans-β-Farnesene	0.006% ± 0.003%	0.006% ± 0.001%	0.025% ± 0.028%
34. α-Humulene	0.014% ± 0.007%	0.024% ± 0.012%	0.075% ± 0.048%
35. Valencene	ND	ND	ND
36. Ledene	ND	ND	ND
37. trans-Nerolidol	0.006% ± 0.002%	0.006% ± 0.002%	0.052% ± 0.048%
38. Caryophyllene Oxide	ND	ND	ND
39. Globulol	ND	ND	ND
40. Viridiflorol	ND	ND	ND

Inflorescences	C1 - CBD	C2 - Intermediate	C3 - THC
41. (-)-Guaiol	0.062% ± 0.025%	0.063% ± 0.023%	0.028% ± 0.026%
42.(+)-Cedrol	ND	ND	ND
43. β-Eudesmol	0.036% ± 0.015%	0.034% ± 0.011%	0.015% ± 0.015%
44. α-Eudesmol	0.021% ± 0.009%	0.018% ± 0.005%	0.010% ± 0.009%
45. α-Bisabolol	0.100% ± 0.061%	0.080% ± 0.028%	0.053% ± 0.030%
Total monoterpenoids*	0.980% ± 0.243%	0.974% ± 0.241%	1.211% ± 0.383%
Total sesquiterpenoids*	0.289% ± 0.124%	0.311% ± 0.099%	0.482% ± 0.275%
Total terpenoids	1.269% ± 0.209%	1.285% ± 0.187%	1.693% ± 0.532%

*Total monoterpenoids = sum of terpenoids 1 – 29.
*Total sesquiterpenoids = sum of terpenoids 30 – 45.
*ND = Not detected or below quantification limit (trace amount).

Table 4.8 Mono- and sesquiterpenoids profile in leaves of 82 plants for three chemotypes.

Leaves	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=18)	(N=9)	(N=43)
1. α-Pinene	0.010% ± 0.006%	0.007% ± 0.006%	0.004% ± 0.004%
2. Camphene	ND	ND	ND
3. Sabinene	ND	ND	ND
4. β-Pinene	0.003% ± 0.002%	0.003% ± 0.001%	0.003% ± 0.002%
5. β-Myrcene	0.008% ± 0.007%	0.003% ± 0.001%	0.004% ± 0.004%
6. α-Phellandrene	ND	ND	ND
7. Δ^3-Carene	ND	ND	ND
8. α-Terpinene	ND	ND	ND
9. p-Cymene	ND	ND	ND
10. Limonene	0.002% ± 0.001%	0.002% ± 0.001%	0.004% ± 0.003%
11. 1,8-Cineole (Eucalyptol)	0.002% ± 0.001%	0.003% ± 0.002%	0.003% ± 0.005%
12. Ocimene	ND	ND	ND
13. γ-Terpinene	ND	ND	ND
14. Sabinene Hydrate	ND	ND	ND
15. Terpinolene	ND	ND	ND
16. Fenchone	ND	ND	ND
17. Linalool	0.001% ± 0.0004%	0.002% ± 0.0005%	0.002% ± 0.002%
18. Fenchol	0.001% ± 0.001%	0.001% ± 0.000%	0.002% ± 0.001%
19. (-)-Isopulegol	ND	ND	ND
20. Camphor	ND	ND	ND
21. Borneol	ND	ND	ND
22. Terpinen-4-ol	ND	ND	ND
23. α-Terpineol	ND	ND	ND
24. (+)-Dihydrocarvone	ND	ND	ND
25. Nerol	ND	ND	ND
26. Pulegone	ND	ND	ND
27. Carvone (isomers)	ND	ND	ND
28. Geraniol	ND	ND	ND

Leaves	C1 - CBD	C2 - Intermediate	C3 - THC
29. Geranyl Acetate	ND	ND	ND
30. (-)-β-Elemene	ND	ND	ND
31. β-Caryophyllene	0.012% ± 0.005%	0.015% ± 0.007%	0.027% ± 0.012%
32. Aromadendrene	ND	ND	ND
33. trans-β-Farnesene	0.003% ± 0.001%	0.003% ± 0.0004%	0.005% ± 0.003%
34. α-Humulene	0.004% ± 0.001%	0.005% ± 0.002%	0.009% ± 0.005%
35. Valencene	ND	ND	ND
36. Ledene	ND	ND	ND
37. trans-Nerolidol	0.001% ± 0.0004%	0.001% ± 0.0004%	0.003% ± 0.002%
38. Caryophyllene Oxide	ND	ND	ND
39. Globulol	ND	ND	ND
40. Viridiflorol	ND	ND	ND
41. (-)-Guaiol	0.010% ± 0.005%	0.012% ± 0.002%	0.004% ± 0.004%
42.(+)-Cedrol	ND	ND	ND
43. β-Eudesmol	0.007% ± 0.004%	0.008% ± 0.002%	0.003% ± 0.002%
44. α-Eudesmol	0.005% ± 0.002%	0.006% ± 0.001%	0.002% ± 0.002%
45. α-Bisabolol	0.036% ± 0.020%	0.045% ± 0.013%	0.023% ± 0.013%
Total monoterpenoids	0.027% ± 0.016%	0.021% ± 0.008%	0.021% ± 0.011%
Total sesquiterpenoids	0.077% ± 0.035%	0.094% ± 0.020%	0.077% ± 0.026%
Total terpenoids	0.104% ± 0.048%	0.115% ± 0.019%	0.099% ± 0.029%

*Total monoterpenoids = sum of terpenoids 1 – 29.
*Total sesquiterpenoids = sum of terpenoids 30 – 45.
*ND = Not detected or below quantification limit (trace amount).

Average total flavonoids as the sum of orientin, vitexin, isovitexin, quercetin, luteolin, kaempferol, and apigenin was highest in leaves, lower in inflorescences, and less than 0.03% in stem bark and roots (**Figure 4.2**). Total flavonoids in inflorescences were between 0.028% and 0.284% with an average of 0.091% ± 0.050%, and in leaves between 0.051% and 0.470% with an average of 0.188% ± 0.098% (**Table 4.3**). Flavonoids exist in cannabis plants as both aglycones and conjugated glycosides and were estimated to be less than 1% in leaves [139] The results of this study was congruent with this estimate, since the flavonoids were not converted to conjugated glycosides. All seven flavonoids were quantifiable in inflorescences in three chemotypes (**Table 4.9**), while quercetin and kaempferol were below the quantification limit in leaves (**Table 4.10**). All flavonoids identified in inflorescences and leaves were less than those reported in other studies [85], possibly due to differences in strains and plant growth stage, since flavonoids content fluctuate with plant age [158].

Table 4.9 Flavonoid profile in inflorescences for three chemotypes.

Inflorescences	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. Orientin (F)	0.051% ± 0.022%	0.021% ± 0.008%	0.014% ± 0.012%
2. Vitexin (F)	0.042% ± 0.018%	0.024% ± 0.006%	0.013% ± 0.011%
3. Isovitexin (F)	0.003% ± 0.001%	0.002% ± 0.001%	0.001% ± 0.001%
4. Quercetin (F)	0.008% ± 0.004%	0.014% ± 0.008%	0.012% ± 0.006%
5. Luteolin (F)	0.027% ± 0.023%	0.021% ± 0.006%	0.018% ± 0.021%
6. Kaempferol (F)	0.0030% ± 0.0004%	0.003% ± 0.001%	0.004% ± 0.001%
7. Apigenin (F)	0.006% ± 0.004%	0.007% ± 0.001%	0.003% ± 0.002%
Total flavonoids	0.140% ± 0.061%	0.092% ± 0.026%	0.065% ± 0.025%

Note: Flavonoids in inflorescences is labelled (F).

Table 4.10 Flavonoid profile in leaves for three chemotypes

Leaves	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. Orientin (L)	0.077% ± 0.064%	0.044% ± 0.032%	0.038% ± 0.039%
2. Vitexin (L)	0.061% ± 0.036%	0.053% ± 0.036%	0.032% ± 0.026%
3. Isovitexin (L)	0.004% ± 0.003%	0.004% ± 0.003%	0.002% ± 0.002%
4. Quercetin (L)	ND	ND	ND
5. Luteolin (L)	0.050% ± 0.040%	0.074% ± 0.046%	0.074% ± 0.068%
6. Kaempferol (L)	ND	ND	ND
7. Apigenin (L)	0.017% ± 0.012%	0.021% ± 0.008%	0.016% ± 0.011%
Total flavonoids	0.213% ± 0.095%	0.208% ± 0.134%	0.170% ± 0.087%

Note: Flavonoids in leaves is labelled (L).

Total sterols content as the sum of three phytosterols, campesterol, stigmasterol, and β-sitosterol was highest in roots, lower in stem bark, and was less than 0.03% in inflorescences and leaves (**Figure 4.2**). Total sterols content in roots was between 0.037% and 0.085% with an average of 0.066% ± 0.009%, and in stem bark was between 0.037% and 0.082% with an average of 0.055% ± 0.013% **(Table 4.3)**. Average total sterols content in stem bark and roots of the three chemotypes are summarized in **Table 4.11** and **Table 4.12.**

Table 4.11 Sterol profile in stem bark for three chemotypes

Stem bark	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. Campesterol	0.012% ± 0.002%	0.012% ± 0.004%	0.011% ± 0.003%
2. Stigmasterol	0.010% ± 0.003%	0.010% ± 0.003%	0.011% ± 0.004%
3. β-sitosterol	0.037% ± 0.008%	0.034% ± 0.008%	0.030% ± 0.008%
Total sterols	0.059% ± 0.012%	0.056% ± 0.013%	0.052% ± 0.014%

Table 4.12 Sterol profile in roots for three chemotypes

Roots	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. Campesterol	0.013% ± 0.001%	0.013% ± 0.001%	0.012% ± 0.002%
2. Stigmasterol	0.012% ± 0.002%	0.013% ± 0.001%	0.013% ± 0.002%
3. β-sitosterol	0.043% ± 0.006%	0.042% ± 0.004%	0.039% ± 0.007%
Total sterols	0.068% ± 0.009%	0.068% ± 0.005%	0.064% ± 0.009%

Total triterpenoids as the sum of β-amyrin, epifriedanol, and friedelin was highest in roots, lower in stem bark, and was less than 0.03% in inflorescences and leaves (**Figure 4.2**). Total triterpenoids in stem bark was between 0.008% and 0.136% with an average of 0.039% ± 0.023%, in roots was between 0.080% and 0.275% with an average of 0.182% ± 0.043% (**Table 4.3**). Average total triterpenoids content in stem bark and roots in the three chemotypes are summarized in **Table 4.13** and **Table 4.14.**

Table 4.13 Triterpenoid profile in stem bark for three chemotypes

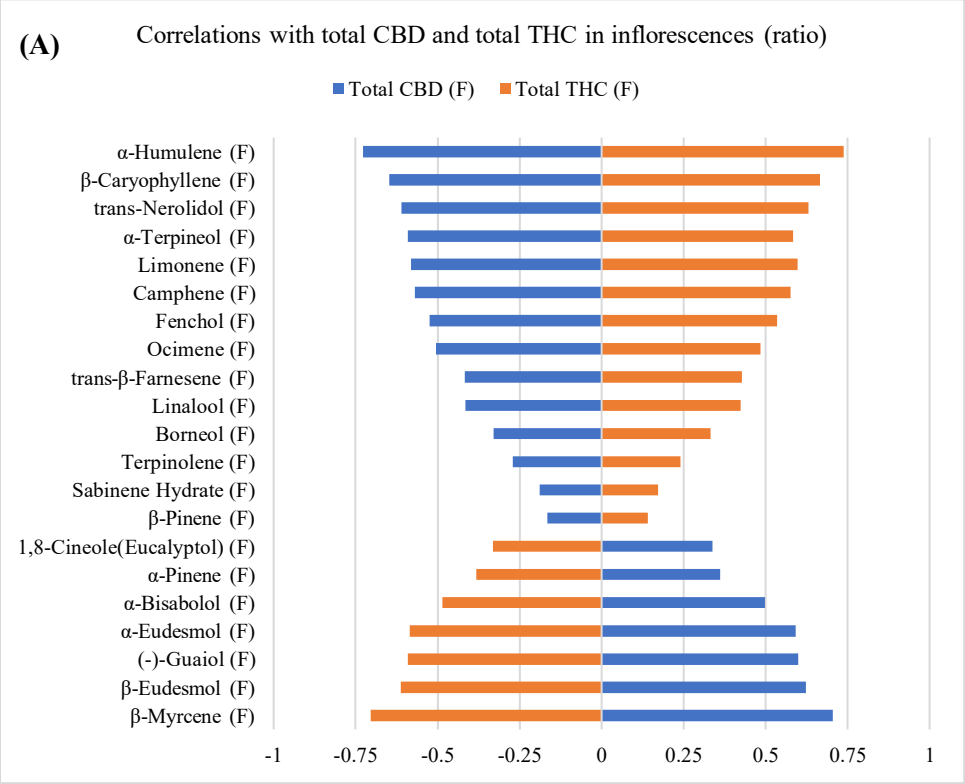
Stem bark	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. β-Amyrin	0.015% ± 0.007%	0.021% ± 0.005%	0.015% ± 0.008%
2. Epifriedanol	0.011% ± 0.008%	0.007% ± 0.004%	0.008% ± 0.008%
3. Friedelin	0.014% ± 0.009%	0.012% ± 0.009%	0.015% ± 0.015%
Total triterpenoids	0.040% ± 0.022%	0.040% ± 0.013%	0.038% ± 0.025%

Table 4.14 Triterpenoid profile in roots for three chemotypes

Roots	C1 - CBD	C2 - Intermediate	C3 - THC
	(N=24)	(N=12)	(N=46)
1. β-Amyrin	0.004% ± 0.001%	0.006% ± 0.001%	0.006% ± 0.001%
2. Epifriedanol	0.055% ± 0.010%	0.064% ± 0.005%	0.062% ± 0.014%
3. Friedelin	0.094% ± 0.024%	0.120% ± 0.011%	0.127% ± 0.034%
Total triterpenoids	0.153% ± 0.032%	0.190% ± 0.016%	0.194% ± 0.046%

The distribution of secondary metabolites in each plant part agreed with conclusions from our last study [204]. Correlation and classification analyses were performed only for metabolites in the plant part where they were present in the highest concentrations representative for that strain. For example, the average terpenoid content in leaves were low (0.103% ± 0.032%) compared to the levels in inflorescences (1.509% ± 0.467%), and only 15 mono- and sesquiterpenoids that were detected in inflorescences were above the quantification limit in leaves (**Table 4.8**). In addition, the correlations between cannabinoids and terpenoids in leaves were like those in inflorescences,

especially for the terpenoids that are abundant in both these two plant parts, including α-pinene, β-pinene, limonene, linalool, β-caryophyllene, trans-β-farnesene, α-humulene, trans-nerolidol, (-) guaiol, β-eudesmol, α-eudesmol, and α-bisabolol (**Figure 4.3**). As such, using the terpenoid profile in inflorescences was adequate for clustering purposes. Flavonoids in inflorescences and leaves were included in the analysis because quercetin and kaempferol were quantifiable in inflorescences but not in leaves. For sterols, the content and ratios of three sterols are similar between stem bark and roots. Because total sterols in roots (0.064%-0.068%) are slightly higher than them in stem bark (0.052%-0.059%), the sterol profiles in roots were used in the data analysis. Triterpenoid profile in roots were used because the content of total triterpenoids was above the threshold for pharmacological interest in all plant parts except in roots. To summarize, the most abundant secondary metabolites in individual plant parts were used in the statistical analysis for identifying differences between the three chemotypes. These metabolites were cannabinoids, terpenoids, and flavonoids in inflorescences; flavonoids in leaves; and sterols and triterpenoids in roots (**Supplementary Table 4.2**).



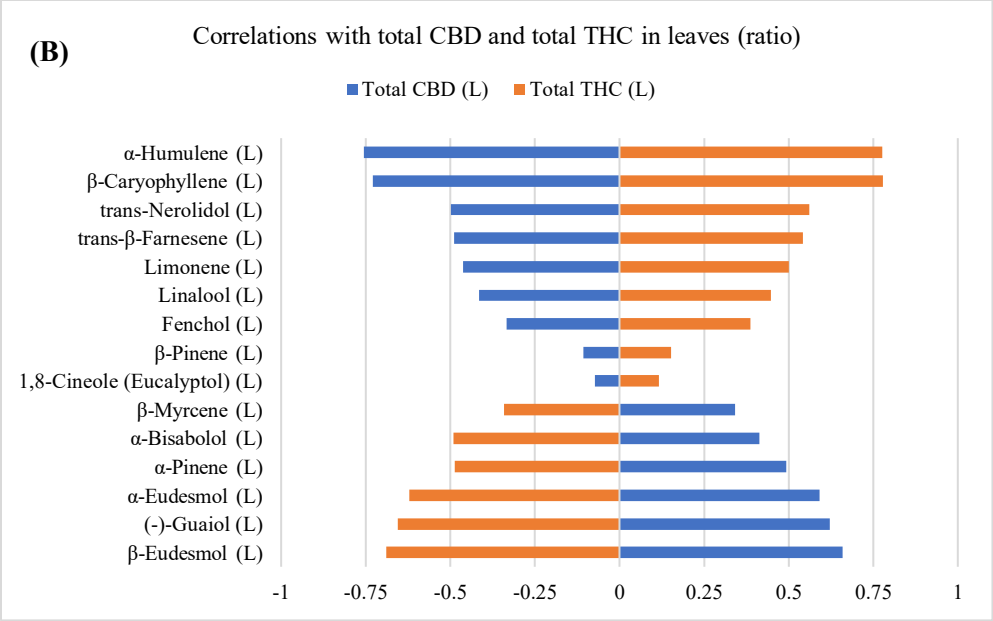


Figure 4.3 Correlations of total THC and total CBD with terpenoids using content ratios (%/%) in (A) inflorescences and in (B) leaves.
Compound quantified in inflorescences are labelled as (F). Compound quantified in leaves are labelled as (L).

4.4.2 Correlation analysis between secondary metabolites

Correlations between total THC or total CBD with individual cannabinoids, terpenoids, flavonoids, sterols and triterpenoids are plotted in **Figure 4.4** and summarized in **Table 4.15**. Calculations were performed on quantifiable compounds using ratios. THC was positively correlated with two cannabinoids (total CBG and total THCV), ten monoterpenoids (α-terpineol, limonene, camphene, fenchol, linalool, ocimene, borneol, terpinolene, β-pinene, and sabinene hydrate), four sesquiterpenoids (α-humulene, β-caryophyllene, trans-nerolidol, and trans-β-farnesene), four flavonoids (quercetin and kaempferol in flowers, luteolin and apigenin in both flowers and leaves), and two triterpenoids (β-amyrin and friedelin). Total CBD was positively correlated with two cannabinoids (total CBDV and total CBC), three monoterpenoids (β-myrcene, 1,8-cineole (eucalyptol), α-pinene), four sesquiterpenoids (β-eudesmol, (-)-guaiol, α-eudesmol, α-bisabolol), three flavonoids (orientin, vitexin, isovitexin in both flowers and leaves), three sterols (campesterol, stigmasterol, β-sitosterol), and one triterpenoid (epifriedanol). Compounds that were positively correlated with THC were all negatively correlated with total CBD, and vice versa. The quantitative correlations are plotted in **Figure 4.5**. Most compounds have similar correlations with total THC and total CBD when calculated using ratios and absolute values.

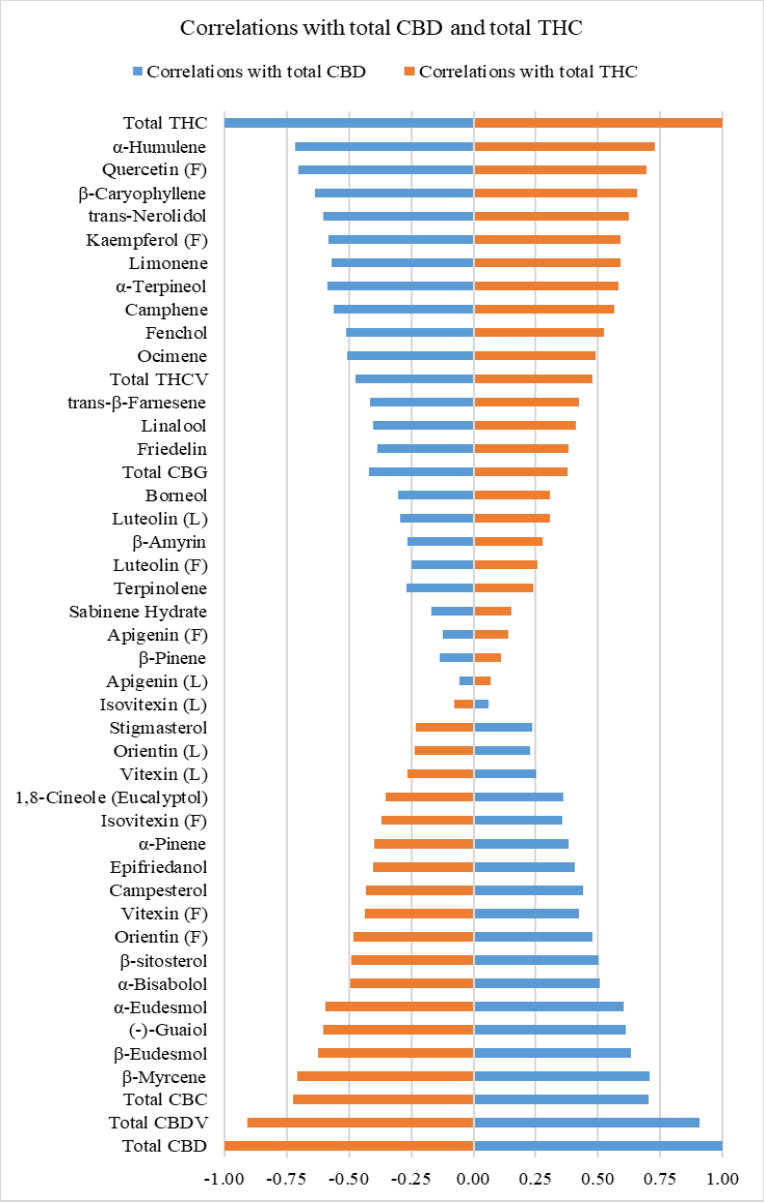


Figure 4.4 Correlations of total THC and total CBD with secondary metabolites in each plant part using ratios.
Flavonoids quantified in inflorescences are labelled (F), and flavonoids in leaf are labelled (L).

Table 4.15 Correlations of total THC and total CBD with minor cannabinoids (in inflorescences), mono- and sesquiterpenoids (in inflorescences), flavonoids (in inflorescences and leaves), sterols and triterpenoids (in roots)

		Correlations with total THC		Correlations with total CBD
Cannabinoids	Total THCV	0.48	Total CBDV	0.91
	Total CBG	0.38	Total CBC	0.70
Monoterpenoids	Limonene	0.59	β-Myrcene	0.71

	α-Terpineol	0.58	α-Pinene	0.38
	Camphene	0.57	1,8-Cineole (eucalyptol)	0.36
	Fenchol	0.52		
	Ocimene	0.49		
	Linalool	0.41		
	Borneol	0.31		
	Terpinolene	0.24		
	Sabinene hydrate	0.15		
	β-Pinene	0.11		
Sesquiterpenoids	α-Humulene	0.73	β-Eudesmol	0.63
	β-Caryophyllene	0.66	(-)-Guaiol	0.61
	trans-Nerolidol	0.63	α-Eudesmol	0.60
	trans-β-Farnesene	0.42	α-Bisabolol	0.51
Flavonoids*	Quercetin (F)	0.70	Orientin (F)	0.48
	Kaempferol (F)	0.59	Vitexin (F)	0.42
	Luteolin (L)	0.31	Isovitexin (F)	0.36
	Luteolin (F)	0.26	Vitexin (L)	0.25
	Apigenin (F)	0.14	Orientin (L)	0.23
	Apigenin (L)	0.07	Isovitexin (L)	0.06
Sterols			β-sitosterol	0.50
			Campesterol	0.44
			Stigmasterol	0.23
Triterpenoids	Friedelin	0.38	Epifriedanol	0.40
	β-Amyrin	0.28		

Note: Flavonoids in inflorescences is labelled (F), and flavonoids in leaves is labelled (L). Only positive correlations are shown.

