Canadian Science Publishing | Genome

Genomics-based taxonomy to clarify cannabis classification

Journal:	*Genome*
Manuscript ID	gen-2023-0005.R1
Manuscript Type:	Review
Date Submitted by the Author:	18-Apr-2023
Complete List of Authors:	Lapierre, Éliana; Université Laval Monthony, Adrian S.; Université Laval Torkamaneh, Davoud; Université Laval Faculté des sciences de l'agriculture et de l'alimentation, Phytologie
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)
Keyword:	cannabis, taxonomy, genomics, phylogeny, population structure

Genomics-based taxonomy to clarify cannabis classification

Éliana Lapierre, Adrian S. Monthony, Davoud Torkamaneh*

Département de phytologie, Université Laval, Québec, Québec, Canada

Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada

Centre de recherche et d'innovation sur les végétaux (CRIV), Université Laval, Québec, Canada

Institut intelligence et données (IID), Université Laval, Québec, Canada

ORCID:

É. Lapierre: 0000-0003-3911-7989

A.S. Monthony: 0000-0003-0536-691X

D. Torkamaneh: 0000-0002-9782-5695

*Corresponding author: davoud.torkamaneh.1@ulaval.ca

Abstract

In the 18th century, Carolus Linnæus created a formalized system of classification of living organisms based on their anatomic relationships, which we know as taxonomic nomenclature. Historically, the genus Cannabis has been described three ways under this system: *Cannabis sativa* by C. Linnæus in 1753, *Cannabis indica* by J. B. Lamarck in 1785, and *Cannabis ruderalis* by D. E. Janischewsky in 1924, with these taxonomic classifications having been derived from physical, morphological, chemical and geographical data. Today, this confusing taxonomy has led to an ongoing debate about whether the genus *Cannabis* consists of a single species or multiple distinct species or subspecies. Recently, genome sequencing and bioinformatics have provided greater resolution of taxonomic assignments at the species level. As a result, some previously discussed classification frameworks have been brought into question. The aim of this review is to provide a historical context for the confusion surrounding the taxonomy of the genus *Cannabis* and highlight recent research on genomics-based taxonomical approaches to clarify the question of *Cannabis* taxonomy. We suggest that the latest evidence shifts away from the previous multiple species framework and points towards the genus *Cannabis* consisting of a highly diverse monotypic species.

Key words: cannabis, taxonomy, genomics, phylogeny, population structure.

2

1 Introduction

Cannabis sativa L. has a long history of human use for a variety of applications, including fiber, food, medicine, and for its psychoactive properties (Kalant 2001). Cannabis was domesticated around 8000 B.C. in Central and Eastern Asia (Berenji et al. 2013; Ren et al. 2023). The species is frequently classified and regulated based on the quantity of psychoactive cannabinoids produced and has been given many informal names to refer to different commercial and medicinal uses. In this review, when referring to the plant, we will refer to it using its scientific genus name *Cannabis*, but we will also use certain species-specific terminologies such as *indica, sativa, hemp*, *drug type* and *hybrid* in *italics* throughout the review. From a legal and regulatory standpoint, most countries where cannabis is produced choose to regulate it according to the amount of psychoactive cannabinoid delta-9-tetrahydrocannabinol (THC) produced; plants with less than 0.3% of THC are classified as *hemp*, while plants that produce 0.3% or more THC are categorized as *drug type* (Monthony et al. 2021; Torkamaneh and Jones 2021).THC has been the main point of focus for governments crafting legislation governing cannabis, despite the fact that cannabis produces over 120 different cannabinoids (Mudge et al. 2018; Hurgobin et al. 2021).

The taxonomic classification of cannabis has long been controversial. Due to its centuries of use and informal breeding, present-day cannabis varieties represent *hybrid*s of different genotypes with an incredibly diverse gene pool resulting in diverse phenotypes (Figure 1) (Peng and Shahidi 2021).

Figure 1 : Visual representation of the phenotypic diversity in mature inflorescences of 10 *Cannabis sativa* L. drug-type genotype varieties (A-J), and 2 hemp-type (K-L). Varieties are as follows: A) Hyperion 7, B) Trop2683, C) Mandarin Zkittlez, D) Grape Pupil, E) Purple Majik 5, F) Funky Cherries, G) White Wedding, H) Cres 805, I) End Game R2, J) White Truffle Pupil, K) Finola, L) Felina32. A-J varieties we're grown under controlled environment in the high-performance greenhouses of Laval University by Davoud

Torkamaneh lab. K and L were grown at University College Dublin and images were generously provided by C. Dowling, J. Shi and R. Melzer at the UCD Flower Power Lab.