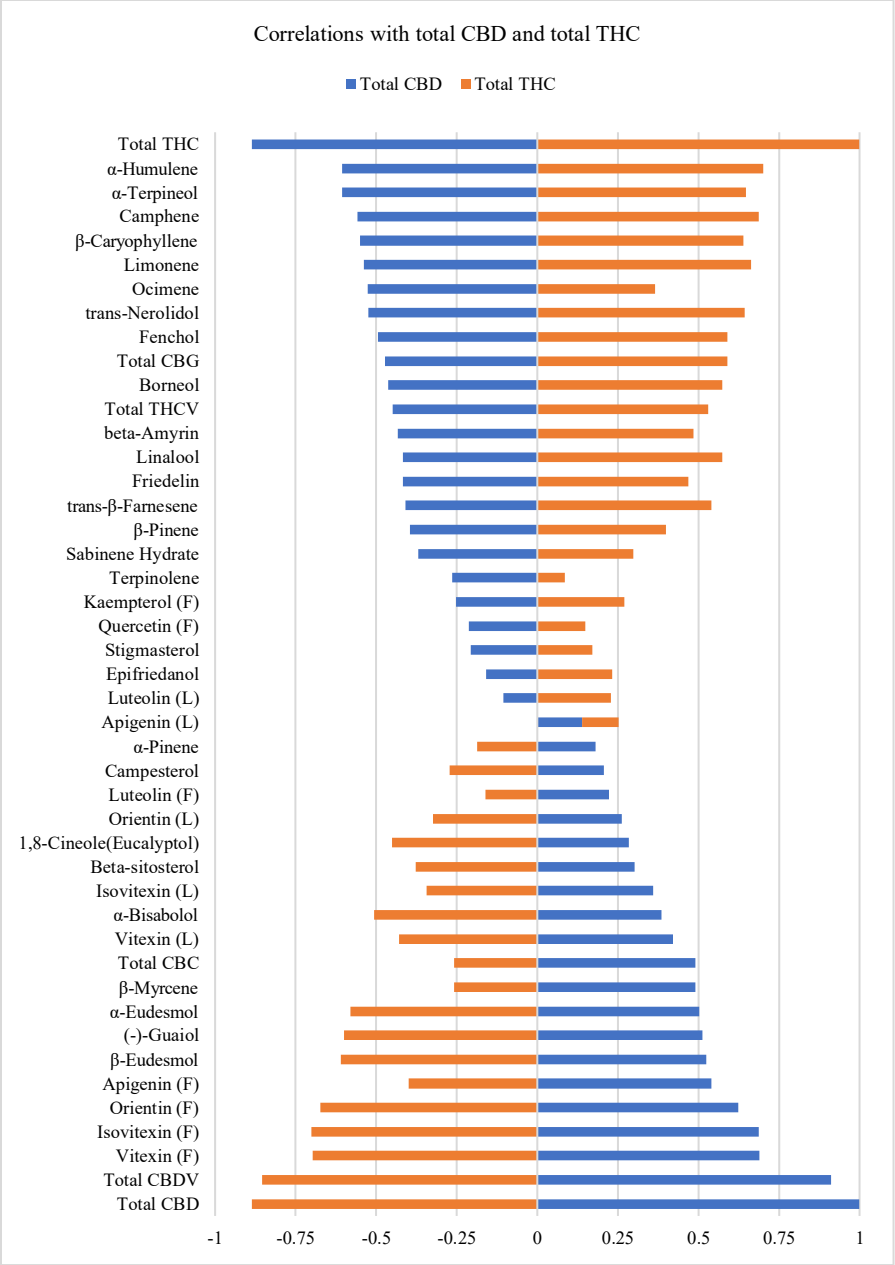


Figure 4.5 Correlations of total THC and total CBD with cannabinoids (in inflorescences), mono- and sesquiterpenoids (in inflorescences), flavonoids (in inflorescences and leaves), sterols and triterpenoids (in roots) on quantifiable compounds using absolute value Supplementary

Flavonoids quantified in inflorescences is labelled (F), and flavonoids in leaf is labelled (L).

4.4.3 Unsupervised hierarchical clustering

The same set of data was used to build a dendrogram of the 82 plants using hierarchical clustering, where almost all plants of the same strains were clustered together, except for one 5-CBD plant that was mixed with 4-CBD plants and plants of 15-THC that were mixed with 23-THC plants (**Figure 4.6**). The dendrogram shows two major branches: CBD dominant strains and intermediate strains together as one major branch, and THC dominant strains as the other. The dendrogram using absolute values of the secondary metabolites is shown in **Figure 4.7**. These results both confirmed the minimum within-strain variation (between plants within each strain) and between-cluster variation (between strains within each chemotypes). The full spectrum of secondary metabolites without total THC and total CBD resulted in a dendrogram with the same grouping results (**Figure 4.8**).

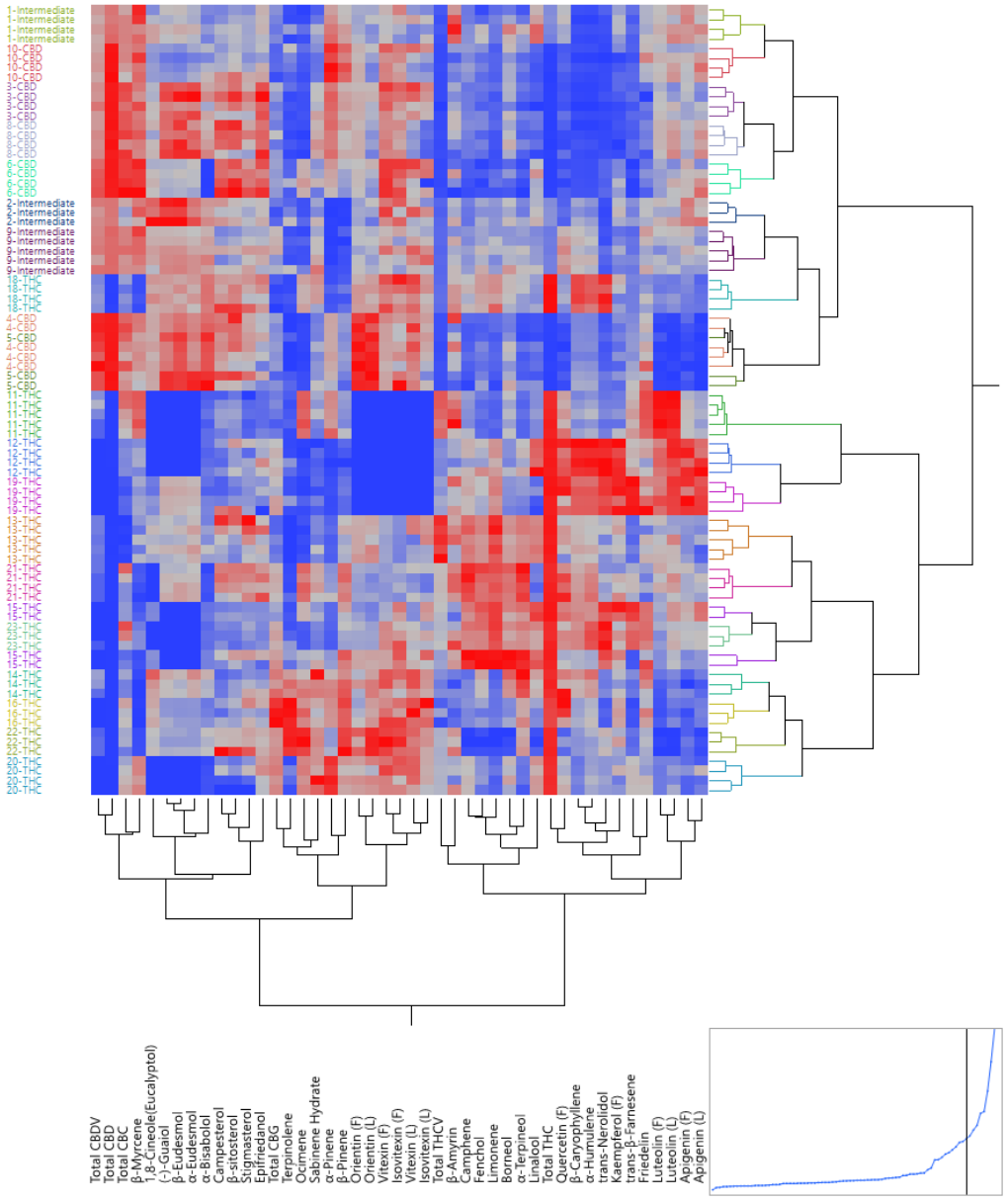


Figure 4.6 Dendrogram by hierarchical clustering analysis using the full spectrum of secondary metabolites (in ratios) of 82 plants representing 21 strains.

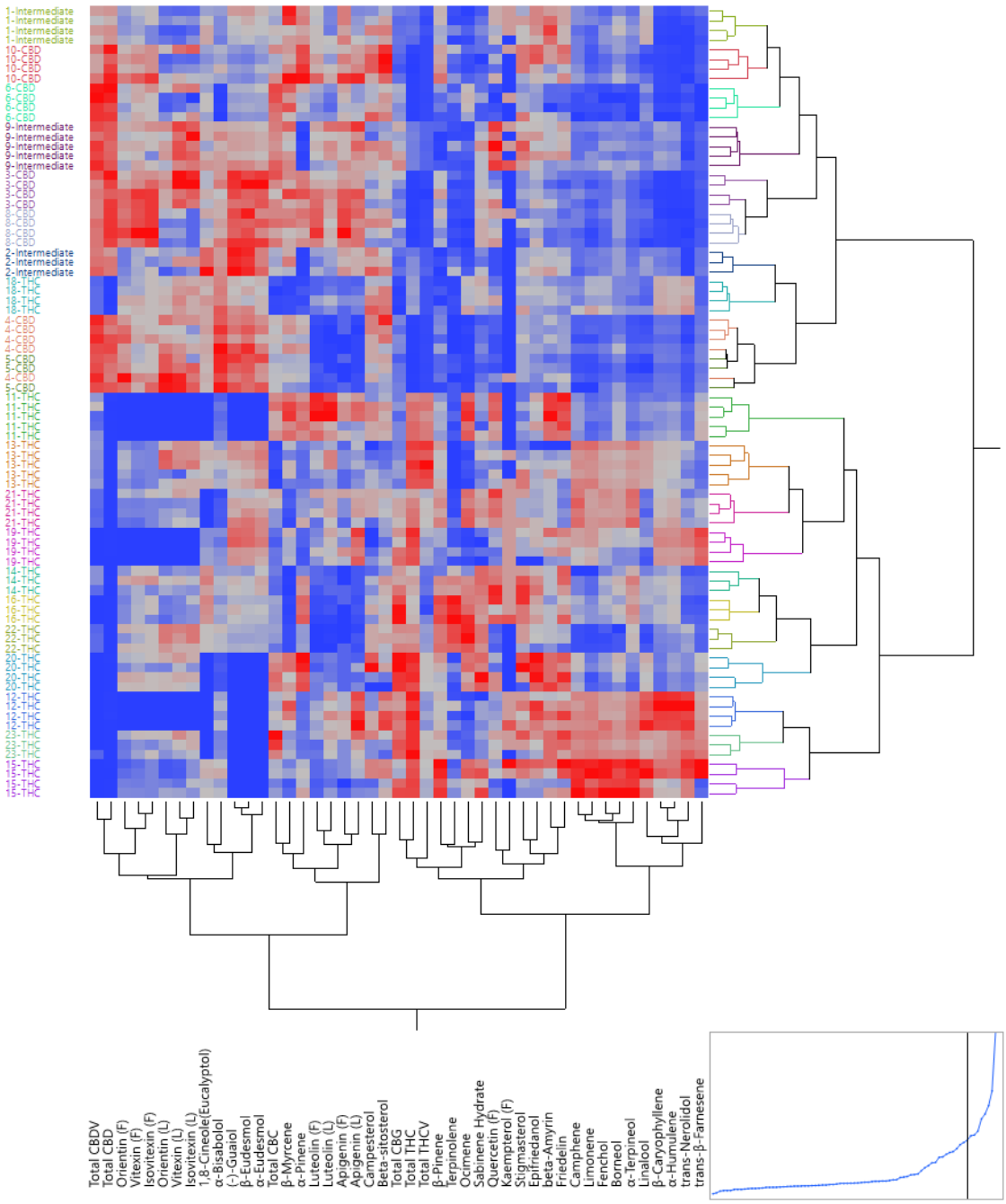


Figure 4.7 Dendrogram by hierarchical clustering analysis using the full spectrum of secondary metabolites (absolute values) of 82 plants representing 21 strains.

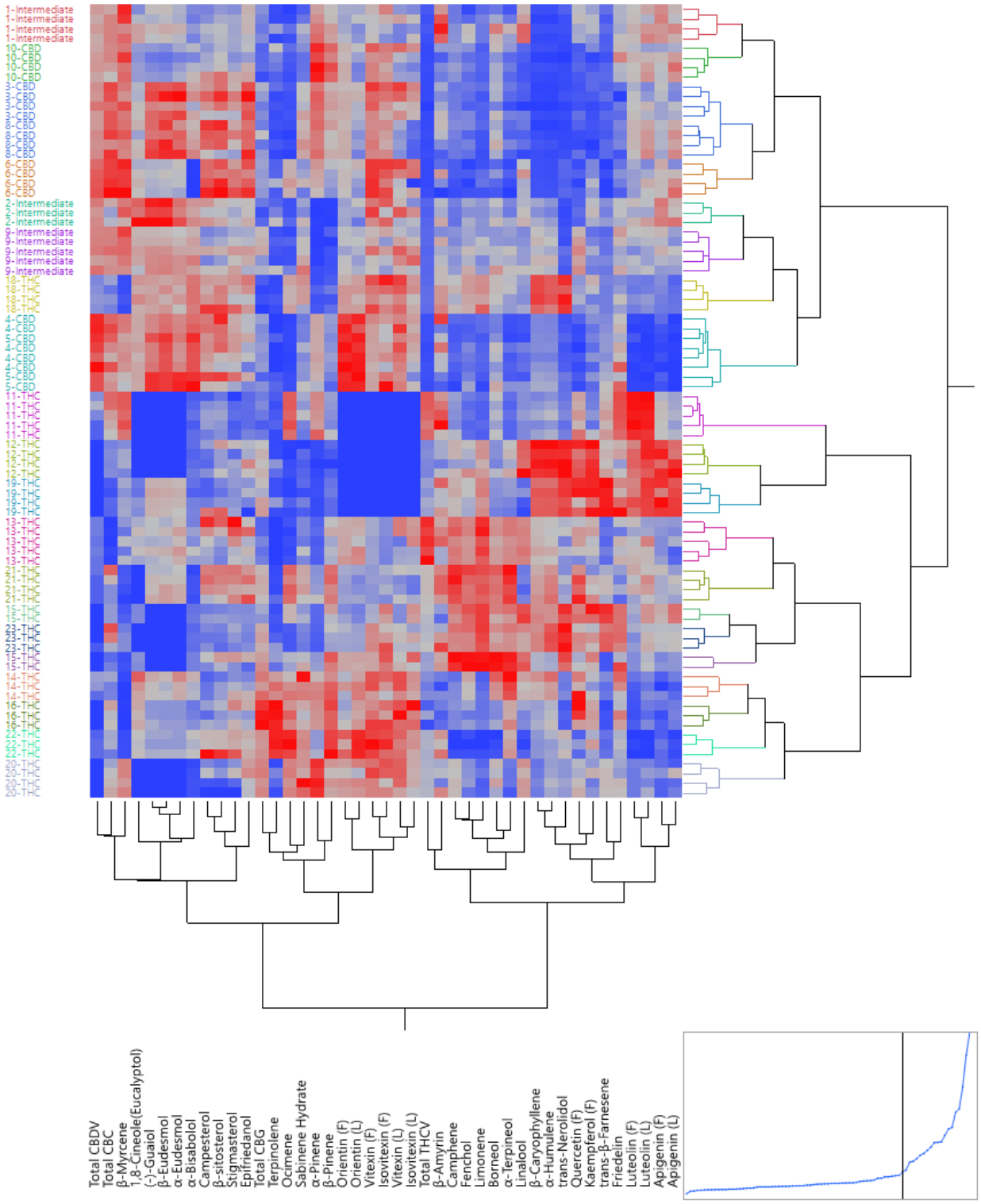


Figure 4.8 Dendrogram by hierarchical clustering analysis using the full spectrum of secondary metabolites (using ratios) without total THC and total CBD.

.

4.4.4 Unsupervised principal component analysis

Figure 4.9 shows a scatterplot of 82 plants along two principal components (PC), where PC1 and PC2 explained 33.8% and 16.4% of the total variance, respectively. Plants of the same strains tended to occupy the same region on the plot. THC dominant strains (C3) mainly occupied the left side the plot and CBD dominant (C1) and intermediate strains (C2) occupied the lower right quadrant. The loading matrix in **Table 4.16** lists the compounds that contributed most to the separations along PC1 and PC2 with the absolute value of loadings equal to or greater than 0.45. PC1 was positively correlated with three cannabinoids (total CBD, total CBDV, total CBC), one monoterpenoid (1,8-cineole (eucalyptol)), four sesquiterpenoids (β-eudesmol, (-)-guaiol, α-eudesmol, α-bisabolol), three flavonoids (orientin, vitexin, isovitexin), three sterols (campesterol, stigmasterol, β-sitosterol), and one triterpenoid (epifriedanol), which were compounds identified as positively correlated with total CBD. PC1 was negatively correlated with one cannabinoid (total THC), four monoterpenoids (limonene, camphene, fenchol, and linalool), four sesquiterpenoids (α-humulene, β-caryophyllene, trans-nerolidol, and trans-β-farnesene,), four flavonoids (quercetin, kaempferol, and apigenin), and one triterpenoid (friedelin), which were compounds identified as positively correlated with total THC. THC dominant strains were scattered in both lower left quadrant and upper right quadrant along PC2. Compounds positively correlated with PC2 and negatively correlated with PC1 (PC1<0 and PC2>0), including total THC, total CBG, total THCV, α-terpineol, camphene, fenchol, linalool, ocimene, borneol, α-humulene, β-caryophyllene, trans-nerolidol, quercetin and kaempferol, were more abundant in THC dominant strains than those in CBD dominant and intermediate strains. β-Myrcene was negatively correlated with PC2 and positively correlated with PC1, which means it was more abundant in CBD dominant and intermediate strains. Two flavonoids, luteolin and apigenin, were negatively correlated with PC1 and PC2, and were more abundant in THC dominant strains in the left lower quadrant than other THC dominant strains. Although some compounds were more correlated with CBD, they may be more abundant in some THC dominant strains. For example, compounds positively correlated with PC2 and positively correlated with PC1, including orientin (L), vitexin (L), isovitexin (L), were more abundant in THC dominant strains in the upper right quadrant than strain in C1 and C2, even though these flavonoids were positively correlated with CBD. This may be the result of extensive strain crossing and hybridization. PCA using absolute values of the secondary metabolites are also shown in **Figure 4.10**. The full spectrum of secondary metabolites without total THC and total

CBD resulted in a similar PCA scatter plot where PC1 and PC2 explained 32.6% and 16.1% of the total variance, respectively (**Figure 4.11**).

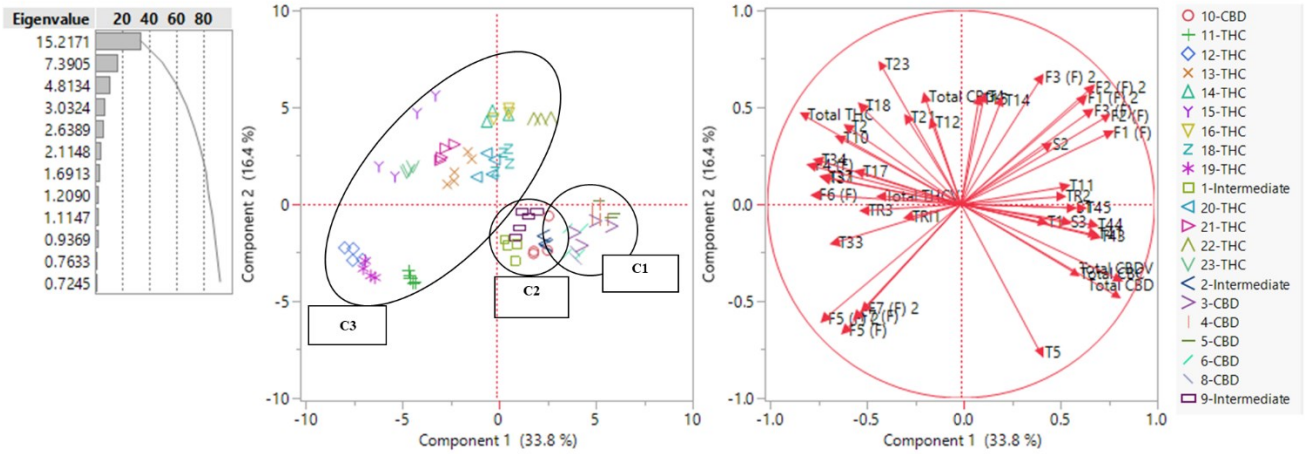


Figure 4.9 PCA scatter plot (left) and loading plot (right) using the full spectrum of secondary metabolites (in ratios) of 82 plants representing 21 strains.
Terpenoids are labelled with T and the number assigned in **Table 4.7**. Flavonoids are labelled as F and the number assigned in **Table 4.9**. Flavonoids quantified in inflorescences are labelled (F) and flavonoids in leaf are labelled (L). Sterols are labelled as S and the number assigned in **Table 4.11**. Triterpenoids are labelled as TRI and the number assigned in **Table 4.13**.

Table 4.16 Formatted loading matrix for PC1 and PC2.

PC1				PC2			
Compound	**Positive loadings**	**Compound**	**Negative loadings**	**Compound**	**Positive loadings**	**Compound**	**Negative loadings**
Total CBDV	0.82	Total THC	-0.81	α-Terpineol	0.72	β -Myrcene	-0.77
Total CBD	0.81	Quercetin (F)	-0.77	Isovitexin (L)	0.65	Luteolin (F)	-0.65
Orientin (F)	0.77	Kaempferol (F)	-0.75	Vitexin (L)	0.60	Luteolin (L)	-0.60
Vitexin (F)	0.76	α-Humulene	-0.74	β-Pinene	0.56	Apigenin (F)	-0.58
β-Eudesmol	0.70	Luteolin (L)	-0.71	Total CBG	0.55	Apigenin (L)	-0.55
α-Eudesmol	0.69	trans-Nerolidol	-0.71	Orientin (L)	0.55	Total CBD	-0.47
(-)-Guaiol	0.68	β-Caryophyllene	-0.70	Terpinolene	0.54		
Vitexin (L)	0.68	trans-β-Farnesene	-0.65	Sabinene Hydrate	0.53		
Isovitexin (F)	0.67	Limonene	-0.63	Fenchol	0.50		
Orientin (L)	0.64	Luteolin (F)	-0.60	Isovitexin (F)	0.46		
α-Bisabolol	0.62	Camphene	-0.59	Vitexin (F)	0.45		
Total CBC	0.59	Apigenin (F)	-0.54	Borneol	0.45		
Campesterol	0.57	Linalool	-0.53				
β-sitosterol	0.55	Fenchol	-0.52				
1,8-Cineole (Eucalyptol)	0.54	Apigenin (L)	-0.50				
Epifriedanol	0.52	Friedelin	-0.50				
Stigmasterol	0.45						

* Formatted loading matrix for PC1 and PC2 (only compounds with absolute loadings > 0.45 are listed)

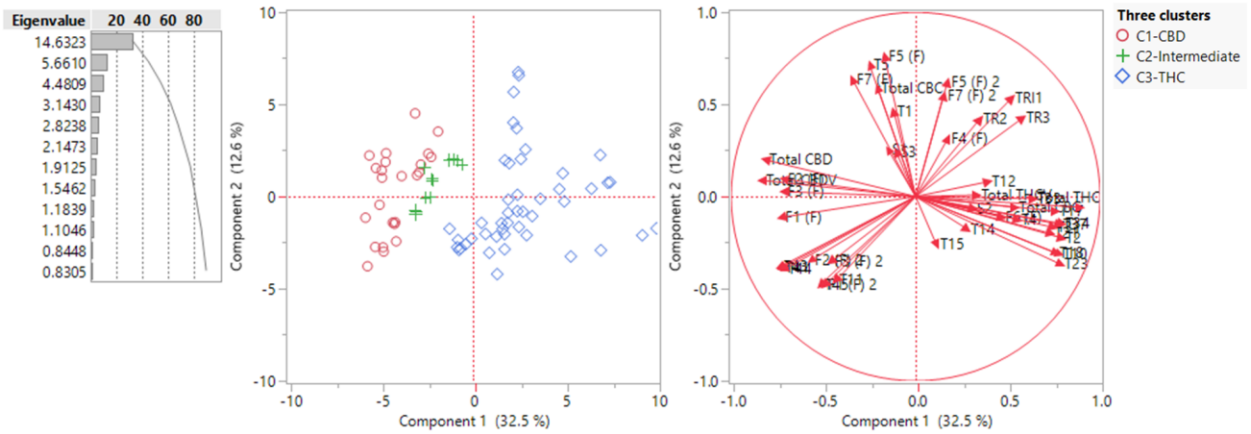


Figure 4.10 PCA scatter plot (left) and loading plot (right) using the full spectrum of secondary metabolites (absolute values) of 82 plants representing 21 strains.

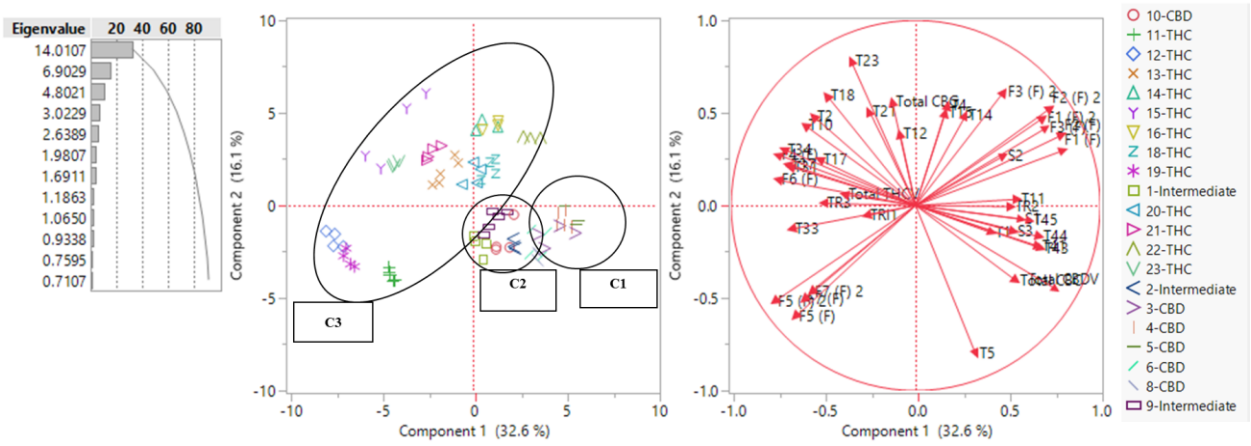


Figure 4.11 PCA scatter plot (left) and loading plot (right) using the full spectrum of secondary metabolites (using ratios) without total THC and total CBD.

4.4.5 Supervised canonical correlation analysis

The canonical correlation analysis of 82 plants showed good separation between the three chemotypes (**Figure 4.12**). Each plant was predicted to be in its originally preassigned cluster with 100% accuracy (**Table 4.17**). Canonical correlation analysis using the absolute values of 45 compounds were also investigated (**Figure 4.13**), with 100% accuracy in sorting each plant into its originally preassigned chemotypes. The full spectrum of secondary metabolites, absent total THC and total CBD, also predicted each plant to be in its originally preassigned cluster with 100% accuracy (**Figure 4.14**). However, the distance between three clusters were smaller along two

canonical axes due to reduced differences in the chemical profiles of three chemotypes after removing the THC and CBD data.

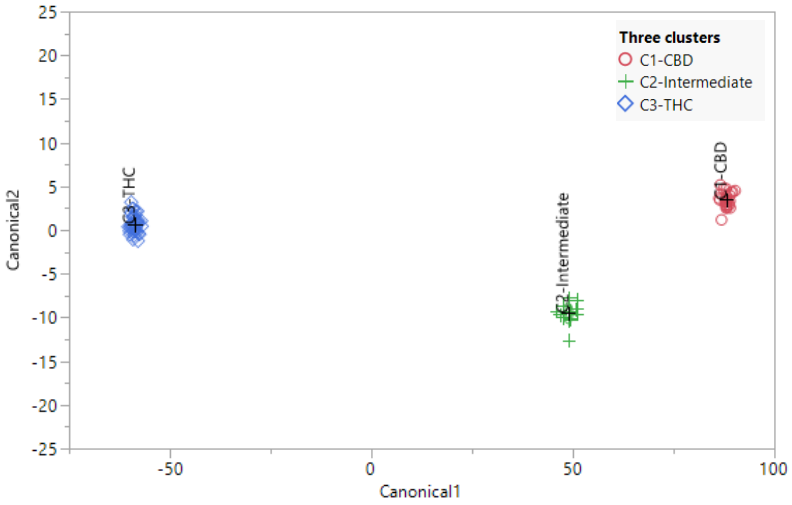


Figure 4.12 Canonical correlation analysis using the full spectrum of secondary metabolites (using ratios) of 82 plants representing 21 strains.
The plants were preassigned to three chemotypes in **Table 4.2**. The observations and the multivariate means of each group ("+") are represented as points on the biplot. An ellipse denoting a 50% contour is plotted for each group, that contains approximately 50% of the observations.

Table 4.17 Summary prediction of 82 plants into preassigned chemotypes using canonical correlation analysis (using ratios)

Preassigned	Predicted		
	C1-CBD	C2-Intermediate	C3-THC
C1-CBD	24	0	0
C2-Intermediate	0	12	0
C3-THC	0	0	46

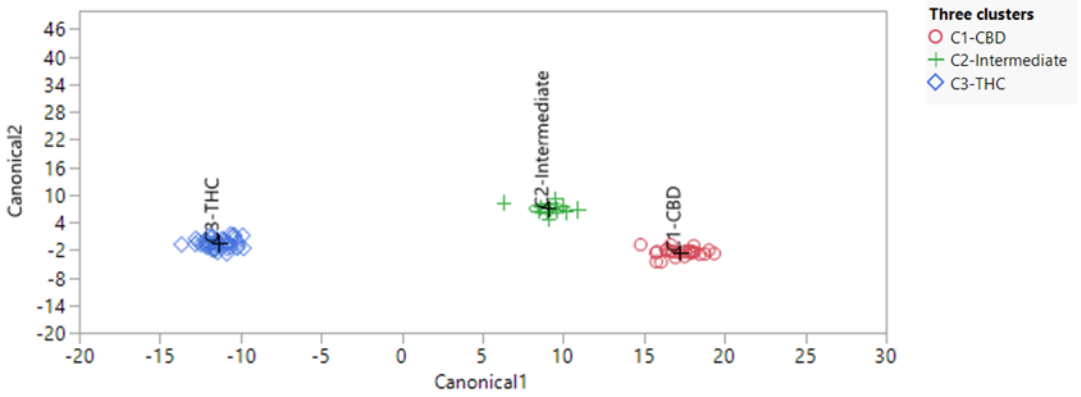


Figure 4.13 Canonical correlation analysis using the full spectrum of secondary metabolites (absolute values) of 82 plants representing 21 strains.

The plants were preassigned to three chemotypes in **Table 4.1**. The observations and the multivariate means of each group ("+") are represented as points on the biplot. An ellipse denoting a 50% contour is plotted for each group, that contains approximately 50% of the observations.

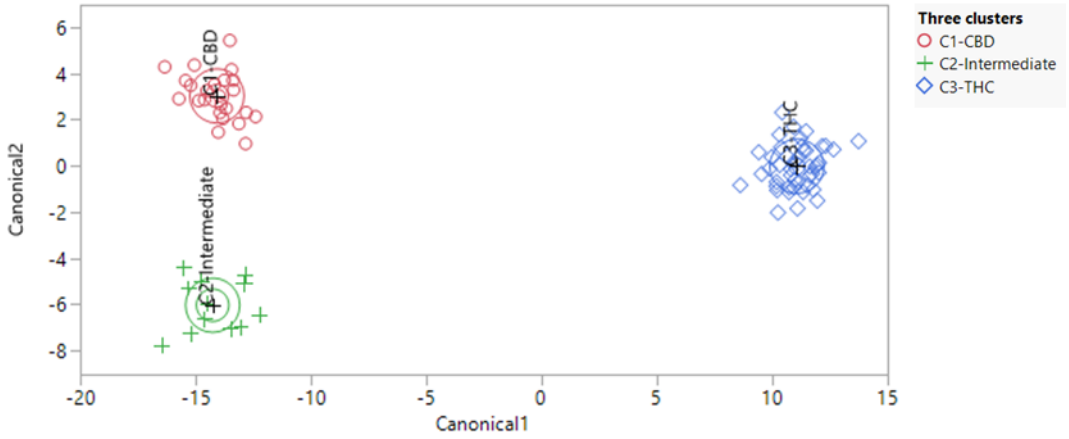


Figure 4.14 Canonical correlation analysis using the full spectrum of secondary metabolites (using ratios) without total THC and total CBD of 82 plants representing 21 strains.

The plants were preassigned to three chemotypes in **Table 4.1**. The observations and the multivariate means of each group ("+") are represented as points on the biplot. A 95% confidence level ellipse is plotted for each mean. An ellipse denoting a 50% contour is plotted for each group, that contains approximately 50% of the observations.

4.4.6 Identification of chemotypic markers for three chemotypes

Means (±SD), Tukey HSD multiple tests at the 0.05 significance level, and p value of one-way ANOVA of 45 quantifiable compounds (using ratios) for each of the three chemotypes are listed in **Table 4.18** and plotted in **Figure 4.15**. The largest number of significant differences (Tukey HSD multiple tests at the 0.05 significance level) was 37, which was between C1 and C3. The most similar pair was C1 and C2, with 14 significant differences. The number of significant differences between C2 and C3 was 23. Strains from C1 had significant higher amount of total CBD, total CBDV, total CBC, α-pinene, β-pinene, β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, α-bisabolol, orientin (F), vitexin (F), isovitexin (F), orientin (L), campesterol, stigmasterol, β-sitosterol, and epifriedanol than in strains of C3, which were all positively correlated with total CBD. Strains from C3 had significant higher amount of total THC, total THCV, total CBG, camphene, limonene, ocimene, linalool, fenchol, borneol, α-terpineol, β-caryophyllene, trans-β-farnesene, α-humulene, trans-nerolidol, quercetin (F), kaempferol (F), β-amyrin, and friedelin, which were all positively correlated with total THC. Most compounds in the C2 strains were at the same level with strains in C1 or C3 or at an intermediate level between C1 and C3.

Table 4.18 Means (±SD) of the ratios of 45 secondary metabolites above quantification limit for 82 plants assigned to C1-CBD dominant, C2-intermediate, and C3-THC dominant.

	Three chemotypes	C1-CBD	C2-Intermediate	C3-THC	ANOVA
	Plant count	N=24	N=12	N=46	p
Cannabinoids	Total CBDV	0.31% ± 0.09% a	0.24% ± 0.04% b	0.04% ± 0.04% c	<0.0001
	Total CBG	2.16% ± 0.55% b	2.45% ± 1.03% b	3.99% ± 2.41% a	0.0004
	Total CBD	78.20% ± 1.27% a	53.58% ± 2.17% b	0.37% ± 0.16% c	<0.0001
	Total THCV	5.42% ± 0.02% b	0.14% ± 0.04% b	0.94% ± 1.05% a	<0.0001
	Total THC	3.40% ± 0.49% c	28.38% ± 3.40% b	80.39% ± 2.41% a	<0.0001
	Total CBC	4.04% ± 0.52% a	3.45% ± 0.46% b	2.31% ± 0.97% c	<0.0001
Monoterpenoids	α-Pinene	14.41% ± 3.82% a	6.15% ± 6.01% b	8.10% ± 5.72% b	<0.0001
	Camphene	0.43% ± 0.08% b	0.51% ± 0.05% b	0.86% ± 0.40% a	<0.0001
	β-Pinene	5.94% ± 1.39% a	4.25% ± 1.61% b	6.16% ± 2.03% a	0.0067
	β-Myrcene	40.09% ± 7.23% a	41.53% ± 6.21% a	17.12% ± 13.39% b	<0.0001
	Limonene	7.18% ± 0.77% b	10.43% ± 1.60% b	17.88% ± 9.41% a	<0.0001
	1,8-Cineole (Eucalyptol)	0.58% ± 0.17% b	0.94% ± 0.60% a	0.33% ± 0.38% c	<0.0001
	Ocimene	1.22% ± 1.61% b	0.77% ± 0.44% b	4.89% ± 3.94% a	<0.0001
	Sabinene Hydrate	0.50% ± 0.18% a	0.55% ± 0.24% a	0.62% ± 0.38% a	0.3066
	Terpinolene	0.64% ± 0.70% a	1.03% ± 1.05% a	3.94% ± 7.51% a	0.0482
	Linalool	2.23% ± 0.86% b	4.00% ± 1.42% a	4.11% ± 2.05% a	0.0001
	Fenchol	1.16% ± 0.21% b	1.65% ± 0.41% ab	2.29% ± 1.10% a	<0.0001
	Borneol	0.47% ± 0.09% b	0.53% ± 0.11% ab	0.58% ± 0.20% a	0.0179
	α-Terpineol	1.48% ± 0.26% b	2.09% ± 0.42% b	3.09% ± 1.31% a	<0.0001
Sesquiterpenoids	β-Caryophyllene	3.63% ± 1.95% b	6.47% ± 3.28% b	12.27% ± 0.60% a	<0.0001
	trans-β-Farnesene	0.47% ± 0.23% b	0.50% ± 0.12% b	1.34% ± 1.23% a	0.0005
	α-Humulene	1.12% ± 0.65% b	1.91% ± 1.00% b	4.16% ± 1.69% a	<0.0001
	trans-Nerolidol	0.45% ± 0.18% b	0.44% ± 0.14% b	2.75% ± 1.95% a	<0.0001
	(-)-Guaiol	4.90% ± 2.02% a	4.94% ± 2.05% a	1.89% ± 1.76% b	<0.0001
	β-Eudesmol	2.83% ± 1.18% a	2.72% ± 0.99% a	1.04% ± 0.97% b	<0.0001
	α-Eudesmol	1.67% ± 0.71% a	1.40% ± 0.44% a	0.66% ± 0.62% b	<0.0001
	α-Bisabolol	8.01% ± 5.28% a	6.34% ± 2.59% a	3.40% ± 2.62% b	<0.0001
Flavonoids	Orientin (F)	38.61% ± 14.12% a	22.48% ± 2.63% b	20.31% ± 14.42% b	<0.0001
	Vitexin (F)	30.41% ± 3.87% a	26.72% ± 4.03% ab	19.50% ± 13.80% b	0.0004
	Isovitexin (F)	2.44% ± 0.44% a	1.97% ± 0.57% ab	1.61% ± 1.76% b	0.0034
	Quercetin (F)	5.42% ± 1.17% c	14.25% ± 5.35% b	19.81% ± 7.54% a	<0.0001
	Luteolin (F)	16.64% ± 9.76% a	22.96% ± 4.07% a	26.69% ± 21.34% a	0.0698
	Kaempferol (F)	2.49% ± 0.97% b	4.01% ± 0.89% b	6.53% ± 3.35% a	<0.0001
	Apigenin (F)	3.99% ± 1.92% b	7.61% ± 2.15% a	5.57% ± 4.25% ab	0.0144
	Orientin (L)	33.16% ± 19.50% a	20.22% ± 4.43% ab	22.45% ± 17.71% b	0.029
	Vitexin (L)	27.64% ± 8.80% a	23.88% ± 8.41% a	20.18% ± 15.02% a	0.0689
	Isovitexin (L)	1.81% ± 0.98% a	1.91% ± 0.83% a	1.68% ± 1.32% a	0.8075
	Luteolin (L)	26.08% ± 18.61% b	37.59% ± 8.40% ab	41.67% ± 24.33% a	0.017

	Three chemotypes	C1-CBD	C2-Intermediate	C3-THC	ANOVA
	Apigenin (L)	9.66% ± 7.72% a	12.08% ± 4.49% a	11.04% ± 9.22% a	0.6795
Sterols	Campesterol	8.64% ± 1.36% a	6.71% ± 1.08% b	6.56% ± 1.88% b	<0.0001
	Stigmasterol	8.04% ± 1.70% a	7.13% ± 0.80% ab	7.11% ± 1.80% b	0.0494
	β-sitosterol	28.98% ± 4.92% a	22.00% ± 2.52% b	21.09% ± 5.99% b	<0.0001
Triterpenoids	β-Amyrin	2.63% ± 0.37% b	3.02% ± 0.38% a	2.92% ± 0.40% a	0.0045
	Epifriedanol	36.60% ± 4.88% a	33.89% ± 1.44% ab	32.12% ± 4.54% b	0.0005
	Friedelin	60.78% ± 4.82% b	63.08% ± 1.55% ab	64.96% ± 4.52% a	0.0011

*Levels not connected by same letter are significantly different.

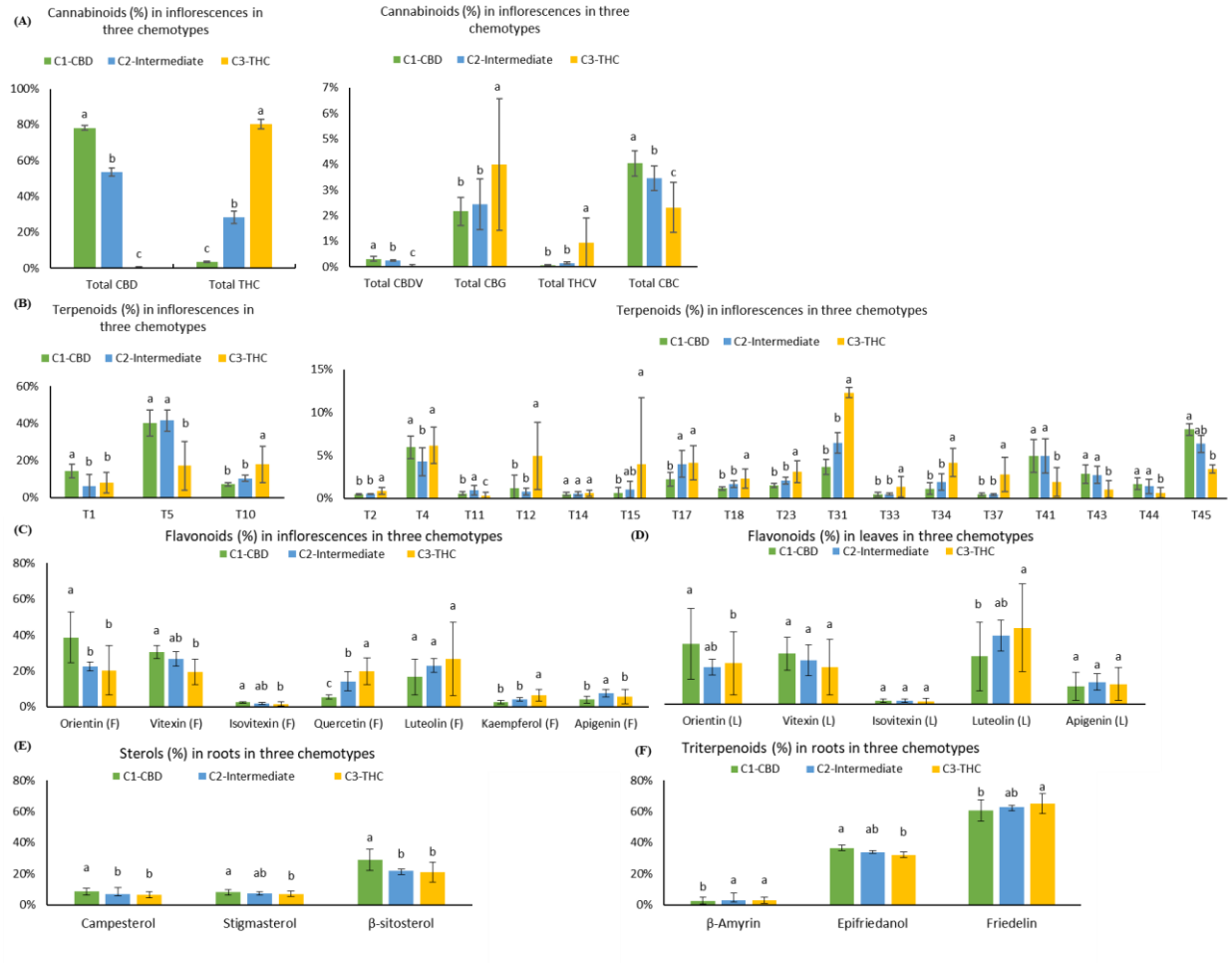


Figure 4.15 Means and standard deviations (±SD) of secondary metabolites in each plant part.

(A) Cannabinoids in inflorescences, (B) mono- and sesquiterpenoids in inflorescences, (C) flavonoids in inflorescences, (D) flavonoids in leaves, (E) sterols in roots, and (F) triterpenoids in roots (in ratios) for each of the three chemotypes C1 – CBD dominant, C2 – intermediate, and C3 – THC dominant. Cluster means were expressed as mean ± SD. *Levels not connected by the same letter are significantly different by Tukey HSD multiple tests at the 0.05 significance level.

Means ±SD, Tukey HSD multiple tests at the 0.05 significance level, and p value of one-way ANOVA of the absolute values of 45 compounds for each cluster were summarized in **Table 4.19**. The largest number of significant differences was 38, which was between C1 and C3. The most similar pair was C1 and C2, with 10 differences. The number of significant differences between C2 and C3 was 23. Cannabinoids, terpenoids, flavonoids, sterols, and triterpenoids that were significantly higher in C1, C2, and C3 were similar to those identified using ratios.

Table 4.19 Means (±SD) of the absolute values of 45 secondary metabolites (mg/mg%) for 82 plants assigned to C1-CBD dominant, C2-intermediate, and C3-THC dominant.

	Three chemotypes	C1-CBD	C2-Intermediate	C3-THC	ANOVA
	Plant count	N=24	N=12	N=46	p
Cannabinoids	Total CBDV	0.042% ± 0.009% a	0.037% ± 0.010% a	0.005% ± 0.006% b	<0.0001
	Total CBG	0.303% ± 0.100% b	0.380% ± 0.212% b	0.682% ± 0.374% a	<0.0001
	Total CBD	10.915% ± 1.686% a	8.049% ± 1.575% b	0.059% ± 0.019% c	<0.0001
	Total THCV	0.007% ± 0.002% b	0.022% ± 0.009% b	0.171% ± 0.203% a	<0.0001
	Total THC	0.471% ± 0.080% c	4.208% ± 0.665% b	13.797% ± 3.750% a	<0.0001
	Total CBC	0.566% ± 0.127% a	0.516% ± 0.120% ab	0.392% ± 0.200% a	0.0003
Monoterpenoids	α-Pinene	0.187% ± 0.067% a	0.083% ± 0.085% b	0.130% ± 0.085% b	0.0010
	Camphene	0.006% ± 0.002% b	0.007% ± 0.001% b	0.015% ± 0.009% a	<0.0001
	β-Pinene	0.077% ± 0.026% b	0.056% ± 0.025% b	0.104% ± 0.039% a	<0.0001
	β-Myrcene	0.516% ± 0.143% a	0.548% ± 0.160% a	0.297% ± 0.228% b	<0.0001
	Limonene	0.092% ± 0.019% b	0.134% ± 0.023% b	0.326% ± 0.223% a	<0.0001
	1,8-Cineole (Eucalyptol)	0.007% ± 0.002% b	0.012% ± 0.007% a	0.005% ± 0.005% b	<0.0001
	Ocimene	0.015% ± 0.018% b	0.009% ± 0.005% b	0.077% ± 0.061% a	<0.0001
	Sabinene Hydrate	0.006% ± 0.002% b	0.007% ± 0.002% b	0.010% ± 0.004% a	0.0007
	Terpinolene	0.008% ± 0.009% a	0.012% ± 0.012% a	0.063% ± 0.123% a	0.0411
	Linalool	0.028% ± 0.008% b	0.052% ± 0.019% ab	0.078% ± 0.063% a	0.0003
	Fenchol	0.015% ± 0.003% b	0.021% ± 0.004% b	0.041% ± 0.028% a	<0.0001
	Borneol	0.006% ± 0.001% b	0.007% ± 0.001% b	0.010% ± 0.005% a	<0.0001
	α-Terpineol	0.019% ± 0.004% b	0.027% ± 0.004% b	0.054% ± 0.027% a	<0.0001
Sesquiterpenoids	β-Caryophyllene	0.045% ± 0.022% b	0.081% ± 0.039% b	0.223% ± 0.163% a	<0.0001
	trans-β-Farnesene	0.006% ± 0.003% b	0.006% ± 0.001% b	0.025% ± 0.028% a	0.0006
	α-Humulene	0.014% ± 0.007% b	0.024% ± 0.012% b	0.075% ± 0.048% a	<0.0001
	trans-Nerolidol	0.006% ± 0.002% b	0.006% ± 0.002% b	0.052% ± 0.048% a	<0.0001
	(-)-Guaiol	0.062% ± 0.025% a	0.063% ± 0.023% a	0.028% ± 0.026% b	<0.0001
	β-Eudesmol	0.036% ± 0.015% a	0.034% ± 0.011% a	0.015% ± 0.015% b	<0.0001
	α-Eudesmol	0.021% ± 0.009% a	0.018% ± 0.005% a	0.010% ± 0.009% b	<0.0001
	α-Bisabolol	0.100% ± 0.061% a	0.080% ± 0.028% ab	0.053% ± 0.030% b	<0.0001
Flavonoids	Orientin (F)	0.051% ± 0.022% a	0.021% ± 0.008% b	0.014% ± 0.012% b	<0.0001

	Three chemotypes	C1-CBD	C2-Intermediate	C3-THC	ANOVA
	Vitexin (F)	0.042% ± 0.018% a	0.024% ± 0.006% b	0.013% ± 0.011% c	<0.0001
	Isovitexin (F)	0.003% ± 0.001% a	0.002% ± 0.001% b	0.001% ± 0.001% b	<0.0001
	Quercetin (F)	0.008% ± 0.004% b	0.014% ± 0.008% a	0.012% ± 0.006% a	0.0012
	Luteolin (F)	0.027% ± 0.023% a	0.021% ± 0.006% a	0.018% ± 0.021% a	0.2290
	Kaempferol (F)	0.0030% ± 0.0004% b	0.003% ± 0.001% ab	0.004% ± 0.001% a	0.0156
	Apigenin (F)	0.006% ± 0.004% a	0.007% ± 0.001% a	0.003% ± 0.002% b	<0.0001
	Orientin (L)	0.077% ± 0.064% a	0.044% ± 0.032% ab	0.038% ± 0.039% b	0.0061
	Vitexin (L)	0.061% ± 0.036% a	0.053% ± 0.036% ab	0.032% ± 0.026% b	0.0010
	Isovitexin (L)	0.004% ± 0.003% a	0.004% ± 0.003% ab	0.002% ± 0.002% b	0.0098
	Luteolin (L)	0.050% ± 0.040% a	0.074% ± 0.046% a	0.074% ± 0.068% a	0.2586
	Apigenin (L)	0.017% ± 0.012% a	0.021% ± 0.008% a	0.016% ± 0.011% a	0.5547
Sterols	Campesterol	0.013% ± 0.001% a	0.013% ± 0.001% a	0.012% ± 0.002% a	0.1279
	Stigmasterol	0.012% ± 0.002% b	0.013% ± 0.001% ab	0.013% ± 0.002% a	0.0361
	β-Sitosterol	0.043% ± 0.006% a	0.042% ± 0.004% ab	0.039% ± 0.007% b	0.0169
Triterpenoids	β-Amyrin	0.004% ± 0.001% b	0.006% ± 0.001% a	0.006% ± 0.001% a	<0.0001
	Epifriedanol	0.055% ± 0.010% a	0.064% ± 0.005% a	0.062% ± 0.014% a	0.0477
	Friedelin	0.094% ± 0.024% b	0.120% ± 0.011% a	0.127% ± 0.034% a	0.0001

*Levels not connected by same letter are significantly different.

Although numerous significant differences in compounds were found amongst CBD dominant, intermediate, and THC dominant strains, the group means of some compounds differed by less than a factor of two. In addition, some compounds may be significantly different qualitatively in ratios but not quantitatively in absolute values. For example, all three sterols (campesterol, stigmasterol, and β-sitosterol), were significantly higher in roots of CBD dominant strains than in THC dominant strains by ratios (one-way ANOVA $p < 0.0001$, $p = 0.1279$, and $p < 0.0001$, respectively), but they were not significantly different by absolute values (one-way ANOVA $p = 0.1279$, $p = 0.0361$, and $p = 0.0169$, respectively). Compounds significantly different (one-way ANOVA $p < 0.05$) with two or more than two-fold higher in terms of both ratios and absolute values in the identified clusters than in the clusters with the lowest values were selected as chemotypic markers. These included three cannabinoids (total CBD, total CBDV, and total CBC), six terpenoids (α-pinene, β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, and α-bisabolol), and three flavonoids (orientin, vitexin, and isovitexin) for CBD dominant strains, three cannabinoids (total THC, total THCV, and total CBG), twelve terpenoids (camphene, limonene, ocimene, sabinene hydrate, terpinolene, linalool, fenchol, α-terpineol, β-caryophyllene, trans-β-farnesene, α-humulene, and trans-nerolidol), and two flavonoids (quercetin and kaempferol) for THC

dominant strains. Intermediate strains are more similar to CBD dominant strains than THC dominant strains with higher amounts of β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, and α-bisabolol. There are more mono- and sesquiterpenoids that are significantly higher in the THC dominant cluster than in the CBD dominant and intermediate clusters. The simultaneous presence of a collection of compounds can be used to differentiate types of plants.