Prohibition has had a great impact on domestication attempts, as years of clandestine selection and breeding by cultivators have increased the difficulty of distinguishing between modern *hybrid*s based on morphological and anatomical criteria (Mudge et al. 2018; Peng and Shahidi 2021). The popular taxonomy of *drug-type* plants, "*sativa*" and "*indica*", has also created confusion in the nomenclature of cannabis (McPartland and Guy 2017). Hybridization between "sativa" and "indica" and the prolonged restrictions on cannabis production also led to the difficulty of differentiating between them (McPartland 2018). In taxonomic studies and legal documents over time, confusion appeared around the uses of vernacular "*sativa*" with "*C. sativa*" and vernacular "*indica*" with "*C. indica*" (McPartland 2018). The names "*sativa*" and "*indica*" have been used in an ambiguous and contradictory manner and have different meanings from a nomenclature or production/consumption point of view. For example, taxonomists historically used the term "*sativa*" to describe non-narcotic plants like hemp, whereas "*indica*" has been used to describe narcotic species like marijuana (McPartland and Guy 2017). However, growers employ these terms to describe a plant's chemotype or psychoactive effects. They attribute "*sativa*" to varieties of plants which have a high THC and little or low cannabidiol (CBD) and "*indica*" to plants with medium THC and CBD (Sawler et al. 2015). This discrepancy between scientific literature and common names in the *Cannabis* genus has been seen at all levels of scientific writing including high ranking academic journals, causing confusion for consumers and scientists alike (Gould 2015; McPartland 2018). The legalization of

recreational and medicinal cannabis in many jurisdictions has created a market where producers have tried to come up with user-friendly vocabulary that helps inform consumers what to expect when consuming their products, opting for words like *indica*, *sativa* or *hybrid* (Small 2015). These naming conventions which typically tell consumers the expected THC/CBD ratio, is further complicated by markets where high THC cannabis is illegal, but the non-psychoactive cannabinoid CBD is permitted, such as in many US states(Barrus et al. 2016). The lack of an established consensus in terminology between growers, taxonomists and consumers presents a major challenge to the modern cannabis market, where a standard lexicon would help ensure that all stakeholders are on the same page and consumers are making informed choices when consuming cannabis.

In the last few decades, advances in genomics have revolutionized the field of phylogenetics and taxonomy by making available the whole genomes of the thousands of species. Currently in plants, in the public genome sequencing repositories, more than 1000 plant genome sequences, representing 788 distinct species are available (Kersey 2019; Sun et al. 2021). The full genomic information provides a unique and comprehensive source of data for an accurate and precise taxonomic classification. Here, we argue that growing availability of genomic resources developed in the last few years has provided more rigorous evidence towards ending the debate about the taxonomy of the *Cannabis* genus. We present a review of contemporary studies, many of which present strong cases for a single phenotypically diverse species in the genus that we hope will help advance this ongoing debate.

2 Modern taxonomy

Modern taxonomy is based on the widely accepted classification of living organisms involving the hierarchic ranks and categories proposed by Linnæus. He proposed an organized classification, nomenclature for naming taxa, based on a binomial system in which a strict hierarchy was established and introduced in his book *Philosophia botanica*. Every species based on his binomial system has given a name that contains two parts: the generic name (1st) and the specific name (2nd) (Mariette Manktelow 2010). Initially, the classification of living organisms was solely based on the Linnean morphological approach. However, in light of the theory of evolution proposed by Charles Darwin and Alfred Russed Wallace (Mariette Manktelow 2010) and cladistics, the Linnaean system has been greatly modified to account for evolution and the genealogical study of relationships between individuals (Ereshefsky 1997). Over the years, many parts of the Linnaean theory were abandoned or adapted to make place for more specialized analysis based on neo-systematics and cladistics concept (Systematics 1982). For the most part, what was retained of the original system of Linnæus are the rules of nomenclature, how species are named and the use of morphological traits to classify species (Ereshefsky 1997) . In the past few decades, the field of taxonomy has significantly evolved, now taxonomy is not solely based on the Linnaean morphological model, but on a more systematic and multidisciplinary approach. Today, taxonomic classifications are informed by information gained through advanced technologies such as genome sequencing, electron microscopy, chromatography, mathematical and statistical models (Cuperus 2022). A great example of this is the Angiosperm Phylogeny Group's updated classification for the orders and families of flowering plants, known as APG IV, which reflects the ongoing changes and updates in

the field of taxonomy, as well as the incorporation of new data and methods from various scientific disciplines (Chase et al. 2016). Despite ongoing efforts to modernize taxonomic system there is still an urgent need for clear and high resolutions phylogenies in many domains, including botany to elucidate and resolve remaining longstanding debates on some species relationships (Meiri and Mace 2007).