4.5 Discussions

4.5.1 Cannabinoids as chemotypic markers

In this study, the average THC to CBD ratios in the three chemotypes were 247 ± 79, 0.5 ± 0.1, and 0.04 ± 0.01, respectively. These ratios showed that THC levels in THC dominant strains were greater than CBD levels in CBD dominant strains. This bias towards higher THC is due to the long history of extensive hybridization for recreational purposes [69]. A THC/CBD ratio of 247:1 in THC dominant strains matched with those in "Sativa" and "Indica" strains that were almost devoid of CBD [70–72,83,149,204]. Due to CBD's therapeutic potential without psychoactive effects [96,231–234], breeding for high CBD concentrations began only recently by integrating hemp-type CBD acid synthase gene clusters into a background of drug-type cannabis to elevate CBDA production [197,216]. The CBD to THC ratios in intermediate strains were similar to 1.8:1 in our previously reported values [204], and also matched with the reported cannabinoid profile of intermediate strains available in the database. These intermediate strains may have been created by crossing purebred THC dominant types with CBD dominant types [42]. Chemotaxonomic research in minor cannabinoids of the three chemotypes are sparse in the current literature. In this study, minor cannabinoids were mostly less than 1% in all three chemotypes and several minor cannabinoids were more abundant in one chemotypes relative to others.

4.5.2 Mono- and sesquiterpenoids as chemotypic markers

In general, sesquiterpenoids are considered as more stable markers because monoterpenoids are more volatile [69]. In this study, (-)-guaiol, β-eudesmol, α-eudesmol, and α-bisabolol were identified as chemotypic markers in CBD and intermediate strains. These compounds were also noted by Hillig as signature peaks on chromatograms for pre-hybridization Afghani WLD landraces [27] and modern "Indica" dominant strains (WLD), but were present in lower amounts in pre-hybridization NLD landraces and modern "Sativa" dominant strains (NLD) [70,72]. CBD dominant strains and pre-hybridization Afghani WLD landraces are similar in that they both have elevated CBD

concentrations compared to their THC dominant counterparts. According to the correlation analysis in this study, these chemotypic markers for CBD dominant strains and intermediate strains may be related to CBD production. For modern "Indica" dominant strains (WLD), which are nearly devoid of CBD, even though these sesquiterpenoids were considered to be inherited from their WLD landrace ancestors despite selection for elevated THC/CBD ratios, these compounds were detected only in trace amounts [70–72,83,149]. In this study, terpinolene, β-caryophyllene, and trans-β-farnesene, were identified as chemotypic markers in THC dominant strains. These compounds were also noted by Hillig as signature peaks on chromatograms for pre-hybridization NLD landraces [27] and modern "sativa" dominant strains (NLD), but were present in lower amounts in pre-hybridization WLD landraces and modern "Indica" dominant strains (WLD) [70,72]. THC dominant strains and pre-hybridization NLD landraces both have elevated THC concentrations and are almost devoid of CBD. These chemotypic markers for THC dominant strains and intermediate strains may be correlated with THC production when CBD is not produced.

Studies have shown that terpenoids in cannabis are derived from two pathways: the plastidial methylerythritol phosphate (MEP) pathway and the cytosolic mevalonate (MVA) pathway [93,218,219]. Geranyl diphosphate (GPP) is typically derived from the MEP pathway and is the precursor for cannabinoid and monoterpenoid biosynthesis. Farnesyl diphosphate (FPP) is commonly produced from MVA pathway and is the precursor for sesquiterpenoids, triterpenoids and sterols. Although it is hypothesized that the identified chemotypic markers may be related to CBD or THC production, currently there are no biomedical studies on these correlations. Future studies are needed on the biochemical relationship between CBD or THC production and individual terpenoid production.

Of the strains with a reported Sativa/Hybrid/Indica ancestry label, CBD dominant strains contained two "Sativa" strains, intermediate strains contained one "Sativa" strain and one "Indica" strain, and THC dominant strains contained ten "Indica" strains and one "50/50 hybrid" strain. Based on the reported ancestry, the results of this study seem to contradict other studies. The terpenoids markers in CBD dominant strains (reported as "Sativa" due to narrow leaflets) were similar to those identified in "Indica" dominant strains but different from those identified in "Sativa" dominant strains in other studies [70–72,83,149]. Similarly, the terpenoids markers in THC dominant strains (reported as "Indica" due to wide leaflets) were similar to those identified in "Sativa"

dominant strains but different from those identified in “Indica” dominant strains in other studies. These conflicting results reflects the unreliability of the vernacular “Sativa” and “Indica” categories, which are based on the visual determination of leaflet shape, often with no reference data for categorization [235]. This may lead to mixed results in separating modern strains genetically or chemically [75,236]. Another explanation for the discrepancy is that instead of separating "Sativa" vs "Indica", which are often THC dominant strains, this paper focused on the differentiation between three chemotypes. Because no “Sativa” strains were reported for THC dominant strains in this study, whether (-)-guaiol, β-eudesmol, α-eudesmol, and α-bisabolol are more abundant in “Indica” dominant strains and terpinolene, β-caryophyllene, and trans-β-farnesene are more abundant in “Sativa” dominant strains as described in other studies could not be verified.

4.5.3 Flavonoids as chemotypic markers

Flavonoid variation in cannabis was investigated by Clark and Bohm in 1979, the only such study that used flavonoids for chemotaxonomy and for supporting a two-species hypothesis: where luteolin was more often detected in *C. sativa* L. but not in *C. indica* Lam. [157]. There have yet to be chemotaxonomic studies of flavonoids across the three cannabis chemotypes. We found that orientin, vitexin, and isovitexin were the signature flavonoids of CBD dominant strains, and quercetin and kaempferol were detected only in inflorescences and tended to be higher in THC dominant strains.

4.5.4 Sterols and triterpenoids as chemotypic markers

The role of sterols and triterpenoids in the chemotaxonomy of cannabis have not yet been investigated. In this study, CBD dominant strains had significantly higher ratios of three sterols, but they differed by less than a factor of two and may not provide a firm basis for chemotaxonomic distinction. Similarly, for triterpenoids, although the ratio of epifriedanol was higher in CBD dominant strains and friedelin was higher in THC dominant strains, the differences were not sufficiently large for these compounds to be used as chemotype markers.

4.5.5 The potential of developing holistic cannabis-based products and medications

Because cannabinoids are concentrated in cannabis inflorescences, cannabis leaves, stems, and roots are normally discarded by cannabis growers. However, in traditional Chinese medicine, cannabis leaves were used for treating conditions such as malaria, panting, roundworm, scorpion

stings, hair loss, greying of hair. Cannabis stem bark were used for strangury and physical injury. Cannabis roots were used for gout, arthritis, joint pain, fever, skin burns, hard tumors, childbirth, and physical injury [115–117]. Their traditional uses may serve as points of reference for investigating the medical potential of what is currently a byproduct or plant waste.

To link the traditional therapeutic uses for each part with the chemistry, we had identified the major groups of compounds in each plant part for correlation with benefits described in the literature. Cannabinoids, including THC, CBD, CBG, CBC, THCV, CBN, and CBDV, in both acid and neutral forms all have broad therapeutic potential, including anti-inflammatory [128,130,177,237–239], analgesic [95,130,240], anticonvulsant [241–243], antioxidant, and neuroprotective properties [244]. Increasing numbers of studies have shown that minor cannabinoids significantly contribute to the variance among cannabis extract, which further alter or enhance targeted therapeutic effects comparing to pure THC or CBD alone [223,224].

Terpenoids are widely distributed in highly fragrant fruits, plants, and herbs and they have anti-inflammatory [245,246], antirheumatic [247], pain relieving [246,248], anti-oxidant and neuroprotective [249], gastroprotective [250,251], and larvicidal properties [252]. If a cannabinoid-terpenoid entourage effect exists, it may not be at the CB1 or CB2 receptor level, but rather the terpenoids may act at different molecular targets in neuronal circuits [253].

Flavonoids share a wide range of biological effects with cannabinoids and terpenoids, including anti-inflammatory [166,167,254–256], antirheumatic [257–259], analgesic [166,260], and antioxidant and neuroprotective properties [166,168,256,261–263]. Ginkgo leaves are one of the prominent sources of flavonoids, with 0.4% total flavonoids in terms of total aglycones [161]. In this study, the mean of total flavonoids was 0.19% ± 0.09%, which makes cannabis leaves a promising source for flavonoids extraction.

Sterols and triterpenoids are mainly present in cannabis stem bark and roots. Friedelin is the most abundant and most studied triterpenoids in cannabis, and has anti-inflammatory, antioxidant, estrogenic, anti-cancer, and liver protectant properties [115]. β-sitosterol, stigmasterol, and campesterol are the most abundant phytosterols in the human diet. Phytosterols are widely recognized as lowering the levels of low-density lipoprotein cholesterol [135,136]. They are also studied for anti-inflammatory, anti-oxidant, and pain relieving properties [137].

These groups of identified bioactive compounds may underpin the traditional applications indicated for each plant part, but most of the therapeutic properties for these individual compounds have been studied in other herbal medicine and not in cannabis. The pharmaceutical values and the potential synergies of these bioactive compounds need to be directly investigated using cannabis material. Well-designed clinical studies are necessary to convert each part of the cannabis plant into evidence-based medicine. The chemotypic markers identified in this study will facilitate strain selection in research and clinical studies when the optimal combination of the chemical compounds is determined for treating certain conditions.

4.6 Conclusions

The chemical variation in CBD dominant and intermediate strains has yet to be studied or compared to THC dominant strains in the literature. This comprehensive chemotaxonomic investigation profiled cannabinoids, terpenoids, flavonoids, sterols, and triterpenoids in inflorescences, leaves, stem bark, and roots in 82 plants of 21 cannabis strains. These chemical data were subjected to correlation analysis, unsupervised clustering analysis (hierarchical clustering and PCA) and supervised canonical correlations analysis. In unsupervised clustering, 82 plants were clustered in accordance with their chemotypes. Canonical correlation analysis classified 82 plants into three chemotypes with 100% accuracy using full spectrum of secondary metabolites. Numerous significant differences that could be used as chemotypic markers were found amongst CBD dominant, intermediate, and THC dominant strains. These identified compounds were largely consistent with results from correlation analysis, hierarchical clustering, PCA, and by comparing concentration and ratio averages between chemotypes. At each step of the clustering analysis, it was found that secondary metabolites without total THC and total CBD could continue to sort strains into their defined chemotypes and achieve the same clustering results. This demonstrated that the clustering results were not solely driven by THC and CBD content or ratio, and that other metabolites can be used as chemotypic markers. However, the robustness of these markers should be tested in different growing environments to truly elucidate the chemical differences in terms of chemotypes or intra-chemotype sub-clusters. The results of this study provide a proof-of-concept for further collaboration between academia and the industry for leveraging chemotypic markers in medical studies and clinical trials.

Supplementary Table 4.1 Secondary metabolites in all plant parts (absolute values)

(Deposited in ERA: https://doi.org/10.7939/r3-eq07-rp47)

Supplementary Table 4.2 Secondary metabolites used in correlation analysis and classification analysis

(Deposited in ERA: https://doi.org/10.7939/r3-eq07-rp47)

Chapter 5 Identification of Phenotypic Characteristics in Three Chemotype Categories in the Genus *Cannabis*

5.1 Abstract

Modern *Cannabis* cultivars are morphologically distinguished by their leaflet shapes (wide for "Indica" and narrow for "Sativa") by users and breeders. However, there are no scientific bases or references for determining the shape of these leaflets. In addition, these two categories contained mostly THC dominant (high THC) cultivars while excluded CBD dominant (high CBD) and intermediate (intermediate level of both THC and CBD) cultivars. This study investigated the phenotypic variation in 21 *Cannabis* cultivars covering three chemical phenotypes, referred to as chemotypes, grown in a commercial greenhouse. Thirty morphological traits were measured in the vegetative, flowering, and harvest stages on live plants and harvested inflorescences. The collected data were subjected to correlation analysis, hierarchical clustering, principal component analysis, and canonical correlation analysis with preassigned chemotypes. Canonical correlation analysis assigned individual plants to their chemotypes with 92.9% accuracy. Significant morphological differences were identified. Traits usable as phenotype markers for CBD dominant cultivars included light-green and narrow leaflets, a greater number of primary and secondary serrations, loose inflorescences, dense and resinous trichomes, and *Botrytis cinerea* resistance. Traits for intermediate cultivars included deep green and medium-wide leaflets, more primary and secondary serrations, medium compact inflorescences, trichomes that are less dense and less resinous, and *Botrytis cinerea* resistance. Traits for THC dominant cultivars included deep-green and wide leaflets, large and compact inflorescences, dense and resinous trichomes, and *Botrytis cinerea* susceptibility. The results of this study provide a comprehensive profile of morphological traits of modern *Cannabis* cultivars and provides the first such profile for CBD dominant and intermediate cultivars. Additionally, this study included the traits of inflorescences, which have not been compared between three chemotypes in the literature. Phenotype markers identified in this study can facilitate preliminary cultivar identification and selection on live plants prior to or as a supplement to chemical and genetic analysis.

(**Published**) Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Phenotypic Characteristics in Three Chemotype Categories in the Genus Cannabis. *HortScience* **1**, 1–10 (**2021**).

5.2 Introduction

Cannabis is an annual, normally dioecious, flowering plant with staminate plants tending to be taller than pistillate plants [79]. The height can vary from 0.2 m to 6 m, with most of the plants reaching between 1 m to 3 m. Plant stems are erect, hollow, and grooved. *Cannabis* plant is taprooted, but taproots do not develop on vegetatively propagated plants. *Cannabis* plant has palmate leaves with 3 to 9 linear-lanceolate leaflets with serrations. Each female flower has one ovary encapsulated by bract and bracteoles (alternately called a calyx), out of which projects two long stigmas. Bract, bracteoles, stigmas, small leaves that grow out of inflorescences leaves are densely covered by capitate stalked trichomes, where most cannabinoids and terpenoids are biosynthesized and stored. Trichomes are denser in female plants and decrease in density from inflorescences, leaves, stems, to roots (which are devoid of cannabinoids). *Cannabis* can host disease-causing pathogens, including *Botrytis cinerea*, which causes gray mold in the inflorescences [216,264–267].

Whether genus *Cannabis* is monotypic or polytypic is still a debate. Although several putative *Cannabis* species were discovered and proposed for distinguishing *Cannabis*, only two were widely accepted: *C. sativa* and *C. indica*. Linnaeus described *C. sativa* L. in *Species Plantarum* [28], with loose inflorescences covered with sparse trichomes and resembling a northern European fiber-type landrace [69]. Later in 1785, de Lamarck described a second (or sub-) species, *C. indica* Lam., which was collected in India, with dense trichomes, narrower leaflets, branching habitus, poorer fibre quality, harder stem, and thinner cortex, but stronger psychoactive effects [29]. Schultes travelled to Afghanistan in 1971 and described *C. indica* as having broad leaflets, densely branched with very dense inflorescences for hashish (resin) production, which deviated from Lamarck's original taxonomic concept [31]. Anderson drew illustrations of *C. indica* and *C. sativa*, the former represented as short, conical, densely branched, with broad leaflets and the latter as relatively tall, laxly branched, with narrow leaflets [32], which aligned with Schultes but differed from Lamarck. *Cannabis* can be assigned as one of three chemotypes based on THC and CBD content [59]. Chemotype I is THC dominant, with more than 0.3% THC and less than 0.5% CBD. Chemotype II is intermediate, with high contents of both CBD (more than 0.5% THC) and THC (more than 0.3% THC). Chemotype III is CBD dominant with less than 0.3% THC. This quantitative approach was further developed into a qualitative measure using THC/CBD ratios: chemotype I has THC/CBD > 1, chemotype II has THC ≈ CBD, and chemotypes III has THC/CBD < 1 [42,221,222].

Hillig carried out genetic, chemical, and morphological analysis on 157 accessions of diverse geographic origin before large scale hybridization, classifying them into two species, *C. sativa* and *C. indica*, and seven putative taxa, including *C.indica* narrow-leaflet drug (NLD) biotype, *C.indica* wide-leaflet drug (WLD) biotype, *C.indica* hemp biotype, *C.indica* feral biotype, *C. sativa* hemp biotype, *C. sativa* feral biotype, and putative ruderal populations [268,269,73,270,27]. NLD biotype included landraces of Indian heritage (including varieties of the Indian subcontinent, Africa, and other drug producing regions), corresponding to Lamarck's *C. indica*. WLD biotype included landraces from Afghanistan and Pakistan, corresponding with Schultes's *C. indica*. *C. indica* hemp biotype included landraces from southern and eastern Asia while *C. sativa* hemp biotype included landraces from Europe, Asia Minor, and Central Asia.

By the end of 1980, nearly all drug-type cannabis cultivated in the USA, Canada, and Europe are cross-bred to achieve high THC content cultivars, called "sinsemilla" (meaning seedless) [69]. Cannabis breeders and users use vernacular "Sativa" to describe cultivars with narrow leaflets and "Indica" for cultivars with broad or wide leaflets, based on illustrations of Anderson which deviated from the original botanical nomenclature [69]. Even so, researchers have tried to differentiate these two categories genetically and chemically [75,76,86]. However, these vernacular categories are unreliable for medical applications due to extensive cross-breeding and unreliable labelling during unrecorded hybridization [69]. CBD dominant and intermediate varieties were also excluded from these studies despite getting increasing attention from the therapeutic potential of CBD [183–186], especially the indication of regulatory-approved prescription CBD (marketed as Epidiolex®) to treat epilepsy [271–273]. In addition to delimiting "Sativa" and "Indica" plants, recent studies tried to differentiate three chemotypes by profiling secondary metabolites (Jin et al., 2020), developing genetic markers [274–278], and comparing sequence and copy number variation of THC acid synthase and CBD acid synthase [279,280].

Although phenotypic differences are essential for delimiting plant species, this aspect of the plants has been largely ignored or limited when studying modern *Cannabis* cultivars. The description of leaflet shapes, if any, were determined visually and subjectively from sample providers without quantitative measures. There are no tangible data on determining what leaves are narrow leaflet and what leaves are wide leaflet, or whether an intermediate category exists between them due to hybridization. Lastly, plant morphology includes both qualitative and quantitative traits.

Qualitative traits are ratios of two measurements, for example, the ratio of the width and length of a central leaflet on a node. Quantitative traits are absolute measurements, such as plant height and yield. Qualitative traits are usually determined by a single genetic locus, while quantitative traits usually result from interactions between several genes and environmental variables [69,270]. In order to study the phenotypic variation on plants, it is necessary to grow them in a single location under identical conditions to control environmental variation (Small et al., 1976).

The objectives of this study were to:

1. Investigate whether modern *Cannabis* cultivars, including CBD dominant, intermediate, and THC dominant cultivars, can be differentiated using morphological traits, and

2. Identify qualitative phenotypic markers, supplied with quantitative markers, that can be leveraged to select and distinguish chemotypes.

5.3 Materials and Methods

5.3.1 Plant material

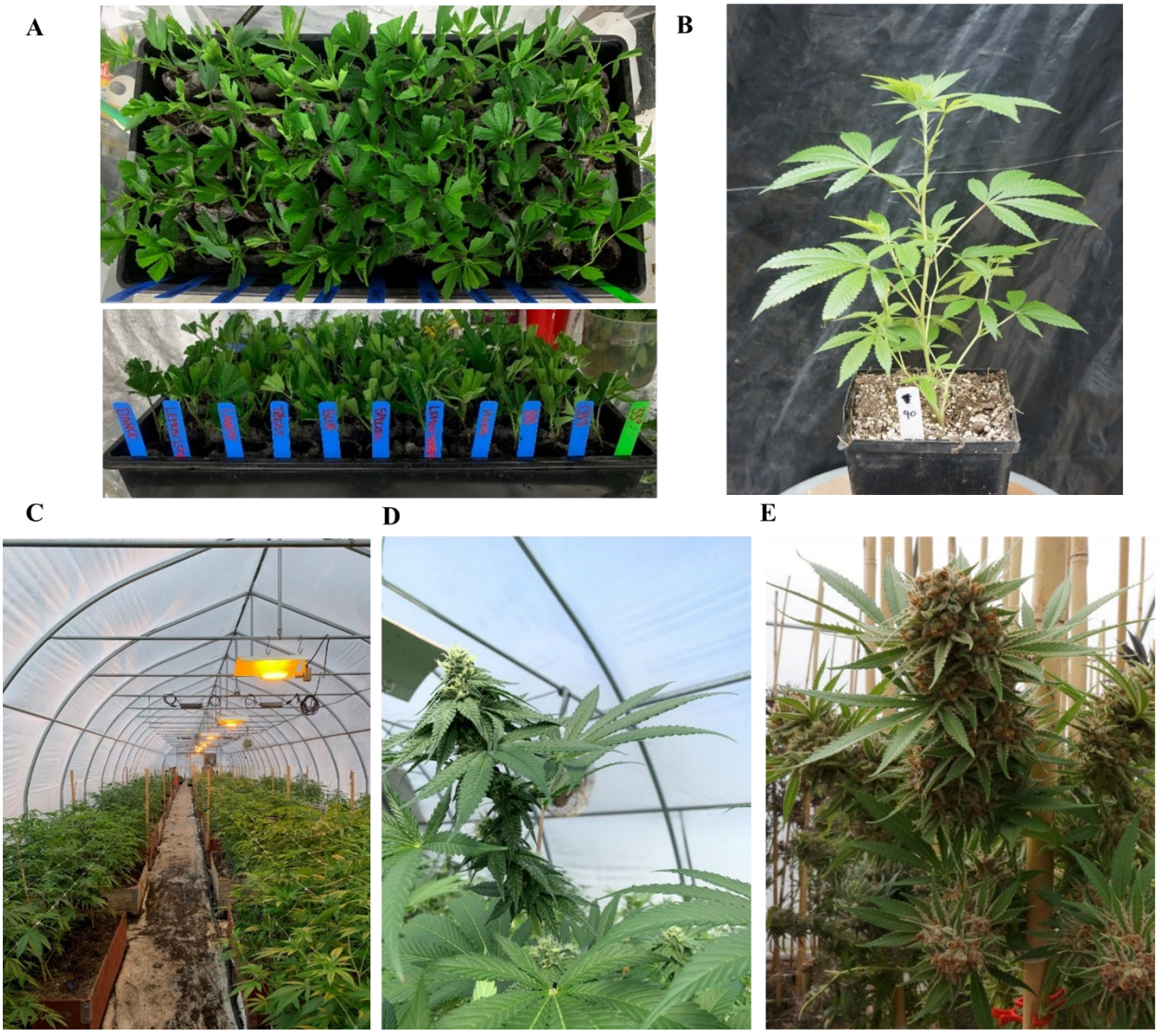


Figure 5.1 Cannabis plants grown in a greenhouse from rooting to flowering.
(A) Five clones per cultivar rooted in Jiffy 7 plugs. (B) Vegetative traits were measured 40 days after rooting. (C) After two months of vegetative growth, plants were transplanted into planters with 12 plants per planter. (D) Light regime changed to 12 hours per day and flowers began to grow. (E) After two months of flowering, plants were ready for harvest.

In this project, 23 commercially available cultivars were grown in a commercial greenhouse (**Figure 5.1**) under a research license issued by Health Canada. Plants for two cultivars were not rooted successfully and were excluded in the study. Where possible, the reported ancestry ("Sativa", "Indica", or "Sativa-dominant" and "Indica-dominant") was obtained from the Leafly online database (https://www.leafly.ca/) or from the licensed producer providing the cultivar (**Table 5.1**). Each cultivar was analyzed for chemical composition using methods established in previous work [204] and labelled as "THC dominant", "CBD dominant", or "intermediate". Five

cuttings per cultivar were propagated in Jiffy 7 Peat Pellets Seed Starting Plugs (Jiffy, Pokemouche, NB, Canada) under SunBlaster T5HO fluorescent lighting (SunBlaster Holdings ULC, Langley, BC, Canada) and a 24-hour light photoperiod for two weeks. 85 successful rooted plants were then transplanted to 6-inch pots with Dark Matter Super Soil (Destiny Grow Systems, Grand Forks, BC, Canada) for vegetative growth. Each plant was numbered, then placed adjacent to one another in grids, the order of which was determined using a random number generator. The random placement of plants was intentional to smooth out the impact of environmental variations within the room on the resultant data. Natural light was supplemented with artificial lighting using adjustable Gavita Pro 1000e DE HPS (Gavita, Vancouver, WA, USA). The photosynthetically active radiation (PAR) was measured using an Apogee MQ-200 Quantum Separate Sensor (Apogee Instrument, Logan, UT, USA) and was determined to be 200 ± 68 $\mu mol \cdot m^{-2} \cdot s^{-1}$ at the canopy level, on average. The greenhouse temperature was set constant at 22°C. The plants were watered with Alaska Fish Fertilizer (N-P-K Ratio 5:1:1) (Alaska, Canada) every 3-4 days when the soil was visually dry. After two months vegetative growth, plants were transferred to 9 wooden planters, each measuring 150cm × 150cm and filled with soil, with 12 plants per planter and a 12-hour photoperiod to induce flowering. Natural light was supplemented with adjustable Gavita Pro 1000e DE HPS (Gavita, Vancouver, WA, USA). The average PAR at canopy level, measured over three days, was 559 ± 71 $\mu mol \cdot m^{-2} \cdot s^{-1}$ in the morning and 1159 ± 198 $\mu mol \cdot m^{-2} \cdot s^{-1}$ at noon. The highest PAR readings ranged from 1016 ± 295 to 1390 ± 104 $\mu mol \cdot m^{-2} \cdot s^{-1}$ at the canopy level for the nine planters, and the relative standard deviation was 11.9%, indicating a relatively even light distribution. The greenhouse temperature was set constant at 28°C during this phase. Relative humidity ranged between 35% and 60%. The plants were watered every 3-4 days when the soil was visually dry. After two months of flowering, the whole plants were harvested and hung to dry in a closed environment. Horticultural fans were used to maintain air circulation, and the temperature was kept under 35°C. The plants were dried for 7 days until the leaves and stems became brittle. At this time, the plants' moisture content is usually below 10-15% (mg/mg) [226,227]. Dried material was stored at room temperature until analysis.

Table 5.1 Information and assignment of 21 cultivars into three chemotypes based on THC and CBD ratio

Cultivar number	Cultivar name	Chemotypes	Total THC* Ratio (mean ± SD)	Total CBD* Ratio (mean ± SD)	Clusters	"Sativa" or "Indica"
1	Lemon Garlic OG	1-Intermediate	29.0% ± 2.4%	53.5% ± 2.2%	C1	"Indica" dominant

2	Royal Medic	2-Intermediate	32.7% ± 1.4%	50.9% ± 1.0%	C3	"Sativa" dominant
3	Blue Hawaiian	3-CBD	3.4% ± 0.2%	77.7% ± 0.8%	C3	"Sativa" dominant
4	Kandy Kush	4-CBD	3.9% ± 0.4%	77.5% ± 1.0%	C3	"Sativa" dominant
5	Special	5-CBD	3.3% ± 0.3%	78.6% ± 0.9%	C3	Not provided
6	NN	6-CBD	3.3% ± 0.2%	77.2% ± 0.5%	C3	Not provided
8	Treat	8-CBD	3.1% ± 0.5%	78.4% ± 1.4%	C3	Not provided
9	High	9-Intermediate	25.2% ± 1.1%	55.3% ± 0.7%	C3	Not provided
10	CB7	10-CBD	3.3% ± 0.7%	79.8% ± 1.2%	C3	Not provided
11	33°	11-THC	79.9% ± 1.0%	0.4% ± 0.2%	C1	Not provided
12	Banana Cake	12-THC	81.7% ± 0.4%	0.4% ± 0.1%	C2	"Indica" dominant
13	Bananium	13-THC	81.7% ± 0.6%	0.3% ± 0.05%	C3	"Indica" dominant
14	Burmese Blueberry	14-THC	78.8% ± 1.1%	0.3% ± 0.02%	C2	"Indica" dominant
15	Divine Banana	15-THC	81.7% ± 1.1%	0.3% ± 0.05%	C2	"Indica" dominant
16	Granddaddy Purple	16-THC	74.1% ± 0.7%	0.4% ± 0.1%	C2	"Indica" dominant
18	Lemon Sorbet	18-THC	84.4% ± 0.6%	0.6% ± 0.3%	C1	"Indica" dominant
19	MeatHead	19-THC	82.0% ± 1.2%	0.2% ± 0.03%	C2	"Indica" dominant
20	Nanitro	20-THC	78.6% ± 0.4%	0.3% ± 0.1%	C1	"Indica" dominant
21	Platinum Jelly Punch	21-THC	80.0% ± 1.0%	0.4% ± 0.1%	C1	"Indica" dominant
22	SBSK2 (Lemon Thai)	22-THC	79.7% ± 0.3%	0.4% ± 0.1%	C3	50/50 hybrid
23	Super Sherbet	23-THC	79.6% ± 1.6%	0.2% ± 0.02%	C1	"Indica" dominant

*Total THC = THCA×0.877 +THC. Total CBD = CBDA × 0.877 + CBD.
*Total THC ratio = Total THC/sum of cannabinoids content

5.3.2 Morphological traits evaluated

Traits on live plants were measured using a ruler and a digital micrometer 40 days after rooting and again at the end of the flowering phase (**Table 5.2**). Because it was difficult to evaluate leaf traits with respect to nodal positions in cuttings, mean values of traits on leaves were averaged from node 3 to node 10 (or highest node number if fewer than 10 nodes). Leaf traits were measured as illustrated in **Figure 5.2**, referencing Anderson's method [27,32]. Petiole width and thickness were measured midway along the petiole. Leaf chlorophyll concentrations were measured using a MC-100 chlorophyll concentration meter (Apogee Instrument, Logan, UT, USA). After harvest and drying, the following traits were measured: inflorescence yield, mean inflorescence weight averaged from at least ten inflorescences, compactness of harvested inflorescences, whether the trichomes were dense and resinous, and incidence of *Botrytis cinerea* in inflorescences. All traits in **Table 5.2** were treated as continuously variable for statistical analysis.

Table 5.2 Phenotypic characteristics evaluated on each plant assigned to three chemotypes.

	Code	Characteristic	Unit/Notes
1	HgtVeg	Plant height 40 days after rooting	cm
2	DiaVeg	Stem diameter at base 40 days after rooting	mm

3	StmClrVeg	Reddish-brown coloration at base of stem of plants 40 days after rooting.	Visually rated: 1-absent, 2-somewhat apparent, 3-present
4	VisGrnVeg	Visual determination of greenness 40 days after rooting.	Visually rated: 1-light green, 2-green, 3-deep green
5	BranchVeg	Extent of branching 40 days after rooting.	Visually rated: 1- less branching, 2-branching, 3-heavily branching
6	StretchVeg	Extent of stretching 40 days after rooting.	Visually determined: 1-compact, 2-normal, 3-very stretching
7	NodeVeg	Number of nodes 40 days after rooting	
8	IntLngVeg	Mean internode length 40 days after rooting	mm
9	LftNumVeg	Mean leaflet number at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
10	CtrLftLngVeg	Mean length of central leaflet at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	mm
11	CtrLftWdtVeg	Mean width of central leaflet at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	mm
12	LftRatioVeg	Mean width/length ratio of central leaflet at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
13	LftShapeVeg	Mean ratio of distance from base of central leaflet to widest point/total length at node n, 40 days after rooting, n=3 to 10 (or highest number < 10)	
14	PetLngVeg	Mean petiole length at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	mm
15	PetWdtVeg	Mean petiole width at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	mm
16	PetRatioVeg	Mean petiole width/thickness ratio at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
17	PriSerVeg	Mean number of primary serrations on central leaflet at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
18	SecSerVeg	Mean number of secondary serrations on central leaflet at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
19	ChlphlVeg	Mean leaf chlorophyll concentration at node n 40 days after rooting, n=3 to 10 (or highest number < 10)	
20	HgtFlw	Final height at the end of flowering stage	cm
21	HgtRat	Ratio of height 40 days after rooting over height at the end of flowering stage	cm
22	DiaFlw	Stem diameter at base at the end of the flowering stage	mm
23	StmClrFlw	Reddish-brown coloration at base of stem of plants at the end of the flowering stage	Visually rated: 1=absent, 2=somewhat apparent, 3=present
24	YieldFlw	Flower yield per plant	g
25	WeightFlw	Mean weight per inflorescence averaged from at least ten inflorescences at the end of flowering stage	mg
26	OvrAprFlw	Overall appearance of inflorescences at the end of flowering stage	Visually rated: loose =1, intermediate =2, compact =3
27	SugrLftClrFlw	Color of leaves in the inflorescence at the end of flowering stage	Visually rated: green =1, mix of green and purple =2, purple = 3
28	CalyxClrFlw	Color of calyx	Visually rated: green =1, mix of green and purple = 2, purple =3
29	ResinFlw	Whether inflorescences, on average of 5 from one plant, are resinous: sparkly, dense, sticky trichomes	Visually rated: 1. non-resin production, 2, intermediate, 3. resin production
30	SickFlw	Sickness at the end of flowering stage	Visually rated: *Botrytis cinerea* present =1, absent =0

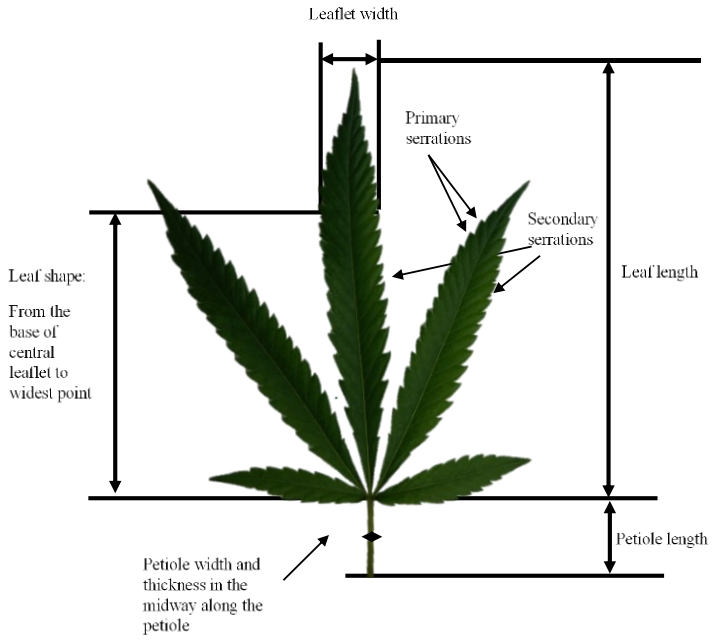


Figure 5.2 Leaf traits measured on the central leaflet of fully expanded leaves while still attached to the living plant.

5.3.3 Statistical analysis

In total, 85 plants representing 21 cultivars were included in the following analysis. First, correlations were calculated between individual morphological traits with THC and CBD (ratios relatively to total cannabinoids), respectively. Ratios were employed because the relative compound proportions are more stable than absolute values, the latter changing between growth stages, plant parts, and environmental factors [27]. Then, unsupervised (no preassigned categories as constraints) hierarchical clustering using Ward's minimum variance method [215] and principal component analysis (PCA) [228] were used to check within-cultivar variation and between-cluster variation. Finally, the data were subjected to supervised canonical correlation analysis with preassigned chemotypes in **Table 5.1**. Canonical correlation analysis is a multiple discriminant analysis that calculates the correlation between preassigned clusters and the set of covariates describing the observations (morphological traits in this study) [88]. Canonical variables are linear combination of the covariates that maximize the multiple correlation between the clusters and the covariates are uncorrelated with each other. The analysis outputs a biplot with the first two

canonical variables that provide maximum separation among the clusters. To identify phenotypic markers that contribute most to each chemotype, one-way ANOVA followed by Tukey honestly significant difference (HSD) post hoc test at the 0.05 significance level were used to determine whether significant differences exist between all clusters and each pair of clusters. Statistical analysis was performed with JMP 14.0.0 (SAS, Cary, NC, USA).

5.4 Results

5.4.1 Correlation analysis of morphological traits with THC and CBD

Correlations of THC and CBD with morphological traits are plotted in **Figure 5.3**. The ratio of THC is positively correlated with leaf width/length ratio (0.77), width of central leaflet (0.55), final height at the end of flowering stage (0.54), inflorescence yield per plant (0.46), leaf chlorophyll concentration (0.42), sickness (*Botrytis cinerea* incidence) (0.41), stem diameter at base at the end of the flowering stage (0.33), and weight per inflorescence (0.32) in decreasing correlations. The ratio of CBD is positively correlated with number of primary serrations on central leaflet (0.52), number of leaflets (0.44), and length of central leaflet averaged from each node (0.35). The traits that were positively correlated with THC were all negatively correlated with CBD, where the absolute values of the correlations were similar, and vice versa. Interestingly, the correlations of the color of the calyx in inflorescences (green or purple) with CBD and THC are -0.12 and 0.09, respectively. The correlations of the color of the leaves in inflorescences (green or purple) with CBD and THC are -0.13 and 0.12, respectively. Neither was highly correlated with CBD and THC production.

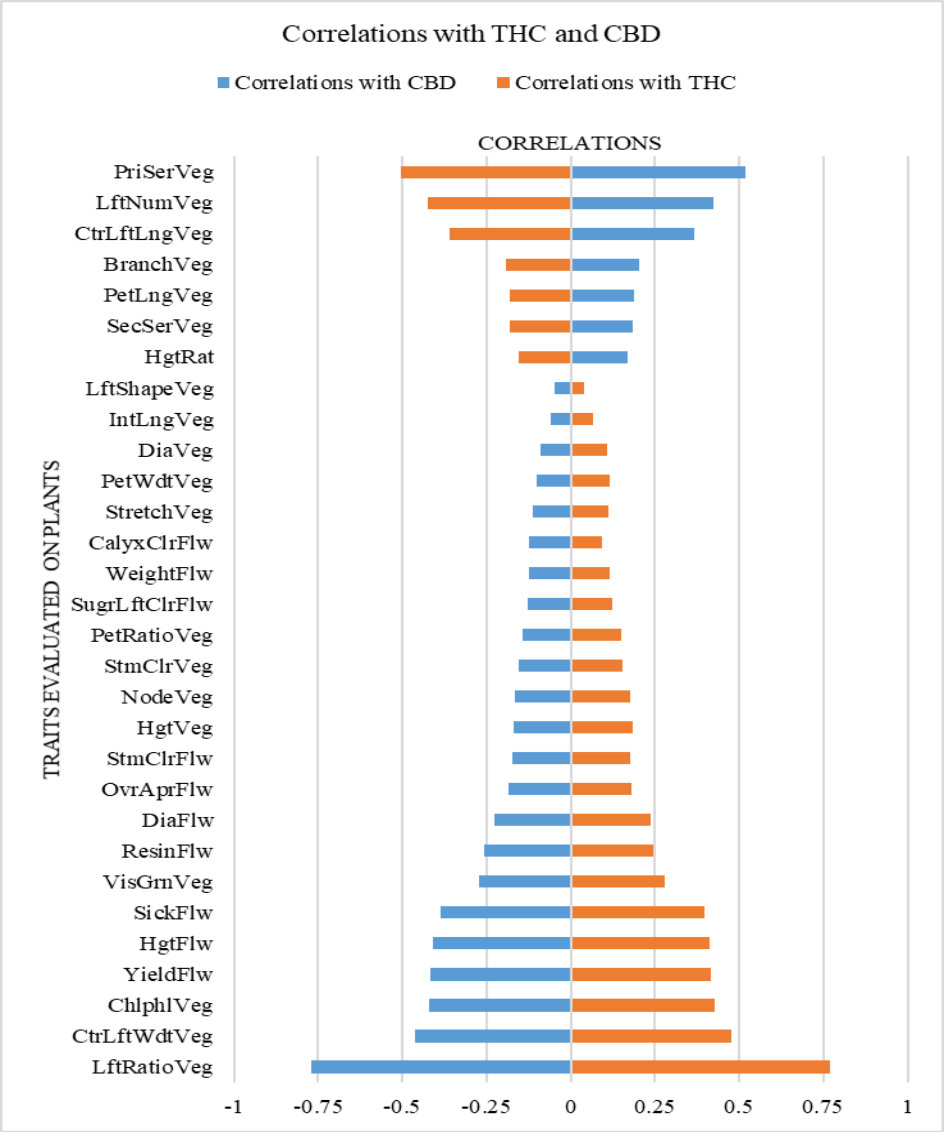


Figure 5.3 Correlations of morphological traits with THC and CBD

5.4.2 Unsupervised hierarchical clustering

A hierarchical clustering dendrogram of the 85 plants is shown in **Figure 5.4**, where most plants of the same cultivars were clustered together. Plants in 2-intermediate cultivar were scattered over the dendrogram. They seemed to experience different growth conditions, possibly related to uneven light interception. The dendrogram shows two major branches: one branch with plants from CBD dominant cluster C1 and the other with plants from THC dominant cluster C3. Plants

from intermediate cluster C2 were distributed over both branches, showing phenotypic similarities with both THC dominant cultivars and CBD dominant cultivars, possibly due to hybridization.

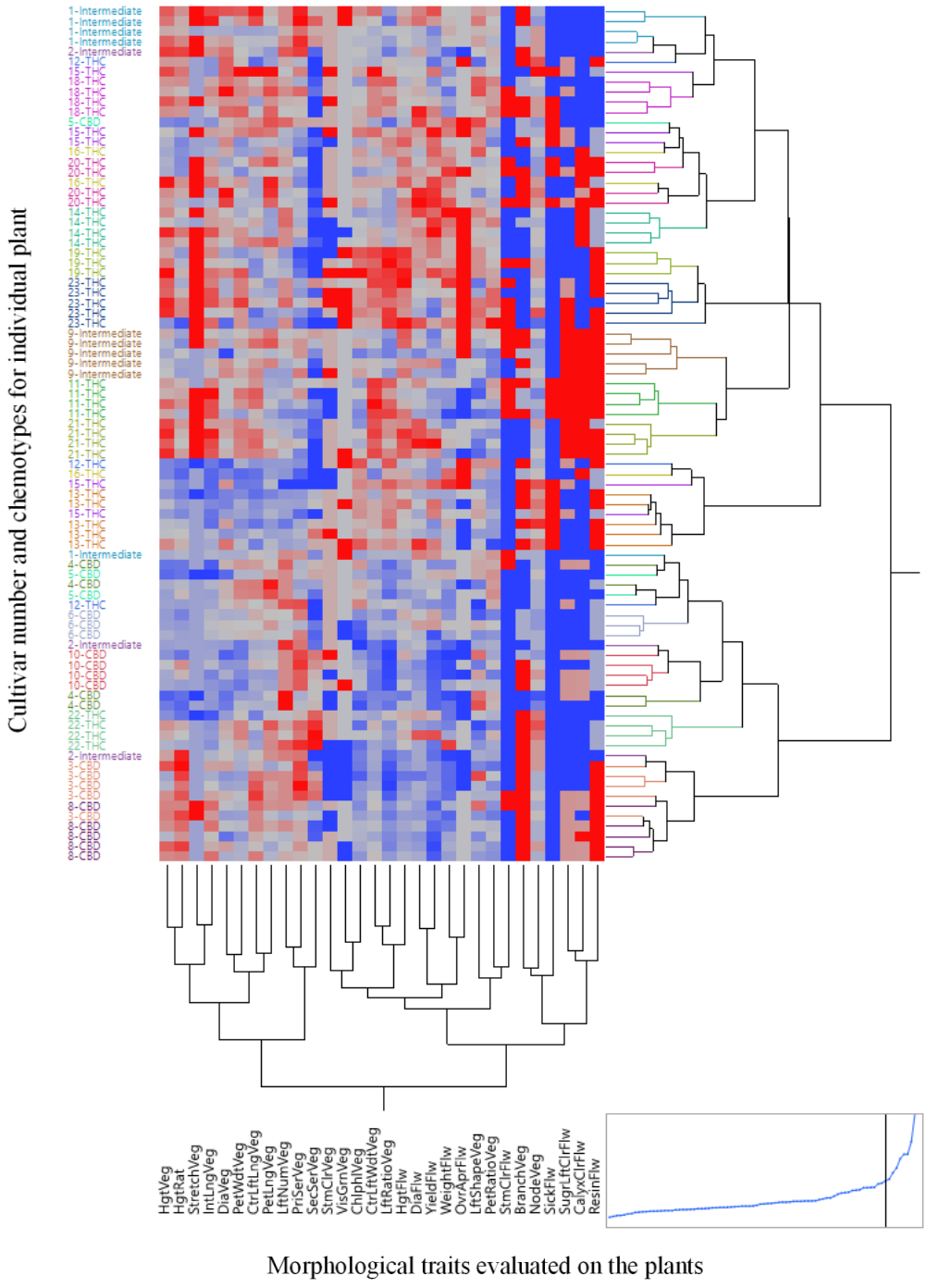


Figure 5.4 Dendrogram by hierarchical clustering analysis using 30 morphological traits of 85 plants representing 21 cultivars.

5.4.3 Unsupervised principal component analysis

Figure 5.5 shows a scatterplot of 85 plants on PC1 and PC2. Plants of the same cultivars tended to occupy the same region on the plot, which shows small within-cultivar variation and relative consistent morphological profiles within each cultivar. PC1 and PC2 explained 21.7% and 13.6% of total variance, respectively. These numbers are comparable to those of Hillig's study, where the numbers were 29.0% and 17.3%, respectively [27]. Plants from THC dominant cluster C3 mainly occupied the right side the plot, while plants from CBD dominant cluster C1 occupied the left. Plants from intermediate cluster C2 occupied the middle of the plot and were mixed with both THC dominant plants and CBD dominant plants. Although plants from the same chemotypes tend to cluster together, the three clusters overlap, which may explain why PC1 and PC2 only explained 35.3% of the total variance. Cultivars assigned to C3 expressed a greater range of phenotypic variation than those assigned to C1 and C2 in the PC scatter plot, which may be due to long history of selection for high THC levels for recreational purposes [69].

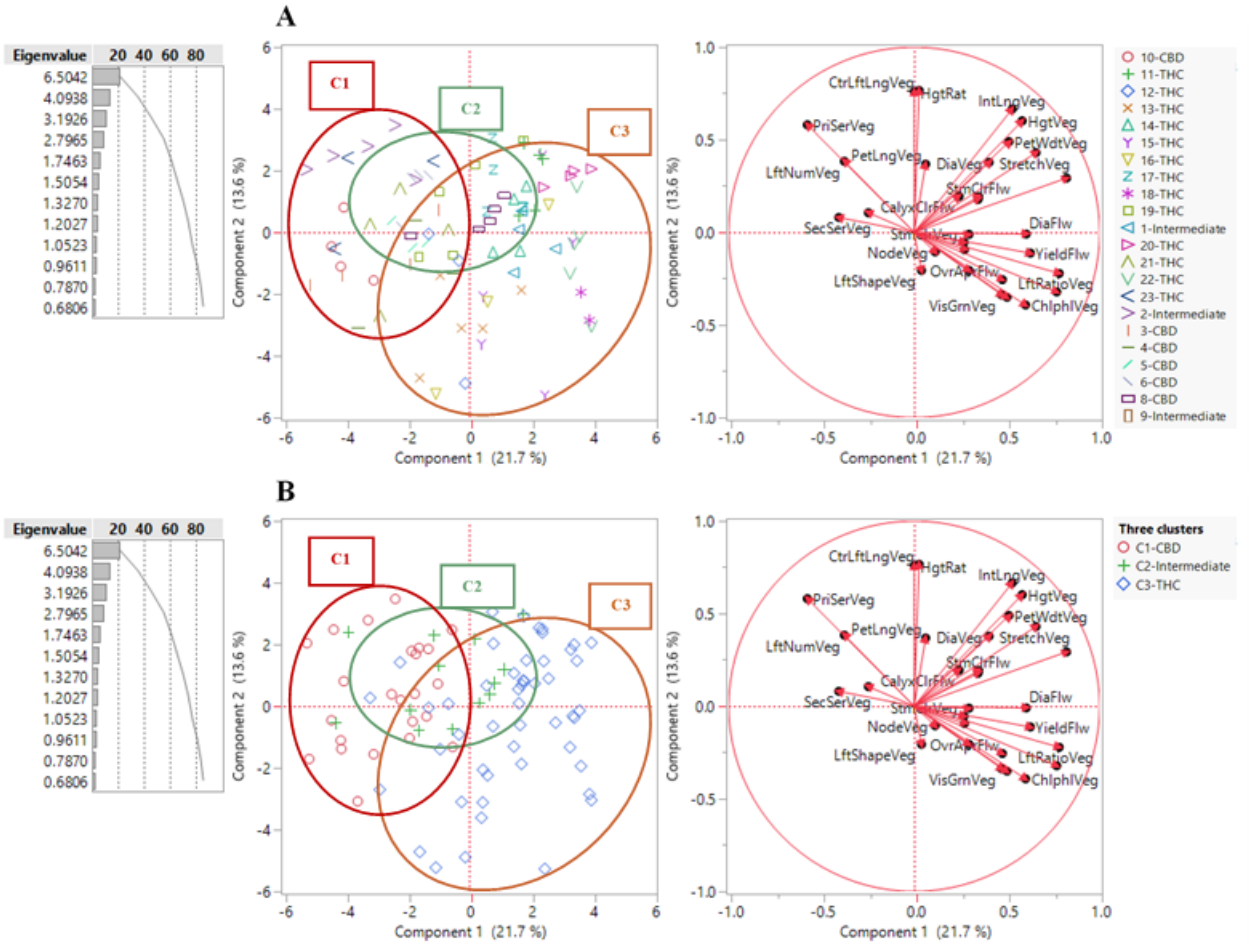


Figure 5.5 PCA scatter plot (left) and loading plot (right) using 30 morphological traits of 85 plants representing 21 cultivars.
(A) using individual cultivar as label and (B) using three chemotypes as label.

Table 5.3 Formatted loading matrix for PC1 and PC2

PC1				PC2			
Traits	**Positive loadings**	**Traits**	**Negative loadings**	**Traits**	**Positive loadings**	**Traits**	**Negative loadings**
CtrLftWdtVeg	0.82	PriSerVeg	-0.58	CtrLftLngVeg	0.77		
HgtFlw	0.78	SecSerVeg	-0.41	HgtRat	0.76		
LftRatioVeg	0.77			IntLngVeg	0.67		
StretchVeg	0.66			HgtVeg	0.60		
YieldFlw	0.63			PriSerVeg	0.58		
DiaFlw	0.60			PetWdtVeg	0.49		
ChlphlVeg	0.60			StretchVeg	0.43		
HgtVeg	0.58						
IntLngVeg	0.53						
PetWdtVeg	0.51						
WeightFlw	0.50						
OvrAprFlw	0.48						
VisGrnVeg	0.47						
DiaVeg	0.40						
BranchVeg	0.40						

* Only compounds with absolute loadings > 0.4 are listed

The loading matrix in **Table 5.3** lists the traits that contributed most to the separations along PC1 and PC2. Loadings with absolute values equal to or greater than 0.4 are listed in the table. PC1 was positively correlated with the width of central leaflet, final height at the end of flowering stage, leaflet width/length ratio, extent of stretching, inflorescences yield per plant, etc., which were traits identified as positively correlated with THC. PC1 was negatively correlated with the number of primary and secondary serrations on central leaflet, which were traits identified as positively correlated with CBD. Traits that were positively correlated with PC2 and nearly vertical with PC1 included length of central leaflet and the height ratio between 40 days after rooting and at the end of flowering stage. These traits overlap with plants from intermediate C2 in the scatter plot. Traits positively correlated with PC2 and positively correlated with PC1 were internode length, plant height 40 days after rooting, and petiole width. These traits are responsible for the location of the THC dominant plants on the upper right quadrant. The number of primary serrations on central leaflet was positively correlated with PC2 and negatively with PC1, which was responsible for CBD dominant plants located on the upper left quadrant of the plot.