3 Taxonomy in the genomics era

Genomics is the study of the genome; the complete genetic information of an organism or a cell (Giani et al. 2020). Genomic studies gained momentum during the 1980s, benefitting from the availability of next generation-sequencing techniques and computer-based analysis of DNA (Giani et al. 2020). The ability to sequence the entire genomes ushered in a new age of genetics and bioinformatics and opened the door to new biotechnologies and greater understanding of plant and animal development and also taxonomy ("A comparative genomics multitool for scientific discovery and conservation" 2020). In relation to taxonomy and classification, comparative genomics, the branch of genomics that emerged from whole-genome sequencing (WGS) analysis, has provided unprecedented knowledge on the interrelation between species (Koonin et al. 2000) by tracking regions of genes or genomes that share similarities or differences to better understand the function and evolution of the genetic information (Hardison 2003). This has led to a new era of genomics-based taxonomy that links phylogeny and genomic data, so called phylogenomic, which has led to the reclassification and clarification of many taxonomic relationships in several species (Giani et al. 2020; Genereux et al. 2020). One such example is a phylogenomic study on the classification of the acacia tree and shrubs by Kyalangalilwa et al. (2013). The authors used phylogenomic data to study the

evolutionary relationships among Acacia species in Africa using molecular data from three plastid regions, *matK/trnK*, *trnL-trnF* and *psbA-trnH* to construct a phylogenetic tree. The phylogenetic tree showed that the African Acacia species were not monophyletic as previously thought. Instead, they formed several distinct clades, each with a different common ancestor. Based on molecular data and the phylogenetic tree, the authors proposed a revised classification for the African Acacia species, splitting them into two new genera, *Vachellia* and *Senegalia*.

Similar genomics-based phylogenomic studies have resolved and clarified the conflicting taxonomic classifications of numerous plant and animal species, including wheat (genus *Triticum* L. (Goncharov et al. 2009; Goncharov 2011), brittle star (genus *Ophioderma* (Lessios and Hendler 2022), bamboo (*Phyllostachys edulis* (Zhao et al. 2015)), fern (*Alsophila spinulosa* (Huang et al. 2022)), zokors (genus *Myospalax* ((Liu et al. 2022)), spiny water flea (*Bythotrephes leydig* (Karpowicz et al. 2021)), meadow-rue (*Thalictrum spp.* (Xiang et al. 2022)) and salvia (*Salvia spp.* (Hu et al. 2018)). In all the aforementioned organisms, phylogenomic studies helped with the discovery of additional undescribed species, high-resolution evolutionary history and precise taxonomic classification of complex genera and species, helping fill a "critical gap" in taxonomy.

4 Early taxonomy of *Cannabis*

The classification of the *Cannabis* genus and its species has long been debated (McPartland and Guy 2017). To date, numerous classifications have been proposed based on morphological, environmental, and chemical characteristics, but they broadly fall within

three predominant frameworks (Figure 2). In the first framework (Figure 2A), all the plants from the *Cannabis* genus were classified under one single species with three sub-species. However, the second framework (Figure 2B) suggests that this genus contains multiple distinct species, with many scholars debating whether it is two or three species. Finally, the third framework (Figure 2C) suggests the existence of a single phenotypically diverse species; *Cannabis sativa.* In the following section, we explore the monotypic and polytypic classification systems arising from morphological, and chemical and early molecular characteristics and how they shaped these three broad frameworks for the classification of cannabis, causing much of the confusion experienced today.