5.4.4 Supervised canonical correlation analysis

The canonical correlation analysis showed good separation between the preassigned chemotypes (**Figure 5.6**). Each plant was predicted to be in its originally preassigned clusters C1, C2, and C3 with a 92.9% (79/85) accuracy (**Table 5.4**). Means, standard deviations (±SD), ranges, Tukey HSD multiple tests at the 0.05 significance level, and p value of one-way ANOVA of 30 traits for each of the three clusters were calculated in **Table 5.5**. The largest number of significant differences was 19, which was between C1 and C3. The most similar pair was C1 and C2, with three significant differences. The number of significant differences between C2 and C3 was six. CBD dominant cultivars had more leaflets, longer central leaflets, and more primary and secondary serrations, which are traits positively correlated with CBD production. THC dominant cultivars had higher plant height, larger stem diameter, deeper green colour of leaves, more nodes, wider central leaflet, larger width/length ratio of central leaflet, wider petiole width, and higher chlorophyll concentrations, higher inflorescences yield, larger inflorescence weight, more compact looking inflorescence, and sticky inflorescences, and higher *Botrytis cinerea* incidence. These traits were also positively correlated with THC production. Most traits for intermediate cultivars were at an intermediate level between C1 and C3 or at the same level with cultivars in C1 or C3. Several traits were not significantly different between three groups, including reddish-brown coloration at the base of the stem, extent of branching, internode length, leaf shape, petiole length, petiole width/thickness ratio, plant growth rate, and color of leaves and color of calyx. Example leaflets for C1, C2, and C3, green/purple leaves and calyx, inflorescences infected with *Botrytis cinerea* are shown in **Figure 5.7**.

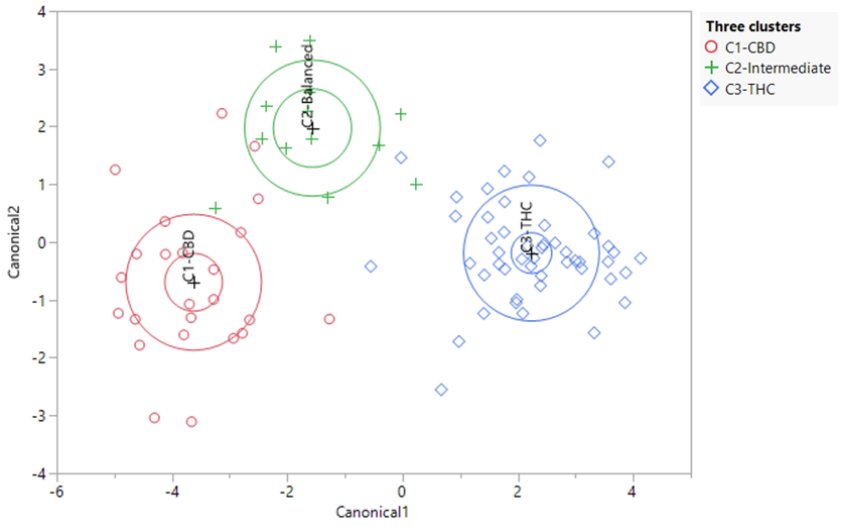


Figure 5.6 Canonical correlation analysis scatterplot of 85 plants representing 21 cultivars on the first and second canonical axes using 30 traits.

The cultivars were preassigned to three chemotypes in Table 1. The observations and the multivariate means of each group ("+") are represented as points on the biplot. A 95% confidence level ellipse is plotted for each mean. An ellipse denoting a 50% contour is plotted for each group, that contains approximately 50% of the observations.

Table 5.4 Canonical correlation analysis summary of preassigned and predicted classifications of 85 plants into three chemotypes using 30 traits

Preassigned	Predicted		
	C1-CBD	C2-Intermediate	C3-THC
C1-CBD	21	3	0
C2-Intermediate	1	12	0
C3-THC	0	2	46

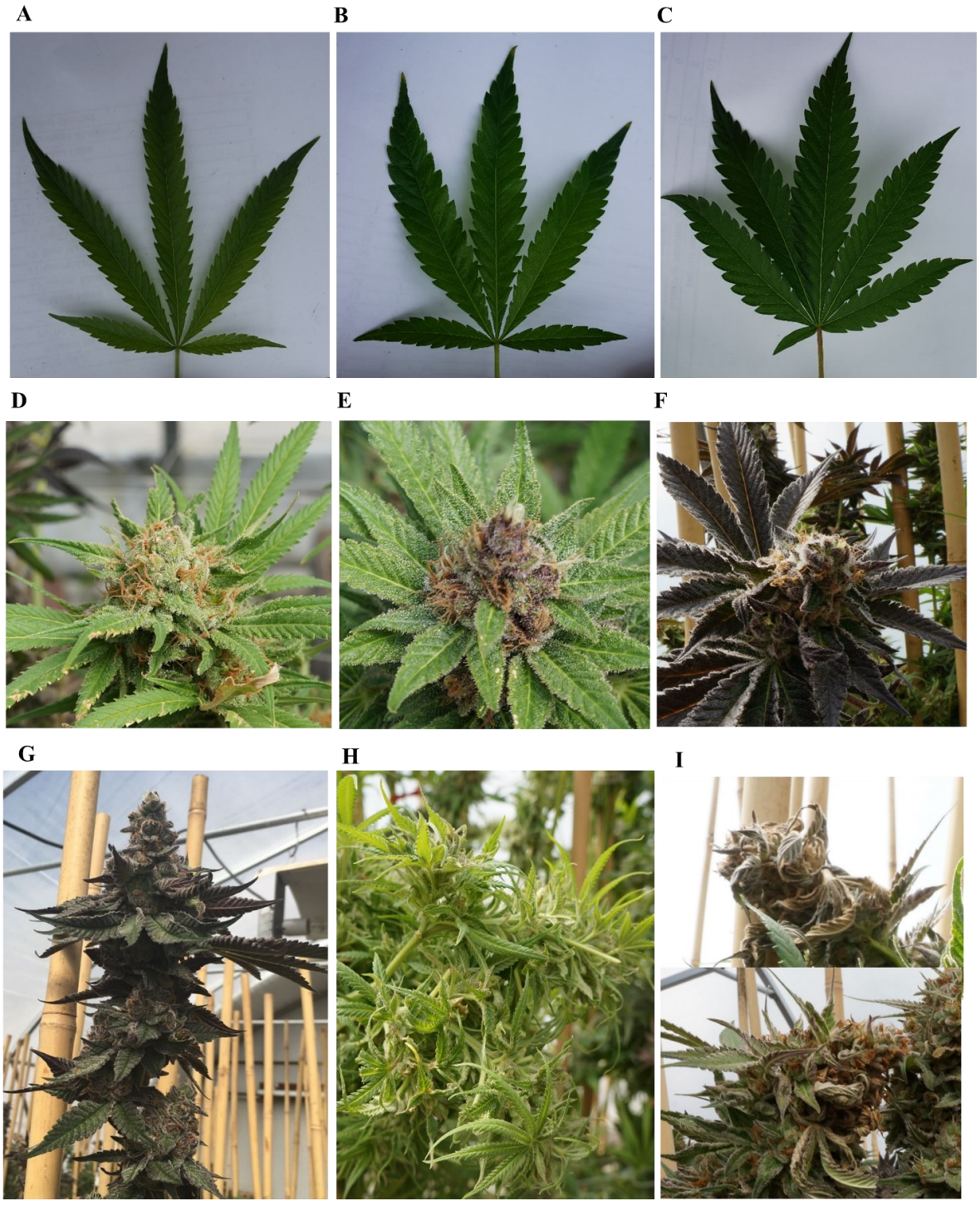


Figure 5.7 Example of inflorescences of three chemotypes
(A) Example leaflets for CBD dominant cultivars, (B) intermediate cultivars, and (C) THC dominant cultivars. (D) Example mature inflorescences of cultivars with green leaves and green calyx, (E) green leaves with purple calyx, and (F) purple leaves with purple calyx. (G) Example of compact inflorescences, (H) loose inflorescences, and (I) inflorescences infected with *Botrytis cinerea*.

Table 5.5 . Means (±SD) and ranges (mean-SD, mean+SD) of 30 traits for 85 plants assigned to C1-CBD dominant, C2-intermediate, and C3-THC dominant.

	Three clusters	C1-CBD	C2-Intermediate	C3-THC	ANOVA
		N = 24	N = 13	N = 48	p
Vegetative growth (40 days after rooted)	HgtVeg (cm)	19.64 ± 5.57 b 14.07 – 25.21	23.47 ± 3.76 ab 19.71 – 27.23	24.05 ± 4.79 a 19.26 – 28.84	0.0012
	DiaVeg (mm)	3.26 ± 0.46 b 2.80 – 3.72	3.57 ± 0.63 ab 2.94 – 4.20	3.68 ± 0.61 a 3.07 – 4.29	0.013
	StmClrVeg	1.67 ± 0.56 a 1.11 – 2.23	2.00 ± 0.41 a 1.59 – 2.41	1.92 ± 0.64 a 1.28 – 2.56	0.1847
	VisGrnVeg	1.71 ± 0.55 b 1.16 – 2.26	2.15 ± 0.55 a 1.60 – 2.70	2.17 ± 0.47 a 1.70 – 2.64	0.0016
	BranchVeg	2.29 ± 0.85 a 1.44 – 3.14	2.46 ± 0.66 a 1.80 – 3.12	2.06 ± 0.91 a 1.15 – 2.97	0.35
	StretchVeg	2.00 ± 0.42 b 1.58 – 2.42	2.46 ± 0.52 a 1.92 – 2.98	2.44 ± 0.61 a 1.83 – 3.05	0.0036
	NodeVeg	8.04 ± 1.33 b 6.71 – 9.37	8.62 ± 1.12 ab 7.50 – 9.74	9.33 ± 1.77 a 7.56 – 11.1	0.008
	IntLngVeg (mm)	24.32 ± 5.10 a 19.22 – 29.42	27.46 ± 4.67 a 22.79 – 32.13	26.52 ± 6.58 a 19.94 – 33.10	0.1715
	LftNumVeg	4.92 ± 0.47 a 4.45 – 5.39	4.81 ± 0.52 ab 4.29 – 5.33	4.34 ± 0.74 b 3.60 – 5.08	0.0009
	CtrLftLngVeg (mm)	95.64 ± 10.93 a 84.71 – 106.57	95.77 ± 13.69 ab 82.08 – 109.45	84.74 ± 16.24 b 68.51 – 100.98	0.0071
	CtrLftWdtVeg (mm)	17.53 ± 2.39 b 15.14 – 19.92	19.31 ± 2.66 b 16.65 – 21.97	22.65 ± 3.51 a 19.14 – 26.16	<0.0001
	LftRatioVeg	0.18 ± 0.02 b 0.16 – 0.20	0.20 ± 0.02 b 0.18 – 0.22	0.25 ± 0.03 a 0.22 – 0.28	<0.0001
	LftShapeVeg	0.51 ± 0.04 a 0.47 – 0.55	0.51 ± 0.02 a 0.49 – 0.53	0.50 ± 0.04 a 0.46 – 0.54	0.9282
	PetLngVeg (mm)	40.20 ± 6.80 a 33.40 – 46.99	38.69 ± 6.82 a 31.87 – 45.51	38.93 ± 8.47 a 30.46 – 47.41	0.7756
	PetWdtVeg (mm)	1.34 ± 0.15 b 1.19 – 1.49	1.41 ± 0.16 ab 1.25 – 1.57	1.47 ± 0.21 a 1.26 – 1.68	0.0232
	PetRatioVeg	0.99 ± 0.03 a 0.96 – 1.02	1.00 ± 0.03 a 0.97 – 1.03	1.01 ± 0.04 a 0.97 – 1.05	0.2934
	PriSerVeg	15.04 ± 1.74 a 13.30 – 16.78	15.97 ± 1.49 a 14.48 – 17.46	12.75 ± 2.06 b 10.69 – 14.81	<0.0001
	SecSerVeg	1.14 ± 0.87 a 0.27 – 2.01	0.99 ± 1.04 a 0 – 2.03	0.34 ± 0.12 b 0.22 – 0.46	0.0002
	ChlphlVeg	16.92 ± 2.49 b 14.43 – 19.41	18.33 ± 3.00 ab 15.33 – 21.33	20.39 ± 3.38 a 17.01 – 23.77	<0.0001
At the end of flowering stage	HgtFlw (cm)	97.21 ± 15.00 b 82.21 – 112.21	115.54 ± 17.67 a 97.87 – 133.21	126.56 ± 13.42 a 113.14 – 139.98	<0.0001
	HgtRat	0.21 ± 0.06 a 0.15 – 0.27	0.21 ± 0.05 a 0.16 – 0.26	0.19 ± 0.04 a 0.15 – 0.23	0.5352
	DiaFlw (mm)	13.65 ± 2.41 b 11.24 – 16.06	15.58 ± 2.61 ab 12.97 – 18.19	16.85 ± 2.95 a 13.90 – 19.80	<0.0001
	StmClrFlw	1.13 ± 0.34 a 0.79 – 1.47	1.31 ± 0.48 a 0.83 – 1.79	1.31 ± 0.48 a 0.83 – 1.79	0.1509
	YieldFlw (g)	36.43 ± 26.73 b 9.70 – 63.16	60.40 ± 30.50 ab 29.90 – 90.90	85.09 ± 38.45 a 46.64 – 123.54	<0.0001
Inflorescences (Harvested)	WeightFlw (mg)	355.04 ± 130.40 b 224.64 – 485.44	411.69 ± 118.04 ab 293.65 – 529.73	542.75 ± 208.40 a 334.35 – 751.15	0.0002
	OvrAprFlw	1.71 ± 0.55 b 1.16 – 2.26	2.00 ± 0.71 ab 1.29 – 2.71	2.19 ± 0.74 a 1.45 – 2.93	0.0345
	SugrLftClrFlw	1.50 ± 0.51 a 0.99 – 2.01	1.77 ± 1.01 a 0.76 – 2.78	1.73 ± 0.83 a 0.90 – 2.56	0.4277
	CalyxClrFlw	1.50 ± 0.59 a 0.91 – 2.09	1.77 ± 1.01 a 0.76 – 2.78	1.77 ± 0.97 a 0.90 – 2.56	0.4341
	ResinFlw	2.17 ± 0.82 ab 1.35 – 2.99	1.77 ± 1.01 b 0.76 – 2.78	2.40 ± 0.72 a 1.68 – 3.12	0.0489
	SickFlw	0.04 ± 0.20 b 0 – 0.24	0.00 ± 0.00 b 0.00 – 0.00	0.40 ± 0.49 a 0 – 0.89	0.0002

* Means for the same trait not connected by same letter are significantly different.

5.5 Discussion

In this study, 21 *Cannabis* cultivars belonging to three chemotypes (THC dominant, intermediate, and CBD dominant) were grown in a greenhouse. Morphological traits were measured, and canonical correlation analysis was used to test the goodness of fit between chemotype labelling and phenotypic variations. This study also identified phenotypic markers for each chemotype. The widespread crossbreeding and introgression in *Cannabis* blurred the differences between NLD and WLD cultivars, as well as made the distinction between their hybrids difficult [69]. However, useful suites of traits were identified for differentiating three chemotypes. These identified traits were largely consistent as confirmed by correlation analysis, PCA, and canonical correlation analysis. The ratio-based qualitative differences may be more consistent between growing environments and therefore more useful for differentiation when applied to conditions different from those used in this study. THC dominant cultivars had the largest mean width/length ratio of 0.25 ± 0.03. Hillig described two drug types (THC dominant) before large scale hybridization in the 90s: NLD with width/length ratio ranged from 0.15 ± 0.02 to 0.24 ± 0.03 (measured with respect to nodal positions), and WLD with width/length ratio ranged from 0.22 ± 0.03 to 0.39 ± 0.06 [27]. The THC dominant cultivars described in this study had width/length ratios at the high end of the ratio for NLD cultivars and the lower end of the ratio for WLD cultivars. This may be attributed to hybridization between NLD and WLD cultivars, which was performed to obtain sinsemilla hybrids with high THC content, low CBD content, and high inflorescences yield [216]. This study is the first to describe leaf traits for CBD dominant and intermediate cultivars. CBD dominant cultivars had narrow leaflets with width/length ratios of 0.18 ± 0.02, whereas intermediate cultivars had intermediate width/length ratios of 0.20 ± 0.02. A complete genome assembly of CBD dominant cultivars revealed that these cultivars were created by integrating hemp-type CBD acid synthase gene clusters into a background of drug-type *Cannabis* to elevate CBD production [187]. The intermediate width/length ratio for intermediate cultivars was likely a result of hybridization between purebred CBD dominant cultivars and THC dominant cultivars.

In addition to qualitative criteria, this study provided additional quantitative and visual phenotypic criteria that may have differentiation power. For example, CBD dominant cultivars were lighter green, THC dominant cultivars were deeper green, and intermediate cultivars had shades in between. Since morphological traits change depending on environment variables, absolute measurements, including the plant height, the leaflet length, and the inflorescence yield, may differ

if grown in a different environment. The visual greenness of leaf colors and chlorophyll concentrations were positively correlated with each other. The inflorescences of NLD/WLD hybrid cultivars in C3 were large and compact – a clear result of artificial selection. However, the major horticultural drawback of these hybrids is their susceptibility to fungal infections [216]. NLD landraces originated from regions with relatively humid conditions (Colombia, India, Jamaica, Thailand, etc.) and evolved natural resistance to fungal infection, whereas WLD landraces adapted to arid the Afghani environment in which fungal resistance was unnecessary [216,266]. More varieties in C3 had *Botrytis cinerea* infection than in C1 and C2, indicating that these cultivars face additional risk in a greenhouse environment. Alternatively, this may reflect the current market in North America – growers and users favour large and compact inflorescences, which are a trait of wide leaflet cultivars.

5.6 Conclusions

This study investigated the phenotypic variation in 21 cannabis plants with preassigned chemotype labels using thirty morphological traits measured on live plants and harvested inflorescences. The data were subjected to correlation analysis, unsupervised clustering analysis (hierarchical clustering and PCA) and supervised canonical correlations analysis. In unsupervised clustering, 85 plants were clustered in accordance with their chemotypes. Canonical correlation analysis classified 85 plants into three chemotypes with 92.9% accuracy. Numerous significant differences identified among chemotypes were largely consistent with results from correlation analysis, hierarchical clustering, PCA, and by comparing group means between chemotypes. The identified suites of phenotypic signatures in this work can be used to determine chemotypes on live plants prior to or as a supplement to chemical and genetic analysis.

Chapter 6 – Conclusions

The vernacular naming convention commonly used by the cannabis community ("Sativa" and "Indica") is inadequate for medical purposes because they are arbitrarily, unreliable, and inconsistent. In addition, they only include high THC strains. CBD dominant strains and balanced strains, which have been getting increased attention due to CBD's use as a therapeutic, have not been systematically studied nor compared to THC strains. The aim of the project was to develop an integrated classification system based on the genetic distance, chemical differences, and quantified morphological characteristics for identifying or selecting strains for clinical research and medicinal production.

This study used a common garden experiment where all plants were grown and uniformly processed in a single location under identical environmental conditions. This common garden experiment was the major advantage of this research because it enabled the study of the genetics, biochemical profiles, and morphological traits in uniformly processed plants. This study first assigned 23 cannabis strains into five clusters based on genetic distance at genome level: one CBD dominant, one balanced, and three THC dominant clusters. Strain differentiation into their assigned chemotype labels was investigated using chemotyping and phenotyping, using the full spectrum of secondary metabolites and morphological traits, respectively. The results have confirmed with the hypothesis that chemotypic and phenotypic treatment aligned with the pattern of genetic variation with high accuracy. The explanation was that genetics differences contribute most to these variations when environmental factors are controlled.

6.1 Genotyping cannabis strains

For genotyping, current cannabis classification studies use genome fragments from different sources, with few or no overlapping markers between datasets. Leveraging the recent release of the 10-chromosome cannabis genome map, this study sequenced the whole genome of 23 cannabis strains and identified 137,858 genome-wide SNPs that provided insight into the distribution of genetic diversity and population structure in three major chemotypes of modern North American cannabis. Balanced (intermediate) strains share a closer gene pool with CBD dominant strains, while THC gene pool is more dispersed. DAPC assigned 23 strains into five chemotype-aligned groups: one CBD dominant, one balanced, and three THC dominant clusters. The study further

identified 344 multiallelic SNPs, mostly triallelic SNPs, by DAPC, which were responsible of separating CBD dominant, balanced, and THC dominant strains. This is the first such report for cannabis in the literature because triallelic SNPs are excluded so far in cannabis population structural analysis. These SNPs were further applied to PCA, NJ tree, and hierarchical clustering, which provided consistent observations and groupings despite the differences in algorithms. The identified SNPs were spread across all 10 chromosomes, with the largest portion of genetic variation occurring (37%) on chromosome 6, where CBDAS and THCAS are located. The remaining variation between these groups may be attributable to the production of other cannabinoids, mono- and sesquiterpenoids, flavonoids and other compounds, or morphological characteristics. The genotyping results indicate that modern hybridized strains can still be separated using genome-wide information. With enough cannabis strains, DAPC has the potential to untangle the currently disordered genetic background of hybridized modern cannabis strains. This can be achieved by identifying the number of genetic clusters (especially within hundreds of and thousands of THC dominant strains), describing clusters by interpreting group memberships, and identifying SNPs that contribute the most to differentiation. These SNPs have the potential to be used as genetic markers or fingerprints for strain classification and identification.

6.2 Chemotyping cannabis strains

Chemotyping was carried out on the same set of plants using hierarchical clustering, PCA, and canonical correlation analysis. Full spectrum of secondary metabolites, including 14 cannabinoids, 45 terpenoids, 7 flavonoids, 3 sterols, and 3 triterpenoids were measured in inflorescences, leaves, stem bark, and roots at harvest. Canonical correlation analysis assigned individual plants into their preassigned chemotypes using both ratios (% within same metabolic categories) and absolute values (mg/mg%) with 100% accuracy. Significant chemical differences were identified for three chemotypes. Cannabinoids, terpenoids, flavonoids had differentiation power while sterols and triterpenoids had none. Chemotype markers for CBD dominant chemotype included total CBD, CBDV, CBC, α-pinene, β-myrcene, (-)-guaiol, β-eudesmol, α-eudesmol, α-bisabolol, orientin, vitexin, and isovitexin. Chemotype markers for THC dominant chemotype included total THC, total THCV, total CBG, camphene, limonene, ocimene, sabinene hydrate, terpinolene, linalool, fenchol, α-terpineol, β-caryophyllene, trans-β-farnesene, α-humulene, trans-nerolidol, quercetin, and kaempferol. Intermediate strains tended to be chemically closer to CBD strains. The content

of all the compounds in intermediate strains were usually equal to or at an intermediate level between CBD dominant strains and THC dominant strains. The results of chemotyping provide a comprehensive profile of bioactive compounds and form a baseline of reference values for understanding the "entourage effect" of cannabis extract. This study also identified chemotype markers as fingerprints that will facilitate strain identification and selection for research and clinical studies.

6.3 Phenotyping cannabis strains

Phenotyping was carried out on the same set of plants using hierarchical clustering, PCA, and canonical correlation analysis. A total of 30 morphological traits, including those on inflorescences, leaves, stems, and overall features, were measured on each plant: 40 days after rooting and at the end of the two month flowering phase. Canonical correlation analysis assigned individual plants into their preassigned chemotypes with 92.9% accuracy. Significant morphological differences were identified for three chemotypes. Phenotypic markers for CBD dominant chemotype included light-green and narrow leaflets, a greater number of primary and secondary serrations, loose inflorescences, dense and resinous trichomes, and *Botrytis cinerea* resistance. Phenotypic markers for intermediate strains included deep green and medium-wide leaflets, more primary and secondary serrations, medium compact inflorescences, trichomes that are less dense and less resinous, and *Botrytis cinerea* resistance. Phenotypic markers for THC dominant strains included deep-green and wide leaflets, large and compact inflorescences, dense and resinous trichomes, and *Botrytis cinerea* susceptibility. The results of phenotyping provide a comprehensive profile of morphological traits of modern cannabis strains, and is the first of its kind for CBD dominant and intermediate strains. These identified phenotype markers may facilitate preliminary strain identification and selection on live plants without using chemical analysis.

The goal of genotyping, chemotyping, and phenotyping was to obtain a classification rules for differentiating an individual plant into chemotypes. The differentiating traits or markers can be used individually or in tandem for strain identification in germplasm, live plants, or marketed products. For example, a plant's chemical profile may be predicted based on leaf color, shapes, compactness of inflorescences, and vice versa.

6.4 Future directions

The shortcoming of the project design is that the total number of strains are limited by strains available in the grower's facility. Characterizing cannabis strains should be a collaborative goal of the cannabis industry. I would consider this project a preliminary attempt to apply this triple-fingerprinting model to modern cannabis.

References

1. Zuardi, A. W. History of cannabis as a medicine: a review. *Revista Brasileira de Psiquiatría* **28**, 153–157 (2006).
2. Li, H.-L. An archaeological and historical account of cannabis in China. *Economic Botany* **28**, 437–448 (1973).
3. Li, H.-L. Hallucinogenic Plants in Chinese Herbals. *Journal of Psychedelic Drugs* **10**, 17–26 (1978).
4. Paris, M. & Nahas, G. Botany: the unstabilized species. in *Marihuana in science and medicine.* 3–36 (Raven Press, 1984).
5. Mikuriya, T. H. Marijuana in medicine: past, present and future. *Calif Med* **110**, 34–40 (1969).
6. Fankhauser, M. History of cannabis in Western Medicine. in *Cannabis and Cannabinoids* Chapter 4. p. 37-51. (The Haworth Integrative Healing Press, 2002).
7. Pinho, A. R. D. *Social and Medical Aspects of the Use of Cannabis in Brazil*. *Cannabis and Culture* 293–302 (De Gruyter Mouton, 1975).
8. Aldrich, M. History of therapeutic cannabis. in *Cannabis in medical practice* 35–55 (McFarland, 1997).
9. Bache, G. B. W. a. F. The Pharmacopoeia of the United States of America. in *Mack Printing* (1851).
10. Kandel, D. B. Marijuana users in young adulthood. *Arch Gen Psychiatry* **41**, 200–209 (1984).
11. Gaoni, Y. & Mechoulam, R. Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish. *J. Am. Chem. Soc.* **86**, 1646–1647 (1964).