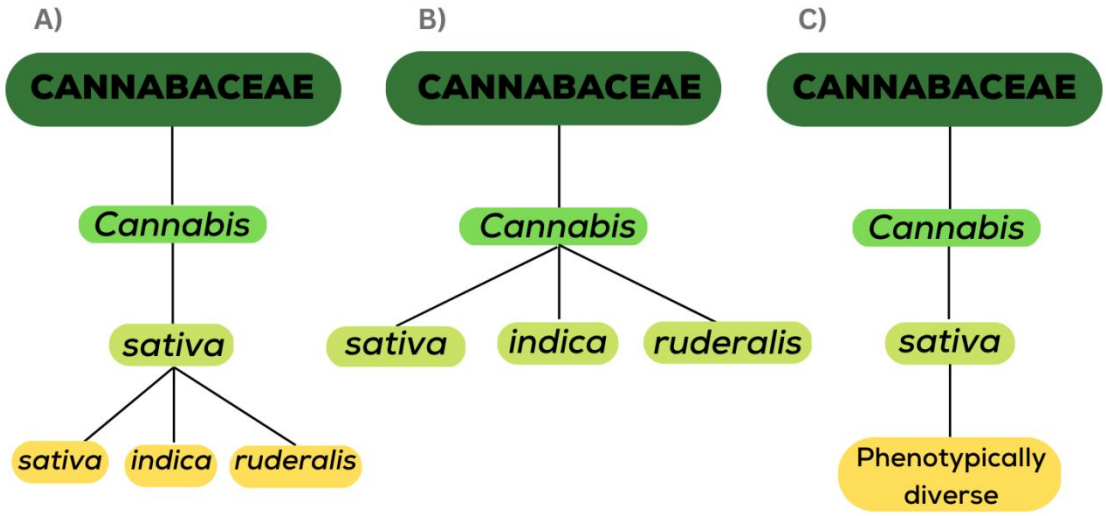


Figure 2: Three theories of classification for the Cannabaceae based on the synthesis and summary of the available articles analyzed in this global literature review. From left to

right, monotypic with three subspecies (A), polytypic consisting of up to three species (B) and single phenotypically diverse species (C).

4.1 Monotypic morphological classification

At first, *Cannabis* was identified with a systemic approach in *Species Plantarum* as *Cannabis sativa* by Linnæus in 1753 as a single species. Linnæus made a brief generic description based on flower parts and seeds (McPartland 2018). Botanist Christiaan Hendrik Persoon was one of the first to publish a monotypic *Cannabis* classification in 1807, reducing the *C. indica* variation to a subspecies of the *C. sativa* species (Erkelens and Hazekamp 2014). In 1838, Lindley also echoed the monotypic approach first suggested by Linnæus, by classifying *C. sativa* as the only member of the genus *Cannabis* (Lindley 2011). In 1843, William O'Shaughnessy in his essay on the medical use of *Cannabis* used the term Indian hemp (*C.indica*), in the title. Despite his use of the *C. indica* moniker, O'Shaughnessy claimed that there was no clear distinction between *C. sativa* and *C. indica* further supporting a monotypic classification. O'Shaughnessy's choice of the *indica* moniker may have been primarily influenced by the choice to use Indian hemp, hence the species name *indica* and could have also been influenced by the plant's roots in the history of medical use of *Cannabis* in India and other neighboring countries. In 1867, De Candolle echoed the monotypic classification of *C.sativa* as a single species. However, they proposed that the species could be grouped into four groups of *C. sativa*; (a) Kif, (b) Vulgaris, (g) Pedemontana, and (d) Chinensis. In his classification, Southern *hemp* with significant psychoactive properties were identified as group A (Kif). Northern *hemp* types

11

utilized for fiber production are classified as groups G (Pedemontana), and D (Chinensis), whereas intermediate varieties containing traits of all groups are classified as group B (Vulgaris) all under the umbrella of the single species *C.sativa* (de Candolle 1867).

4.2 Polytypic morphological classification

The introduction of the *indica* moniker by J. B. Lamarck in 1785 as well as the subsequent discussion of a ruderal species by D. E. Janischewsky in 1924 were based on further morphological observations such as trichomes or leaf shape and geographic localization which created a new framework for thinking of *Cannabis* as a polytypic species. All the authors who have presented a polytypic and monotypic classification are presented in Figure 3 and will be further detailed in this section.

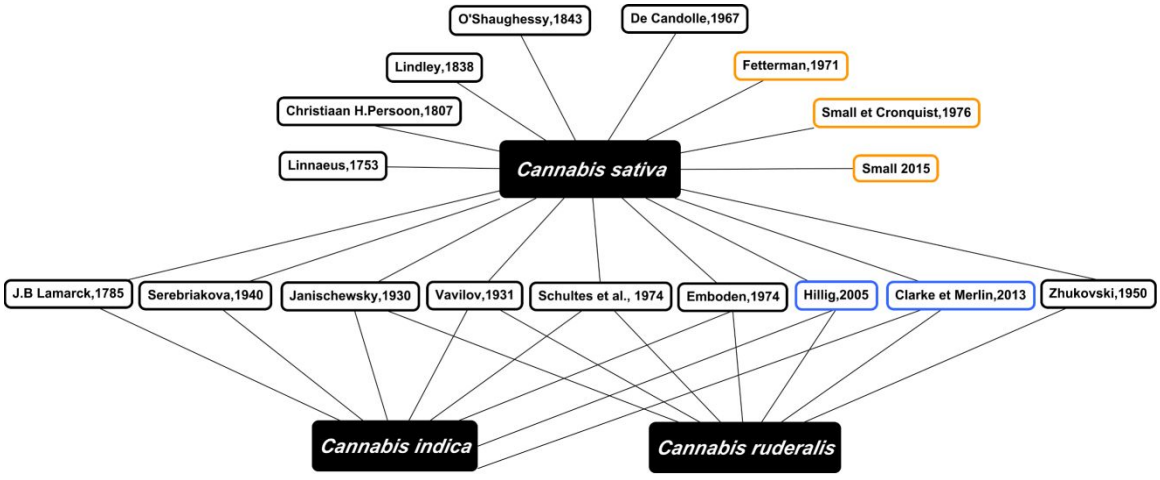


Figure 3: The three primary taxonomic classifications of the genus *Cannabis: C. sativa, C. indica* and *C. ruderalis* proposed since the study of the plant began and the authors that

have endorsed them. Black text boxes distinguish authors whose classification it was carried out using a morphological/anatomic approach, the orange outlined text boxes distinguish authors who used a chemical approach and in the blue boxes the authors used a molecular approach to classify the *Cannabis* genus.

4.2.1 Two species

In 1785, *Cannabis indica* is introduced by J. B. Lamarck which he distinguished from *Cannabis sativa* based on various morphological features including a smaller height, woodier stems, alternate ramifications of the branches, narrow leaflets, and a villous calyx in the female flowers (Lamarck and Poiret 1783; McPartland and Guy 2017). Lamarck proposed a taxonomy with two species of *C. sativa*, and *C. indica* (Lamarck and Poiret 1783; McPartland and Guy 2017). In 1940, Serebriakova has proposed a poly-species taxonomy based on morphological characteristic with two different species *C. sativa* and *C. indica* (Koren et al. 2020) which are then divided into subspecies, proles, variety and forma. Similarly, Zhukovski, in 1950, also proposed a two-species system, but with *C. sativa* L. and *C. ruderalis* first introduce by Janischewsky in 1924 (Zhukovskii 1971).

4.2.2 Three species

The introduction of *Cannabis ruderalis* by D. E. Janischewsky in the second decade of the 20th century was quick to gain attention from biologists, and soon became one of the species often cited in the genus *Cannabis*. Janischewsky's description of *C. ruderalis* was based largely on morphological features of the achenes. In botanical Latin, "ruderalis" means "weedy" or "growing among waste" and generally refers to plant species that is the first to

colonize land or as it is known as "wild species" (Schultes et al. 1975). Vavilov in 1931 similarly theorized that *C. sativa* and *C. indica* were descended from a parental species, *C. ruderalis*, and classified them as three species of the genus *Cannabis* (Koren et al. 2020). Later Schultes et al. (1975) and Emboden (1974) both independently proposed an identical three species genus that includes *C. sativa*, *C. indica*, and *C. ruderalis*. Schultes and his colleagues described *C. indica* as plants primarily found in Afghanistan (Schultes et al. 1975). These plants were characterized by their short stature, dense branching, conical shape, and broad, oblanceolate leaflets, which sharply contrasted with Lamarck's *C. indica*. Lamarck's description was based on taller and less dense plants from India, Southeast and South Africa, leading to confusion regarding the use of the name "indica" to describe different specimens (Lamarck and Poiret 1783; Schultes et al. 1975; McPartland and Guy 2017). This has contributed to the ongoing debate and uncertainty around the taxonomy of *Cannabis* species. The authors have also documented distinct morphological and anatomical characteristics for these three species based on field experiments (Emboden 1974; Schultes et al. 1975).

4.3 Chemotaxonomy of cannabis

In 1971, Fetterman et al., measured the cannabinoid content of nine varieties of *Cannabis* and then proposed a two-way chemical classification of the *Cannabis* genus based on combined physical and chemical characteristics allowing to differentiate between *drug-type* and *fiber-type C. sativa* (Fetterman et al. 1971). In 1976, Small and Cronquist proposed a novel taxonomic grouping of *Cannabis* that took not only the plant's physical characteristics and geographical distribution into account, but also THC content, so called