12. Martin, B. R., Mechoulam, R. & Razdan, R. K. Discovery and characterization of endogenous cannabinoids. *Life Sci* **65**, 573–595 (1999).

13. Legality of cannabis. *Wikipedia* (2021).

14. State Medical Marijuana Laws. https://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx.

15. Branch, L. S. Consolidated federal laws of canada, Marihuana Medical Access Regulations. https://laws-lois.justice.gc.ca/eng/regulations/sor-2001-227/page-1.html (2014).

16. Branch, L. S. Consolidated federal laws of canada, Marihuana for Medical Purposes Regulations. https://laws-lois.justice.gc.ca/eng/regulations/sor-2013-119/page-1.html (2016).

17. Branch, L. S. Consolidated federal laws of canada, Cannabis Act. https://laws-lois.justice.gc.ca/eng/acts/c-24.5/ (2019).

18. Health Canada. Licensed cultivators, processors and sellers of cannabis under the Cannabis Act. https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/industry-licensees-applicants/licensed-cultivators-processors-sellers.html (2018).

19. Canada Gazette – Marihuana for Medical Purposes Regulations. http://gazette.gc.ca/rp-pr/p2/2013/2013-06-19/html/sor-dors119-eng.html (2013).

20. Badowski, M. E. & Yanful, P. K. Dronabinol oral solution in the management of anorexia and weight loss in AIDS and cancer. *Ther Clin Risk Manag* **14**, 643–651 (2018).

21. Ware1, M. A., Daeninck, P. & Maida, V. A review of nabilone in the treatment of chemotherapy-induced nausea and vomiting. *Ther Clin Risk Manag* **4**, 99–107 (2008).

22. Lynch, M. E., Cesar-Rittenberg, P. & Hohmann, A. G. A Double-Blind, Placebo-Controlled, Crossover Pilot Trial With Extension Using an Oral Mucosal Cannabinoid Extract

for Treatment of Chemotherapy-Induced Neuropathic Pain. *Journal of pain and symptom management* **47**, 166–173 (2014).

23. Johnson, J. R., Lossignol, D., Burnell-Nugent, M. & Fallon, M. T. An Open-Label Extension Study to Investigate the Long-Term Safety and Tolerability of THC/CBD Oromucosal Spray and Oromucosal THC Spray in Patients With Terminal Cancer-Related Pain Refractory to Strong Opioid Analgesics. *Journal of Pain and Symptom Management* **46**, 207–218 (2013).

24. Lakhan, S. E. & Rowland, M. Whole plant cannabis extracts in the treatment of spasticity in multiple sclerosis: a systematic review. *BMC Neurol* **9**, 59 (2009).

25. cannabis. *OED Online* (2009).

26. Turner, C. E., Elsohly, M. A. & Boeren, E. G. Constituents of *Cannabis sativa* L. XVII. A review of the natural constituents. *Journal of Natural Products* **43**, 169–234 (1980).

27. Hillig, K. W. A systematic investigation of Cannabis, PhD thesis, Indiana University. (2005).

28. Linnaeus, C. *Species Plantarum 2: 1057.* (Laurentii Salvii, Stockholm, 1753).

29. de Lamarck, J. B. *Encyclopédie Méthodique de Botanique, vol. 1, pt. 2. Paris, France, pp. 694–695*. (1785).

30. Small, E. American law and the species problem in Cannabis: science and semantics. *Bulletin on Narcotics* **27**, 1–20 (1975).

31. Schultes, R. E., Klein, W. M., Plowman, T. & Lockwood, T. E. Cannabis: an example of taxonomic neglect. *Botanical Museum leaflets, Harvard University.* **23**, 337--367 (1974).

32. Anderson, L. C. Leaf variation among cannabis species from a controlled garden. *Botanical Museum Leaflets, Harvard University* **28**, 61–69 (1980).

33. Small, E. & Cronquist, A. A practical and natural taxonomy for Cannabis. *Taxon* 405–435 (1976).

34. Small, E. Interfertility and chromosomal uniformity in Cannabis. *Canadian Journal of Botany* (1972) doi:10.1139/b72-248.

35. Species | Definition of Species by Oxford Dictionary on Lexico.com also meaning of Species. *Lexico Dictionaries | English* https://www.lexico.com/definition/species.

36. Emboden, W. A. The genus *Cannabis* and the correct use of taxonomic categories. *J Psychoactive Drugs* **13**, 15–21 (1981).

37. 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* **35**, 169–177 (1995).

38. Faeti, V., Mandolino, G. & Ranalli, P. Genetic diversity of *Cannabis sativa* germplasm based on RAPD markers. *Plant Breeding* **115**, 367–370 (1996).

39. Jagadish, V., Robertson, J. & Gibbs, A. RAPD analysis distinguishes *Cannabis sativa* samples from different sources. *Forensic Science International* **79**, 113–121 (1996).

40. Shirota, O. *et al.* Random Amplified Polymorphic DNA and Restriction Fragment Length Polymorphism analyses of *Cannabis sativa*. *Journal of Natural Medicines* **52**, 160–166 (1998).

41. Forapani, S. *et al.* Comparison of Hemp Varieties Using Random Amplified Polymorphic DNA Markers. *Crop Science* **41**, 1682–1689 (2001).

42. de Meijer, E. P. M. *et al.* The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics* **163**, 335–346 (2003).

43. Piluzza, G., Delogu, G., Cabras, A., Marceddu, S. & Bullitta, S. Differentiation between fiber and drug types of hemp (*Cannabis sativa* L.) from a collection of wild and domesticated accessions. *Genetic resources and crop evolution* **60**, 2331–2342 (2013).

44. 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* **98**, 86–92 (1999).

45. Flachowsky, H., Schumann, E., Weber, W. E. & Peil, A. Application of AFLP for the detection of sex-specific markers in hemp. *Plant Breeding* **120**, 305–309 (2001).

46. Datwyler, S. L. & Weiblen, G. D. Genetic variation in hemp and marijuana (*Cannabis sativa* L.) according to amplified fragment length polymorphisms. *J Forensic Sci* **51**, 371–375 (2006).

47. Chen Xuan, Y. A. of A. S., Hu Zunhong, Y. A. of A. S. & Guo Hongyan, Y. A. of A. S. Extraction of DNA and construction of AFLP reaction system in hemp. *Molecular Plant Breeding* (2010).

48. Yang, M. Genetic Diversity of *Cannabis sativa* L. Based on AFLP Analysis. *Journal of Plant Genetic Resources* (2012).

49. Hm, H. *et al.* A highly polymorphic STR locus in *Cannabis sativa*. *Forensic Sci Int* **131**, 53–58 (2003).

50. Alghanim, H. J. & Almirall, J. R. Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. *Anal Bioanal Chem* **376**, 1225–1233 (2003).

51. Gilmore, S. & Peakall, R. Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana). *Molecular Ecology Notes* **3**, 105–107 (2003).

52. 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* **131**, 65–74 (2003).

53. Chandra, S. *et al.* Analysis of Genetic Diversity using SSR Markers and Cannabinoid Contents in Different Varieties of *Cannabis sativa* L. *Planta Med* **77**, P_5 (2011).

54. Gao, C. *et al.* Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. *PLoS One* **9**, e110638 (2014).

55. Kojoma, M., Iida, O., Makino, Y., Sekita, S. & Satake, M. DNA fingerprinting of *Cannabis sativa* using inter-simple sequence repeat (ISSR) amplification. *Planta Med* **68**, 60–63 (2002).

56. Hakki, E. E., Kayis, S. A., Pinarkara, E. & Sag, A. Inter simple sequence repeats separate efficiently hemp from marijuana (*Cannabis sativa* L.). *Electronic Journal of Biotechnology* **10**, 570–581 (2007).

57. Lata, H., Chandra, S., Techen, N., Khan, I. A. & ElSohly, M. A. Assessment of the genetic stability of micropropagated plants of *Cannabis sativa* by ISSR markers. *Planta Med* **76**, 97–100 (2010).

58. Zhang, L. G. *et al.* Analysis of the genetic diversity of Chinese native *Cannabis sativa* cultivars by using ISSR and chromosome markers. *Genet Mol Res* **13**, 10490–10500 (2014).

59. Small, E. & Beckstead, H. D. Letter: Cannabinoid phenotypes in *Cannabis sativa*. *Nature* **245**, 147–148 (1973).

60. Small, E. & Beckstead, H. D. Common cannabinoid phenotypes in 350 stocks of Cannabis. *Lloydia* **36**, 144–165 (1973).

61. Fournier, G., Richez-Dumanois, C., Duvezin, J., Mathieu, J. P. & Paris, M. Identification of a new chemotype in *Cannabis sativa*: cannabigerol-dominant plants, biogenetic and agronomic prospects. *Planta Med* **53**, 277–280 (1987).

62. Mandolino, G. & Carboni, A. Potential of marker-assisted selection in hemp genetic improvement. *Euphytica* **140**, 107–120 (2004).

63. de Meijer, E. P. M. & Hammond, K. M. The inheritance of chemical phenotype in *Cannabis sativa* L. (II): Cannabigerol predominant plants. *Euphytica* **145**, 189–198 (2005).

64. de Meijer, E. P. M., Hammond, K. M. & Sutton, A. The inheritance of chemical phenotype in *Cannabis sativa* L. (IV): cannabinoid-free plants. *Euphytica* **168**, 95–112 (2009).

65. de Meijer, E. P. M., Hammond, K. M. & Micheler, M. The inheritance of chemical phenotype in *Cannabis sativa* L. (III): variation in cannabichromene proportion. *Euphytica* **165**, 293–311 (2009).

66. de Meijer, E. P. M. & Hammond, K. M. The inheritance of chemical phenotype in *Cannabis sativa* L. (V): regulation of the propyl-/pentyl cannabinoid ratio, completion of a genetic model. *Euphytica* **210**, 291–307 (2016).

67. McPartland, J. M. & Guy, G. W. Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names. *The Botanical Review* **4**, 327–381 (2017).

68. Rahn, B., Pearson, B. J., Trigiano, R. N. & Gray, D. J. The derivation of modern cannabis varieties. *Critical reviews in plant sciences* **35**, 328–348 (2016).

69. McPartland, J. M. *Cannabis sativa* and *Cannabis indica* versus “Sativa” and “Indica”. in *Cannabis sativa L.-Botany and Biotechnology* 101–121 (Springer, 2017).

70. 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* **71**, 2058–2073 (2010).

71. Hazekamp, A. & Fischedick, J. T. Cannabis - from cultivar to chemovar. *Drug Testing and Analysis* **4**, 660–667 (2012).

72. Hazekamp, A., Tejkalová, K. & Papadimitriou, S. Cannabis: From Cultivar to Chemovar II—A Metabolomics Approach to Cannabis Classification. *Cannabis and Cannabinoid Research* **1**, 202–215 (2016).

73. Hillig, K. W. Genetic evidence for speciation in *Cannabis (Cannabaceae)*. *Genetic Resources and Crop Evolution* **52**, 161–180 (2005).

74. Knight, G. *et al.* The results of an experimental indoor hydroponic Cannabis growing study, using the 'Screen of Green'(ScrOG) method—Yield, tetrahydrocannabinol (THC) and DNA analysis. *Forensic science international* **202**, 36–44 (2010).

75. Sawler, J. *et al.* The genetic structure of marijuana and hemp. *PloS one* **10**, e0133292 (2015).

76. Lynch, R. C. *et al.* Genomic and chemical diversity in Cannabis. *Critical Reviews in Plant Sciences* **35**, 349–363 (2016).

77. Henry, P. Genome-wide analyses reveal clustering in Cannabis cultivars: the ancient domestication trilogy of a panacea. (2015) doi:https://doi.org/10.7287/peerj.preprints.1553v2.

78. Aizpurua-Olaizola, O. S. Evolution of the cannabinoid and terpene content during the growth of *Cannabis sativa* plants from different chemotypes. *Journal of natural products* **79**, 324–331 (2016).

79. Upton, R. *et al. Cannabis inflorescence: cannabis spp.; standards of identity, analysis, and quality control*. (American Herbal Pharmacopoeia, 2014).

80. Russo, E. B. & Marcu, J. Cannabis Pharmacology: The Usual Suspects and a Few Promising Leads. in *Cannabinoid Pharmacology* 67–134 (Elsevier, 2017). doi:10.1016/bs.apha.2017.03.004.

81. Russo, E. B. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *British journal of pharmacology* **163**, 1344–1364 (2011).

82. McPartland, J. M. & Russo, E. B.-B. Cannabis and Cannabis Extracts : Greater Than the Sum of Their Parts ? in (2002).

83. Fischedick, J. T. Identification of Terpenoid Chemotypes Among High (-)-trans-Δ9-Tetrahydrocannabinol-Producing *Cannabis sativa* L. Cultivars. *Cannabis and cannabinoid research* **2**, 34–47 (2017).

84. Jin, D., Jin, S., Yu, Y., Lee, C. & Chen, J. Classification of Cannabis Cultivars Marketed in Canada for Medical Purposes by Quantification of Cannabinoids and Terpenes Using HPLC-DAD and GC-MS. *J Anal Bioanal Tech* **8**, 2 (2017).

85. Flores-Sanchez, I. J. & Verpoorte, R. PKS Activities and Biosynthesis of Cannabinoids and Flavonoids in *Cannabis sativa* L. Plants. *Plant and Cell Physiology* **49**, 1767–1782 (2008).

86. Vergara, D. *et al.* Genetic and genomic tools for *Cannabis sativa*. *Critical Reviews in Plant Sciences* **35**, 364–377 (2016).

87. Henry, P. Cannabis chemovar classification: terpenes hyper-classes and targeted genetic markers for accurate discrimination of flavours and effects. *PeerJ Preprints* **5**, e3307v1 (2017).

88. Hotelling, H. Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936).

89. Ross, S. A. *et al.* Flavonoid glycosides and cannabinoids from the pollen of *Cannabis sativa* L. *Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques* **16**, 45–48 (2005).

90. ElSohly, M. A. & Slade, D. Chemical constituents of marijuana: the complex mixture of natural cannabinoids. *Life sciences* **78**, 539–548 (2005).

91. ElSohly, M. A. & Gul, W. Constituents of *Cannabis Sativa*. in *Handbook of Cannabis* 3–22 (Oxford University Press, 2014). doi:10.1093/acprof:oso/9780199662685.003.0001.

92. Pollastro, F., Minassi, A. & Fresu, L. G. Cannabis Phenolics and their Bioactivities. *Current medicinal chemistry* **25**, 1160–1185 (2018).

93. Andre, C. M., Hausman, J.-F. & Guerriero, G. *Cannabis sativa*: The Plant of the Thousand and One Molecules. *Frontiers in Plant Science* **7**, (2016).

94. Lucas, P. Cannabis as an adjunct to or substitute for opiates in the treatment of chronic pain. *Journal of psychoactive drugs* **44**, 125–133 (2012).

95. Davis, W. M. & Hatoum, N. S. Neurobehavioral actions of cannabichromene and interactions with delta 9-tetrahydrocannabinol. *General pharmacology* **14**, 247–252 (1983).

96. Booz, G. W. Cannabidiol as an emergent therapeutic strategy for lessening the impact of inflammation on oxidative stress. *Free Radical Biology and Medicine* **51**, 1054–1061 (2011).

97. Pacher, P., Bátkai, S. & Kunos, G. The endocannabinoid system as an emerging target of pharmacotherapy. *Pharmacol. Rev.* **58**, 389–462 (2006).

98. Hampson, A. J., Grimaldi, M., Axelrod, J. & Wink, D. Cannabidiol and (-) Δ9-tetrahydrocannabinol are neuroprotective antioxidants. *Proceedings of the National Academy of Sciences* **95**, 8268–8273 (1998).

99. Jones, N. A. *et al.* Cannabidiol displays antiepileptiform and antiseizure properties in vitro and in vivo. *Journal of Pharmacology and Experimental Therapeutics* **332**, 569–577 (2010).

100. Deyo, R. & Musty, R. A cannabichromene (CBC) extract alters behavioral despair on the mouse tail suspension test of depression. in *Proceedings 2003 Symposium on the Cannabinoids. International Cannabinoid Research Society: Cornwall, ON* 146 (2003).

101. Campos, A. C. & Guimarães, F. S. Involvement of 5HT1A receptors in the anxiolytic-like effects of cannabidiol injected into the dorsolateral periaqueductal gray of rats. *Psychopharmacology* **199**, 223 (2008).

102. Resstel, L. B. *et al.* 5-HT1A receptors are involved in the cannabidiol-induced attenuation of behavioural and cardiovascular responses to acute restraint stress in rats. *British journal of pharmacology* **156**, 181–188 (2009).

103. Russo, E. B., Burnett, A., Hall, B. & Parker, K. K. Agonistic properties of cannabidiol at 5-HT1a receptors. *Neurochemical research* **30**, 1037–1043 (2005).

104. Rahn, E. J. & Hohmann, A. G. Cannabinoids as pharmacotherapies for neuropathic pain: from the bench to the bedside. *Neurotherapeutics* **6**, 713–737 (2009).

105. Haney, M. *et al.* Dronabinol and marijuana in HIV-positive marijuana smokers: caloric intake, mood, and sleep. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **45**, 545–554 (2007).

106. Machado Rocha, F. C., Stefano, S. C., De Cassia Haiek, R., Rosa Oliveira, L. M. Q. & Da Silveira, D. X. Therapeutic use of *Cannabis sativa* on chemotherapy-induced nausea and vomiting among cancer patients: systematic review and meta-analysis. *European journal of cancer care* **17**, 431–443 (2008).

107. Leo E. Hollister. Marihuana in Man: Three Years Later. *Science* **172**, 21 (1971).

108. Neff, G. W. *et al.* Preliminary observation with dronabinol in patients with intractable pruritus secondary to cholestatic liver disease. *The American Journal of Gastroenterology* **97**, 2117–2119 (2002).

109. PACHER, P., BÁTKAI, S. & KUNOS, G. The Endocannabinoid System as an Emerging Target of Pharmacotherapy. *Pharmacol Rev* **58**, 389–462 (2006).

110. Kavia, R. B. C., De Ridder, D., Constantinescu, C. S., Stott, C. G. & Fowler, C. J. Randomized controlled trial of Sativex to treat detrusor overactivity in multiple sclerosis. *Mult. Scler.* **16**, 1349–1359 (2010).

111. Petrocellis, L. D. *et al.* Effects of cannabinoids and cannabinoid-enriched Cannabis extracts on TRP channels and endocannabinoid metabolic enzymes. *British Journal of Pharmacology* **163**, 1479–1494 (2011).

112. Eisohly, H. N., Turner, C. E., Clark, A. M. & Eisohly, M. A. Synthesis and antimicrobial activities of certain cannabichromene and cannabigerol related compounds. *J Pharm Sci* **71**, 1319–1323 (1982).

113. Rock, E. M. *et al.* Interaction between non-psychotropic cannabinoids in marihuana: effect of cannabigerol (CBG) on the anti-nausea or anti-emetic effects of cannabidiol (CBD) in rats and shrews. *Psychopharmacology (Berl.)* **215**, 505–512 (2011).

114. Brand, E. J. & Zhao, Z. Cannabis in Chinese Medicine: Are Some Traditional Indications Referenced in Ancient Literature Related to Cannabinoids? *Frontiers in Pharmacology* **8**, (2017).

115. Ryz, N. R., Remillard, D. J. & Russo, E. B. Cannabis Roots: A Traditional Therapy with Future Potential for Treating Inflammation and Pain. *Cannabis and Cannabinoid Research* **2**, 210–216 (2017).

116. Brand, E. & Wiseman, N. *Concise Chinese materia medica*. (Paradigm Publications, 2008).

117. Smith, F. P. & Stuart, G. A. *Chinese materia medica: vegetable kingdom*. (American Presbyterian Mission Press, 1911).

118. Fellermeier, M., Eisenreich, W., Bacher, A. & Zenk, M. H. Biosynthesis of cannabinoids Incorporation experiments with 13C-labeled glucoses. *European Journal of Biochemistry* **268**, 1596–1604 (2001).

119. Gagne, S. J. *et al.* Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proceedings of the National Academy of Sciences* **109**, 12811–12816 (2012).

120. Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W. & Page, J. E. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. *The Plant Journal* (2012) doi:10.1111/j.1365-313x.2012.04949.x.

121. Taura, F., Morimoto, S. & Shoyama, Y. Purification and Characterization of Cannabidiolic-acid Synthase from *Cannabis sativa* L.: Biochemical Analysis of a Novel Enzyme that Catalyzes the Oxidocyclization of Cannabigerolic Acid to Cannabidiolic Acid. *Journal of Biological Chemistry* **271**, 17411–17416 (1996).

122. Taura, F., Morimoto, S., Shoyama, Y. & Mechoulam, R. First direct evidence for the mechanism of delta-1-tetrahydrocannabinolic acid biosynthesis. *J. Am. Chem. Soc.* **117**, 9766–9767 (1995).

123. Morimoto, S., Komatsu, K., Taura, F. & Shoyama, Y. Purification and characterization of cannabichromenic acid synthase from *Cannabis sativa*. *Phytochemistry* **49**, 1525–1529 (1998).

124. Flores-Sanchez, I. J. & Verpoorte, R. Secondary metabolism in cannabis. *Phytochemistry Reviews* **7**, 615–639 (2008).

125. Plazonić, A. *et al.* Identification and quantification of flavonoids and phenolic acids in burr parsley (Caucalis platycarpos L.), using high-performance liquid chromatography with diode array detection and electrospray ionization mass spectrometry. *Molecules* **14**, 2466–2490 (2009).

126. Ross, S. A. *et al.* Flavonoid glycosides and cannabinoids from the pollen of *Cannabis sativa* L. *Phytochemical Analysis* **16**, 45–48 (2005).

127. Pellati, F. *et al.* New Methods for the Comprehensive Analysis of Bioactive Compounds in *Cannabis sativa* L. (hemp). *Molecules* **23**, 2639 (2018).

128. Bolognini, D. *et al.* The plant cannabinoid Δ9-tetrahydrocannabivarin can decrease signs of inflammation and inflammatory pain in mice. *British journal of pharmacology* **160**, 677–687 (2010).

129. Musty, R. E., Karniol, I. G., Shirikawa, I., Takahashi, R. N. & Knobel, E. Interactions of delta-9-tetrahydrocannabinol and cannabinol in man. *The Pharmacology of Marihuana* **2**, 559–563 (1976).

130. Cascio, M. G. & Pertwee, R. G. Known pharmacological actions of nine nonpsychotropic phytocannabinoids. *Handbook of cannabis* **137**, (2014).

131. Turner, C. E., Elsohly, H. N., Lewis, G. S., Lopez-Santibanez, I. & Carranza, J. Constituents of *Cannabis sativa* L., XX: the cannabinoid content of Mexican variants grown in Mexico and in Mississippi, United States of America. *Bulletin On Narcotics* **34**, 45–59 (1982).

132. Andre, C. M. & Evers, Y. L. and D. Dietary Antioxidants and Oxidative Stress from a Human and Plant Perspective: A Review. *Current Nutrition & Food Science* (2010).

133. Plat, J. & Mensink, R. P. Plant stanol and sterol esters in the control of blood cholesterol levels: mechanism and safety aspects. *Am. J. Cardiol.* **96**, 15D-22D (2005).

134. Weingärtner, O., Lütjohann, D., Böhm, M. & Laufs, U. [Cholesterol homeostasis and cardiovascular risk]. *Dtsch. Med. Wochenschr.* **136**, 34–38 (2011).

135. Gylling, H. *et al.* Plant sterols and plant stanols in the management of dyslipidaemia and prevention of cardiovascular disease. *Atherosclerosis* **232**, 346–360 (2014).

136. Ras, R. T., Geleijnse, J. M. & Trautwein, E. A. LDL-cholesterol-lowering effect of plant sterols and stanols across different dose ranges: a meta-analysis of randomised controlled studies. *Br. J. Nutr.* **112**, 214–219 (2014).

137. Kozłowska, M., Gruczyńska, E., Ścibisz, I. & Rudzińska, M. Fatty acids and sterols composition, and antioxidant activity of oils extracted from plant seeds. *Food Chem* **213**, 450–456 (2016).

138. Blasco-Benito, S. *et al.* Appraising the 'entourage effect': Antitumor action of a pure cannabinoid versus a botanical drug preparation in preclinical models of breast cancer. *Biochem. Pharmacol.* **157**, 285–293 (2018).

139. McPartland, J. M. & Russo, E. B. Cannabis and Cannabis Extracts: Greater Than the Sum of Their Parts? *Journal of Cannabis Therapeutics* **1**, 103–132 (2001).

140. Gallily, R., Yekhtin, Z. & Hanuš, L. Overcoming the Bell-Shaped Dose-Response of Cannabidiol by Using Cannabis Extract Enriched in Cannabidiol. *Pharmacology & Pharmacy* **06**, 75 (2015).

141. Ginkgo. in *United States Pharmacopeia and National Formulary* (United States Pharmacopeial Convention, 2015).

142. Zellner, B. d'Acampora *et al.* Linear retention indices in gas chromatographic analysis: a review. *Flavour and Fragrance Journal* **23**, 297–314 (2008).

143. Linstrom, P. F. & Mallard, W. *NIST Chemistry WebBook-SRD 69*. (2001).

144. Bertoli, A., Tozzi, S., Pistelli, L. & Angelini, L. G. Fibre hemp inflorescences: From crop-residues to essential oil production. *Industrial Crops and Products* **32**, 329–337 (2010).

145. Brighenti, V., Pellati, F., Steinbach, M., Maran, D. & Benvenuti, S. Development of a new extraction technique and HPLC method for the analysis of non-psychoactive cannabinoids in fibre-type *Cannabis sativa* L. (hemp). *J Pharm Biomed Anal* **143**, 228–236 (2017).