chemotype (Small and Cronquist 1976) This model has been adapted and updated over time in different publications (Small 1977, 2015; McPartland and Small 2020). The authors have proposed a two-stage hierarchic classification system. In this model, *C. sativa* is first divided into two varieties either *hemp* or *drug type*, which are further separated into two groups of wild and domesticated (Small and Cronquist 1976). In the first step, a distinction between two subspecies (*C. sativa* subsp. *Sativa* and *C. sativa* subsp. *Indica*) based on THC and CBD levels in dried female flowers is formed (Small and Cronquist 1976; McPartland 2018). Then in the second step, varieties in each subspecies are categorized based on their domestication status as wild or cultivated. This model is further proposed to class *Cannabis* into four categories based on cannabinoids levels. The first group is non-narcotic plants domesticated throughout Europe and Western Asia to produce fiber and oilseeds with low THC level but high amount of CBD. The second includes non-narcotic domesticated plants from East Asia to produce fiber and occasionally oilseeds. This group contains low to moderate THC, but a significant CBD content. These are classified as *C. sativa* subsp. *sativa* var. *sativa*. Non-narcotic indigenous plants domesticated in East Asia for the production of fiber and occasionally oilseeds are categorized in another group as *C. sativa* subsp. *sativa* var. *spontanea*. They have low to moderate THC, but a significant CBD content. The third category is made of the narcotic plants that have been domesticated in a vast area of South-Central Asia solely for their high THC concentration. These plants show an acute presence of domestication traits and are classified as *C. sativa* subsp. *indica* var. *indica*. Finally, domesticated naturalized narcotic plants coming from South Asia and the Middle East with similar levels of THC and CBD are identified as *C. sativa* subsp. *indica* var. *karifiristanica*. Recently, plants emerging from hybridization between the two fibrous

groups (1 and 2) and the two narcotic groups (3 and 4) are proposed to be classified in the fifth and sixth groups (Small 2015).

Classifications based on chemical content of plants were evolved and new models were proposed (Turner et al. 1979) to mainly divide the species into *fiber* or *drug-type* plants based on cannabinoids content and phenotype, known as chemovar. De Meijer et al. (2003) identified three different chemotypes based on a chemical and genetic approach which identified loci related to the chemovar profile. However, they proposed that it is highly debatable to use chemotype as a taxonomic criterion. They have also suggested that using a polygenic trait such as total cannabinoid content, rather than only THC or CBD content, is probably a more reliable trait for differentiating subspecies (de Meijer et al. 2003). In 2012, Hazekamp and Fischedick (2012) analyzed 28 primary chemical compounds in five *Cannabis* samples, followed by a principal component analysis (PCA), and proposed a chemotaxonomic classification model. Although they proposed a chemical-based classification model which can be used to distinguish *Cannabis* plants, they found a strong chemical variability between samples which had the potential to negatively impact the classification (Hillig 2005; Hazekamp and Fischedick 2012).

4.4 Molecular taxonomy of cannabis

In recent decades, genomic DNA studies using molecular markers to classify *Cannabis* have been carried out using varieties with different geographical origin. The first system of classification for *Cannabis* based on the use of molecular markers was proposed by Hillig in 2005. Hillig dismisses the notion of *Cannabis* as a monotypic genus and based on

study of allozyme variations at 17 genomic loci. His model is based on two distinct species, which he identified using the C. *sativa* and C. *indica* nomenclature. In addition, his findings suggested a possible third species, *C. ruderalis* (Hillig 2005). Hillig's conclusions were drawn from principal component analyses of the frequencies of 52 alleles in 157 samples from many different regions around the world from which they identified two significant groupings leading to a polytypic theory. Hillig also endorsed the *Cannabis* species theory presented and proposed by Lamarck (1786), Janischevsky (1924), Vavilov (1931), Schultes et al. (1975) and Anderson (1980) in their work. However, they suggested that none of the previous taxonomic theories fully describes the underlying relationships in the *Cannabis* genus. To resolve this matter, Hillig proposed a classification divided into three species *C. sativa*, *C. indica*, *C. ruderalis* and seven putative taxa's; *C. sativa* subsp. *sativa* var. *sativa, C. sativa* subsp. *sativa* var. *spontanea*, *C. sativa* subsp. *indica* var. *kafiristanica, C. indica, C. indica sensu, C. chinensis, C. ruderalis.*

In 2013, Clarke and Merlin proposed a taxonomic model based on a combination of morphological characteristics, molecular marker and chemotypes. They adopted the nomenclature used by Hillig (2005) to their findings of a mutation leading to the production of THC and CBD in the evolution of *Cannabis*. Their model of differentiation is essentially based on "dope versus rope," or narrow leaf drug (NLD), broad leaf drug (BLD), and narrow leaf *hemp* (NLH) order. Though, *hemp* produces negligible amounts of THC, both *hemp* and *drug-type Cannabis* have the allele for THC synthesis (BT) and for CBD synthesis (BD). They found that the allele BT was rather weakly expressed in *C. sativa* samples while the BD allele was highly expressed. On the contrary, BT was highly expressed in *C. indica*. The authors suggested that the drug-type *Cannabis* currently used

should be re-classified under the single species *C. indica* based on their chemotaxonomic and genetic approach in which they found the BT allele, responsible for the THC production, is usually present at higher levels in *C. indica* than in *C. sativa* (Clarke and Merlin 2013). The authors also suggested *C. ruderalis* to be the wild ancestor of *C. sativa* and *C. indica*. However their findings were based on limited sample size and genomic evolutionary analyses.

5 Contemporary genomics-based classification of *Cannabis*

5.1 Cannabis reference genome

The first reference genome assembly of drug-type *Cannabis* (cv. Purple Kush) by Van Bakel et al. in 2011 represented a landmark moment in the study of cannabis and opened the doors to advanced genetic studies and further taxonomic resolution of the genus. This draft assembly of a female plant was produced by using short read sequencing and allowed to cover ~60% (534 Mb) of the cannabis genome and annotation (draft transcriptome) of 30,000 genes (van Bakel et al. 2011). The comparison of the Purple Kush transcriptome with Finola, a *hemp*-type *Cannabis*, showed differential expression of genes encoding proteins implicated in cannabinoid and precursor pathways, suggesting that there might be a broad genetic difference between *hemp* and *drug-type Cannabis* based on cannabinoid synthesis genes. Since then, multiple chromosome-resolved reference genomes using both Purple Kush and Finola (Laverty et al. 2019), a *hemp*-type *Cannabis* (cv. CBDRx) (Grassa et al. 2021), and scaffold-resolved assembly for a wild *Cannabis* (Gao et al. 2020) were produced. Currently more than 13 assemblies (in chromosome-resolved or in scaffold level) using both short and long read sequencing are available in National Center for

18

Biotechnology Information (NCBI). In a recent review by (Hurgobin et al. 2021), the authors have comprehensively reviewed the different *Cannabis* reference genomes. The International Cannabis Genomes Research Consortium (ICGRC; 2020) has recommended using cs10 (GCA_900626175.2) assembly as reference for *Cannabis* genomics studies (Maoz TY. 2020; Grassa et al. 2021; Hurgobin et al. 2021). Kovalchuk et al. (2020) have provided a thorough discussion on the evolutionary background of *Cannabis* and *Humulus* along with a meta-analysis of published genomes. They found that part of the 45S and 5S ribosomal DNA clusters, centromeres, and satellite sequences are missing and unmapped in the currently available *Cannabis* genome assemblies. Their study presents data which highlights the necessity for a coordinated effort to assess the genetic and biochemical characterization of *Cannabis* (Kovalchuk et al. 2020).

Cannabis genome sequencing studies have shed a light onto the *Cannabis* genome organization and cannabinoid-related genes evolution, however they have not investigated the *Cannabis* genus evolution and its taxonomy, as by their nature, they tend to focus on one or two individual plants, rather than entire populations. Nevertheless, the availability of the reference genome, ever-increasing affordability of sequencing and legalization of *Cannabis* enabling large-scale genome sequencing and creating opportunities to revisit the *Cannabis* taxonomic classification. Below, we highlight recent attempts to use genomics-derived data to study *Cannabis* evolution and classification.

5.2 Cannabis phylogenomics

19

Since the publication of the first *Cannabis* reference genome over a decade ago, contemporary studies have sought to move beyond cannabinoid content, and have tried to identify genetic features mostly to distinguish between *hemp* and *drug-type Cannabis*. These represent a growing consensus that *Cannabis* cannot merely be classified based on only THC and CBD content, but rather the entire genome and metabolome play a role in distinguishing individuals within this genus (Jin et al. 2021). Additionally, these recent studies have also focused on using more advanced genomics approches (e.g., whole genome sequencing) to revisit the *sativa, indica* and *ruderalis* classification (Vergara et al. 2016; Lynch et al. 2016; Schwabe et al. 2021; Watts et al. 2021; Ren et al. 2023). Although this question remains far from being answered, these repesent an important advancement in our understanding of cannabis taxonomy and have found new evidence lending creadence to a monotypic theory: as they were not able to distinguish *C. sativa, C. indica* or *C. ruderalis* from each other.

Recently, Watts et al., (2021) genotyped 137 drug-type *Cannabis* plants using 116K single nucleotide polymorphisms (SNPs) derived from genotyping-by-sequencing (GBS) analysis. The authors have conducted a PCA analysis using genomic data and found that *sativa*- and *indica*-labelled samples were genetically indistinct on a genome-wide scale. In fact, the authors noted that based on their findings, the *sativa*/*indica* classification that is currently used to label *Cannabis* poorly captures overall genomic and metabolomic variation (Watts et al. 2021). In a comprehensive study of 110 hemp accessions from across the globe, Ren et al., (2021) investigated the domestication history of *Cannabis* using 12M SNPs derived from whole-genome resequencing. They consider *Cannabis* as a single species and demonstrate that all *hemp*-type *Cannabis* samples are reciprocally

monophyletic to all *drug-type* samples due to different breeding trajectories. Overall, their findings offer a global perspective on *C. sativa* domestication and remarkable genomic resources for functional and molecular studies, and provide evidence for a monotypic *Cannabis* taxonomic classification.