146. De Backer, B. *et al.* Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 4115–4124 (2009).

147. Wells, G., Prest, H. & Russ, C. W. Signal, noise, and detection limits in mass spectrometry. *Technical Note for Agilent Technologies Inc* (2011).

148. *Handbook for calculation of measurement uncertainty in environmental laboratories (NT TR 537 - Edition 3.1) - Nordtest.info*.

149. Fischedick. Cannabinoids and Terpenes as Chemotaxonomic Markers in Cannabis. *Natural Products Chemistry & Research* **03**, (2015).

150. Whalley, B. *et al.* Use of one or a combination of phyto-cannabinoids in the treatment of epilepsy. US Patent US9066920B2. (2015).

151. Parker, L. A., Rock, E. M., Sticht, M. A., Wills, K. L. & Limebeer, C. L. Cannabinoids suppress acute and anticipatory nausea in preclinical rat models of conditioned gaping. *Clin. Pharmacol. Ther.* **97**, 559–561 (2015).

152. Pacifico, D., Miselli, F., Carboni, A., Moschella, A. & Mandolino, G. Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica* **160**, 231–240 (2008).

153. De Backer, B., Maebe, K., Verstraete, A. G. & Charlier, C. Evolution of the Content of THC and Other Major Cannabinoids in Drug Type Cannabis Cuttings and Seedlings During Growth of Plants. *Journal of forensic sciences* **57**, 918–922 (2012).

154. Adams, R. P. The serrate leaf margined Juniperus (Section Sabina) of the western hemisphere: systematics and evolution based on leaf essential oils and Random Amplified Polymorphic DNAs (RAPDs). *Biochemical systematics and ecology* **28**, 975–989 (2000).

155. Lucero, M. E., Estell, R. E. & Fredrickson, E. L. The Essential Oil Composition of Psorothamnus scoparius (A. Gray) Rydb. *Journal of Essential Oil Research* **15**, 108–111 (2003).

156. Lucero, M. E., Fredrickson, E. L., Estell, R. E., Morrison, A. A. & Richman, D. B. Volatile Composition of Gutierrezia sarothrae (Broom Snakeweed) as Determined by Steam Distillation and Solid Phase Microextraction. *Journal of Essential Oil Research* **18**, 121–125 (2006).

157. Clark, M. N. & Bohm, B. A. Flavonoid variation in *Cannabis* L. *Botanical Journal of the Linnean Society* **79**, 249–257 (1979).

158. Vanhoenacker, G., Van Rompaey, P., De Keukeleire, D. & Sandra, P. Chemotaxonomic features associated with flavonoids of cannabinoid-free cannabis (*Cannabis sativa* subsp. *sativa* L.) in relation to hops (Humulus lupulus L.). *Nat Prod Lett* **16**, 57–63 (2002).

159. Werz, O. *et al.* Cannflavins from hemp sprouts, a novel cannabinoid-free hemp food product, target microsomal prostaglandin E2 synthase-1 and 5-lipoxygenase. *PharmaNutrition* **2**, 53–60 (2014).

160. Slatkin, D. J. *et al.* Chemical constituents of *Cannabis sativa* L. root. *Journal of pharmaceutical sciences* **60**, 1891–1892 (1971).

161. Hasler, A., Sticher, O. & Meier, B. Identification and determination of the flavonoids from Ginkgo biloba by high-performance liquid chromatography. *Journal of Chromatography A* **605**, 41–48 (1992).

162. Antonisamy, P., Duraipandiyan, V. & Ignacimuthu, S. Anti-inflammatory, analgesic and antipyretic effects of friedelin isolated from Azima tetracantha Lam. in mouse and rat models. *Journal of Pharmacy and Pharmacology* **63**, 1070–1077 (2011).

163. Chauhan, S. m. s., Singh, M. & Narayan, L. Isolation of 3β-hydroxyolean-12-ene, friedelin and 7-methoxy-4′-5-dihydroxyisoflavone from dry and fresh leaves of Shorea robusta. *Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry* **41**, 1097–1099 (2002).

164. Weston-Green, K. The United Chemicals of Cannabis: Beneficial Effects of Cannabis Phytochemicals on the Brain and Cognition. *Recent Advances in Cannabinoid Research* (2018) doi:10.5772/intechopen.79266.

165. Zhou, X., Wang, F., Zhou, R., Song, X. & Xie, M. Apigenin: A current review on its beneficial biological activities. *Journal of Food Biochemistry* **41**, e12376 (2017).

166. He, M. *et al.* A review on the pharmacological effects of vitexin and isovitexin. *Fitoterapia* **115**, 74–85 (2016).

167. Sharma, A., Kashyap, D., Sak, K., Tuli, H. S. & Sharma, A. K. Therapeutic charm of quercetin and its derivatives: a review of research and patents. *Pharmaceutical patent analyst* **7**, 15–32 (2018).

168. Ashaari, Z. *et al.* The flavone luteolin improves central nervous system disorders by different mechanisms: a review. *Journal of Molecular Neuroscience* **65**, 491–506 (2018).

169. Vázquez, L. H., Palazon, J. & Navarro-Ocaña, A. The Pentacyclic Triterpenes α, β-amyrins: A Review of Sources and Biological Activities. In: Venketeshwer Rao (Eds.), Phytochemicals - A Global Perspective of Their Role in Nutrition and Health, 2012.

170. Baron, E. P. Comprehensive Review of Medicinal Marijuana, Cannabinoids, and Therapeutic Implications in Medicine and Headache: What a Long Strange Trip It's Been …. *Headache: The Journal of Head and Face Pain* **55**, 885–916 (2015).

171. Kuhathasan, N. *et al.* The use of cannabinoids for sleep: A critical review on clinical trials. *Experimental and Clinical Psychopharmacology* **27**, 383–401 (2019).

172. Gibbs, M. *et al.* Cannabis use and mania symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders* **171**, 39–47 (2015).

173. Grotenhermen, F. & Müller-Vahl, K. Medicinal Uses of Marijuana and Cannabinoids. *Critical Reviews in Plant Sciences* **35**, 378–405 (2016).

174. Ben Amar, M. Cannabinoids in medicine: A review of their therapeutic potential. *J Ethnopharmacol* **105**, 1–25 (2006).

175. Rock, E. M., Sticht, M. A. & Parker, L. A. Effect of phytocannabinoids on nausea and vomiting. *Handbook of Cannabis; Pertwee, RG, Ed.; Oxford University Press: Oxford, UK* 435–454 (2014).

176. Parker, L. A., Rock, E. M. & Limebeer, C. L. Regulation of nausea and vomiting by cannabinoids. *British journal of pharmacology* **163**, 1411–1422 (2011).

177. Brierley, D. I., Samuels, J., Duncan, M., Whalley, B. J. & Williams, C. M. Cannabigerol is a novel, well-tolerated appetite stimulant in pre-satiated rats. *Psychopharmacology* **233**, 3603–3613 (2016).

178. Rock, E. M. & Parker, L. A. Effect of low doses of cannabidiolic acid and ondansetron on LiCl-induced conditioned gaping (a model of nausea-induced behaviour) in rats. *British journal of pharmacology* **169**, 685–692 (2013).

179. Rock, E. M. *et al.* A comparison of cannabidiolic acid with other treatments for anticipatory nausea using a rat model of contextually elicited conditioned gaping. *Psychopharmacology* **231**, 3207–3215 (2014).

180. Rock, E. M. & Parker, L. A. Synergy between cannabidiol, cannabidiolic acid, and Δ^9-tetrahydrocannabinol in the regulation of emesis in the Suncus murinus (house musk shrew). *Behavioral neuroscience* **129**, 368 (2015).

181. Rock, E. M., Connolly, C., Limebeer, C. L. & Parker, L. A. Effect of combined oral doses of Δ9-tetrahydrocannabinol (THC) and cannabidiolic acid (CBDA) on acute and anticipatory nausea in rat models. *Psychopharmacology* **233**, 3353–3360 (2016).

182. Rock, E. M., Kopstick, R. L., Limebeer, C. L. & Parker, L. A. Tetrahydrocannabinolic acid reduces nausea-induced conditioned gaping in rats and vomiting in Suncus murinus. *British journal of pharmacology* **170**, 641–648 (2013).

183. McGuire, P. *et al.* Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: a multicenter randomized controlled trial. *Am J Psychiatry* **175**, 225–231 (2018).

184. French, J. *et al.* Cannabidiol (CBD) significantly reduces drop seizure frequency in Lennox-Gastaut syndrome (LGS): results of a multi-center, randomized, double-blind, placebo controlled trial (GWPCARE4)(S21.001). *Neurology* **88**, S21-001 (2017).

185. Avraham, Y. *et al.* Cannabidiol improves brain and liver function in a fulminant hepatic failure-induced model of hepatic encephalopathy in mice. *British journal of pharmacology* **162**, 1650–1658 (2011).

186. Bloomfield, M. A. P. *et al.* The effects of acute cannabidiol on cerebral blood flow and its relationship to memory: An arterial spin labelling magnetic resonance imaging study. *Journal of psychopharmacology (Oxford, England)* 269881120936419 (2020) doi:10.1177/0269881120936419.

187. Kovalchuk, I. *et al.* The genomics of cannabis and its close relatives. *Annu. Rev. Plant Biol.* **71**, 713–739 (2020).

188. Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H. & Satoh, S. Characterization; Genome Sizes and Morphology of Sex Chromosomes in Hemp (*Cannabis sativa* L.). *Cytologia* **63**, 459–464 (1998).

189. Shastry, B. S. SNPs in disease gene mapping, medicinal drug development and evolution. *Journal of Human Genetics* **52**, 871–880 (2007).

190. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).

191. Casci, T. SNPs that come in threes. *Nature Reviews Genetics* **11**, 8–8 (2010).

192. Westen, A. A. *et al.* Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Science International: Genetics* **3**, 233–241 (2009).

193. Kalbfleisch, T. *et al.* Using triallelic SNPs for determining parentage in North American yak (Bos grunniens) and estimating cattle (B. taurus) introgression. *F1000Res* **9**, 1096 (2020).

194. Phillips, C. *et al.* A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. *Forensic Science International: Genetics* **46**, 102232 (2020).

195. Henry, P. *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. *Journal of Cannabis Research* **2**, 26 (2020).

196. Laverty, K. U. *et al.* A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res* **29**, 146–156 (2019).

197. Grassa, C. J. *et al.* A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv* 458083 (2018) doi:10.1101/458083.

198. Jenkins, C. & Orsburn, B. Constructing a Draft Map of the Cannabis Proteome. *bioRxiv* 577635 (2019) doi:10.1101/577635.

199. Jenkins, C. & Orsburn, B. The Cannabis Multi-Omics Draft Map Project. *bioRxiv* 753400 (2019) doi:10.1101/753400.

200. Jenkins, C. & Orsburn, B. The First Publicly Available Annotated Genome for Cannabis plants. *bioRxiv* 786186 (2019) doi:10.1101/786186.

201. Benevenuto, J., Ferrão, L. F. V., Amadeu, R. R. & Munoz, P. How can a high-quality genome assembly help plant breeders? *Gigascience* **8**, (2019).

202. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94 (2010).

203. Henry, P. The genetic basis of the human-cannabis relationship. *bioRxiv* 287938 (2018).

204. Jin, D., Dai, K., Xie, Z. & Chen, J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Scientific Reports* **10**, 3309 (2020).

205. ASM23057v4 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCA_000230575.4/.

206. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

207. ASM23057v5 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCA_000230575.5.

208. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

209. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/. (2018).

210. Liu, N. & Zhao, H. A non-parametric approach to population structure inference using multilocus genotypes. *Human Genomics* **2**, 353 (2006).

211. Konishi, S. & Kitagawa, G. *Information Criteria and Statistical Modeling*. (Springer-Verlag, 2008). doi:10.1007/978-0-387-71887-3.

212. Deperi, S. I. *et al.* Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLoS One* **13**, e0194398 (2018).

213. *Cannabis sativa* Annotation Report. https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Cannabis_sativa/100/.

214. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425 (1987).

215. Ward, J. H., Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **58**, 236–244 (1963).

216. Clarke, R. C. & Merlin, M. D. Cannabis domestication, breeding history, present-day genetic diversity, and future prospects. *Critical Reviews in Plant Sciences* **35**, 293–327 (2016).

217. Van Bakel, H. *et al.* The draft genome and transcriptome of *Cannabis sativa*. *Genome biology* **12**, R102 (2011).

218. Zager, J. J., Lange, I., Srividya, N., Smith, A. & Lange, B. M. Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis. *Plant Physiology* **180**, 1877–1897 (2019).

219. Booth, J. K., Page, J. E. & Bohlmann, J. Terpene synthases from *Cannabis sativa*. *PLOS ONE* **12**, e0173911 (2017).

220. Hilyard, A., Lewin, S., Johnson, S., Henry, P. & Orser, C. Application of a Simple Genetic Assay to Discriminate Hemp from Drug-Type Cannabis. *Cannabis Science Tech* https://www.cannabissciencetech.com/view/application-simple-genetic-assay-discriminate-hemp-drug-type-cannabis (2019).

221. Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P. & de Meijer, E. Qualitative and quantitative aspects of the inheritance of chemical phenotype in Cannabis. *Journal of Industrial Hemp* **8**, 51–72 (2003).

222. Turner, C. E., Elsohly, M. A., Cheng, P. C. & Lewis, G. Constituents of *Cannabis sativa* L., XIV: Intrinsic problems in classifying cannabis based on a single cannabinoid analysis. *J. Nat. Prod.* **42**, 317–319 (1979).

223. Baram, L. *et al.* The heterogeneity and complexity of Cannabis extracts as antitumor agents. *Oncotarget* **10**, 4091–4106 (2019).

224. Berman, P. *et al.* A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in Cannabis. *Sci Rep* **8**, 1–15 (2018).

225. Jin, D., Jin, S. & Chen, J. Cannabis Indoor Growing Conditions, Management Practices, and Post-Harvest Treatment: A Review. *American Journal of Plant Sciences* **10**, 925–946 (2019).

226. Potter, D. J. The propagation, characterisation and optimisation of *Cannabis sativa* L. as a phytopharmaceutical. (King's College, 2009).

227. Caplan, D. M. Propagation and root zone management for controlled environment cannabis production. (University of Guelph, 2018).

228. Jolliffe, I. T. *Principal Component Analysis*. (Springer-Verlag, 2002). doi:10.1007/b98835.

229. Jikomes, N. & Zoorob, M. The Cannabinoid Content of Legal Cannabis in Washington State Varies Systematically Across Testing Facilities and Popular Consumer Products. *Scientific Reports* **8**, 4519 (2018).

230. Richins, R. D., Rodriguez-Uribe, L., Lowe, K., Ferral, R. & O'Connell, M. A. Accumulation of bioactive metabolites in cultivated medical Cannabis. *PLOS ONE* **13**, e0201119 (2018).

231. Mallada Frechín, J. Effect of tetrahydrocannabinol: cannabidiol oromucosal spray on activities of daily living in multiple sclerosis patients with resistant spasticity: a retrospective, observational study. *Neurodegenerative disease management* (2018).

232. Callejas, G. H. *et al.* Maternal administration of cannabidiol promotes an anti-inflammatory effect on the intestinal wall in a gastroschisis rat model. *Brazilian Journal of Medical and Biological Research* **51**, (2018).

233. Couch, D. G., Tasker, C., Theophilidou, E., Lund, J. N. & O'Sullivan, S. E. Cannabidiol and palmitoylethanolamide are anti-inflammatory in the acutely inflamed human colon. *Clinical Science* **131**, 2611–2626 (2017).

234. Vallée, A., Lecarpentier, Y., Guillevin, R. & Vallée, J.-N. Effects of cannabidiol interactions with Wnt/β-catenin pathway and PPARγ on oxidative stress and neuroinflammation in Alzheimer's disease. *Acta biochimica et biophysica Sinica* **49**, 853–866 (2017).

235. Jin, D., Henry, P., Shan, J. & Chen, J. Identification of Phenotypic Characteristics in Three Chemotype Categories in the Genus Cannabis. *HortScience* **1**, 1–10 (2021).

236. Elzinga, S., Fischedick, J., Podkolinski, R. & Raber, J. C. Cannabinoids and Terpenes as Chemotaxonomic Markers in Cannabis. (2015) doi:10.4172/2329-6836.1000181.

237. Borrelli, F. *et al.* Beneficial effect of the non-psychotropic plant cannabinoid cannabigerol on experimental inflammatory bowel disease. *Biochemical pharmacology* **85**, 1306–1316 (2013).

238. DeLong, G. T., Wolf, C. E., Poklis, A. & Lichtman, A. H. Pharmacological evaluation of the natural constituent of *Cannabis sativa*, cannabichromene and its modulation by Δ9-tetrahydrocannabinol. *Drug and alcohol dependence* **112**, 126–133 (2010).

239. De Petrocellis, L. *et al.* Cannabinoid actions at TRPV channels: effects on TRPV3 and TRPV4 and their potential relevance to gastrointestinal inflammation. *Acta Physiologica* **204**, 255–266 (2012).

240. Evans, F. J. Cannabinoids: the separation of central from peripheral effects on a structural basis. *Planta medica* **57**, S60–S67 (1991).

241. Dwivedi, C. & Harbison, R. D. Anticonvulsant activities of Δ-8 and Δ-9 tetrahydrocanabinol and uridine. *Toxicology and applied pharmacology* **31**, 452–458 (1975).

242. Hill, T. D. M. *et al.* Cannabidivarin-rich cannabis extracts are anticonvulsant in mouse and rat via a CB1 receptor-independent mechanism. *Br J Pharmacol* **170**, 679–692 (2013).

243. Hill, A. J. *et al.* Δ9-Tetrahydrocannabivarin suppresses in vitro epileptiform and in vivo seizure activity in adult rats. *Epilepsia* **51**, 1522–1532 (2010).

244. Gugliandolo, A., Pollastro, F., Grassi, G., Bramanti, P. & Mazzon, E. In Vitro Model of Neuroinflammation: Efficacy of Cannabigerol, a Non-Psychoactive Cannabinoid. *International journal of molecular sciences* **19**, 1992 (2018).

245. Miguel, M. G. Antioxidant and anti-inflammatory activities of essential oils: a short review. *Molecules* **15**, 9252–9287 (2010).

246. Xiao, R.-Y. *et al.* Screening of analgesic and anti-inflammatory active component in Fructus Alpiniae zerumbet based on spectrum–effect relationship and GC–MS. *Biomedical Chromatography* **32**, e4112 (2018).

247. Ames-Sibin, A. P. *et al.* β-Caryophyllene, the major constituent of copaiba oil, reduces systemic inflammation and oxidative stress in arthritic rats. *J. Cell. Biochem.* **119**, 10262–10277 (2018).

248. Gouveia, D. N. *et al.* α-Terpineol reduces cancer pain via modulation of oxidative stress and inhibition of iNOS. *Biomedicine & Pharmacotherapy* **105**, 652–661 (2018).

249. Shahriari, M., Zibaee, A., Sahebzadeh, N. & Shamakhi, L. Effects of α-pinene, trans-anethole, and thymol as the essential oil constituents on antioxidant system and acetylcholine esterase of Ephestia kuehniella Zeller (Lepidoptera: Pyralidae). *Pesticide Biochemistry and Physiology* (2018).

250. Klopell, F. C. *et al.* Nerolidol, an antiulcer constituent from the essential oil of Baccharis dracunculifolia DC (Asteraceae). *Z. Naturforsch., C, J. Biosci.* **62**, 537–542 (2007).

251. Tambe, Y., Tsujiuchi, H., Honda, G., Ikeshiro, Y. & Tanaka, S. Gastric cytoprotection of the non-steroidal anti-inflammatory sesquiterpene, beta-caryophyllene. *Planta Med.* **62**, 469–470 (1996).

252. Govindarajan, M., Rajeswary, M., Hoti, S. L., Bhattacharyya, A. & Benelli, G. Eugenol, α-pinene and β-caryophyllene from Plectranthus barbatus essential oil as eco-friendly larvicides against malaria, dengue and Japanese encephalitis mosquito vectors. *Parasitol. Res.* **115**, 807–815 (2016).

253. Santiago, M., Sachdev, S., Arnold, J. C., McGregor, I. S. & Connor, M. Absence of Entourage: Terpenoids Commonly Found in *Cannabis sativa* Do Not Modulate the Functional Activity of Δ9-THC at Human CB1 and CB2 Receptors. *Cannabis Cannabinoid Res* **4**, 165–176 (2019).

254. Yao, Z.-H., Yao, X., Zhang, Y., Zhang, S. & Hu, J. Luteolin Could Improve Cognitive Dysfunction by Inhibiting Neuroinflammation. *Neurochemical research* **43**, 806–820 (2018).

255. Hayasaka, N. *et al.* Absorption and Metabolism of Luteolin in Rats and Humans in Relation to in Vitro Anti-inflammatory Effects. *Journal of agricultural and food chemistry* **66**, 11320–11329 (2018).

256. Wang, J. *et al.* Kaempferol protects against gamma radiation-induced mortality and damage via inhibiting oxidative stress and modulating apoptotic molecules in vivo and vitro. *Environmental Toxicology And Pharmacology* **60**, 128–137 (2018).

257. Chirumbolo, S. & Bjørklund, G. Quercetin in collagen-induced arthritis. Some comments. *International immunopharmacology* **62**, 335 (2018).

258. Yang, Y. *et al.* Quercetin attenuates collagen-induced arthritis by restoration of Th17/Treg balance and activation of Heme Oxygenase 1-mediated anti-inflammatory effect. *Int. Immunopharmacol.* **54**, 153–162 (2018).

259. Haleagrahara, N. *et al.* Therapeutic effect of quercetin in collagen-induced arthritis. *Biomed. Pharmacother.* **90**, 38–46 (2017).

260. Strada, C. L. *et al.* Isovitexin as marker and bioactive compound in the antinociceptive activity of the Brazilian crude drug extracts of Echinodorus scaber and E. grandiflorus. *Revista Brasileira de Farmacognosia* **27**, 619–626 (2017).

261. An, F., Yang, G., Tian, J. & Wang, S. Antioxidant effects of the orientin and vitexin in Trollius chinensis Bunge in D-galactose-aged mice. *Neural regeneration research* **7**, 2565 (2012).

262. Zheng, Y.-Z., Chen, D.-F., Deng, G., Guo, R. & Fu, Z.-M. The surrounding environments on the structure and antioxidative activity of luteolin. *Journal Of Molecular Modeling* **24**, 149–149 (2018).

263. Tsai, M.-S. *et al.* Kaempferol protects against propacetamol-induced acute liver injury through CYP2E1 inactivation, UGT1A1 activation, and attenuation of oxidative stress, inflammation and apoptosis in mice. *Toxicology Letters* **290**, 97–109 (2018).

264. Kusari, P., Kusari, S., Spiteller, M. & Kayser, O. Endophytic fungi harbored in *Cannabis sativa* L.: diversity and potential as biocontrol agents against host plant-specific phytopathogens. *Fungal Diversity* **60**, 137–151 (2013).

265. McPartland, J. M. McPartland: Common names for diseases of *Cannabis sativa* L. *Plant Dis.* 75:226-227 (1991).

266. McPartland, J. M. McPartland: A review of Cannabis diseases. *Journal of the International Hemp Association* 3(1): 19-23 (1996).

267. Punja, Z. K. *et al.* Pathogens and Molds Affecting Production and Quality of *Cannabis sativa* L. *Front. Plant Sci.* **10**, (2019).

268. Hillig, K. W. & Mahlberg, P. G. A chemotaxonomic analysis of cannabinoid variation in Cannabis (Cannabaceae). *American Journal of Botany* **91**, 966–975 (2004).

269. Hillig, K. W. A chemotaxonomic analysis of terpenoid variation in Cannabis. *Biochemical Systematics and Ecology* **32**, 875–891 (2004).

270. Hillig, K. W. A Combined Analysis of Agronomic Traits and Allozyme Allele Frequencies for 69 Cannabis Accessions. *Journal of Industrial Hemp* **10**, 17–30 (2005).

271. Billakota, S., Devinsky, O. & Marsh, E. Cannabinoid therapy in epilepsy. *Current opinion in neurology* **32**, 220–226 (2019).

272. Kaplan, E. H., Offermann, E. A., Sievers, J. W. & Comi, A. M. Cannabidiol treatment for refractory seizures in Sturge-Weber syndrome. *Pediatric neurology* **71**, 18–23 (2017).

273. Szaflarski, J. P. *et al.* Long-term safety and treatment effects of cannabidiol in children and adults with treatment-resistant epilepsies: Expanded access program results. *Epilepsia* **59**, 1540–1548 (2018).

274. Borna, T., Salami, S. A. & Shokrpour, M. High resolution melting curve analysis revealed SNPs in major cannabinoid genes associated with drug and non-drug types of cannabis. *Biotechnology & Biotechnological Equipment* **31**, 839–845 (2017).

275. Kojoma, M., Seki, H., Yoshida, S. & Muranaka, T. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in “drug-type” and “fiber-type” *Cannabis sativa* L. *Forensic Science International* **159**, 132–140 (2006).

276. Pacifico, D. *et al.* Genetics and Marker-assisted Selection of the Chemotype in *Cannabis sativa* L. *Mol Breeding* **17**, 257–268 (2006).

277. Rotherham, D. & Harbison, S. A. Differentiation of drug and non-drug Cannabis using a single nucleotide polymorphism (SNP) assay. *Forensic Sci Int* **207**, 193–197 (2011).

278. Toth, J. A. *et al.* Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy* **12**, 213–222 (2020).

279. McKernan, K. J. *et al.* Single molecule sequencing of THCA synthase reveals copy number variation in modern drug-type *Cannabis sativa* L. *bioRxiv* 028654 (2015) doi:10.1101/028654.

280. 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* **116**, 57–68 (2015).

281. Ernest Small, Perry Y. Jui, & L. P. Lefkovitch. A Numerical Taxonomic Analysis of Cannabis with Special Reference to Species Delimitation. *Systematic Botany* **1**, 67–84 (1976).