A series of recent studies conducted by Vergara et al. investigated the question of *Cannabis* modern taxonomic classification using chemical and genetic tools (Vergara et al. 2016, 2017, 2021a, 2021b). First in their 2017 study, they compared the chemical profiles of *Cannabis* plants available in the National Institute on Drug Abuse (NIDA) and those available through commercial dispensaries in the United States. The study found significant variations between cannabinoids content of *Cannabis* used in federally funded medical research and *Cannabis* that was available to the general public for recreational consumption. Importantly this highlighted the chemical diversity of contemporary "research grade" cannabis from "recreational" cannabis, which could have important implications for the findings of clinical studies on the effects of cannabinoids in human health (Vergara et al. 2017). Building off of their work in 2017, a 2021 study conducted by Vergara et al. (2021) explored the genomes of NIDA produced *Cannabis* and widely accessible medicinal, recreational and industrial *Cannabis* varieties. Once again, they showed notable differences in the genomes of NIDA and legal market *Cannabis* including a lack of diversity in the single copy region of nucleotide variation, in different cannabinoid genes and in the repetitive genome elements. Further highlighting how NIDA's varieties may not be suitable to be used to generalize the effects of *Cannabis* use in the medical context. They also found that only few plants with a 100% genetically assignment probability of their classified taxonomic group (broad-leaf marijuana-type (BLMT),

narrow-leaf marijuana-type (NLMT), *hemp*, and *hybrid*). In a subsequent study published in 2021, they demonstrated that the terms "*indica*" and "*sativa*" -as they are currently used- do not correlate to the scientific subspecies of the same names and that evolutionary links are not reflected by these conventional divisions (Vergara et al. 2021a). These conclusions were drawn from a study where 297 individuals from a first-generation backcross between a female "Carmagnola" *hemp-type* and a male "Afghan Kush" *drug-type* plant were quantified for 18 phenotypic traits. They used morphological characteristics measured by the *Cannabis* industry to distinguish and assess distinct plants and discovered that the majority of the phenotypic traits measured were not genetically associated (Vergara et al. 2021a), further suggesting a phenotypically diverse but single species.

A similar study (Schwabe and McGlaughlin 2019) aimed to find whether the distinctions that are frequently cited and commercially used, *sativa*, *indica* and *hybrid* can be genetically corroborated and whether samples bearing the same variety designation are consistent group together when collected from different suppliers. Accordingly, they used the "Purple Kush" genome, to create ten new microsatellite markers to investigate possible genetic diversity in 30 different plants gathered from dispensaries in three different US states. They found that the commonly used *sativa*, *indica*, and *hybrid* designations did not have a clear genetic difference. Within samples carrying the same variety name, significant genetic variations were found, suggesting inconsistent identification of what is offered to consumers (Schwabe and McGlaughlin 2019). In a subsequent 2021 publication by the same group, they outline a comparative study of the genetic structure of *C. sativa* with feral, wild, and cultivated samples in the USA. Using ten variable microsatellite loci developed in their 2019 study, the authors found that the present *sativa* and *indica* scale for

classifying *Cannabis* does not adequately reflect global genomic and metabolomics variations. They also suggested that the terms *sativa* and *indica* derived from taxonomic names that were originally used to describe plants based on their genealogy, but now they have been adopted by popular *Cannabis* culture and are now being used differently (Schwabe et al. 2021). Overall, they found no clear genetic distinction between *sativa*, *hybrid*, and *indica* subcategories within retail marijuana samples, underscoring the confusion that these naming conventions create for scientists and consumers.

Lynch et al. (2016) tried to explore the diversity of the nuclear genome of 340 different *Cannabis* varieties, including wild populations, high-cannabinol *drug-type*, *hemp* for fiber and *hemp* used for seed and oil production by re-mapping of WGS and GBS sequence data to the current "Purple Kush" reference genome. They concluded that there are at least three primary categories of plants broad leaflet drug-types (BLDT), narrow leaflet drug-types (NLDT) and *hemp* similar to what Clarke and Merlin proposed in their 2013 model. European *hemp* varieties were found to be more closely connected to NLDTs than to BLDTs. The BLDT group appeared to have less diversity than the NLDT, which is consistent with the NLDTs' wider geographic distribution and raises the possibility that the BLDTs were domesticated more recently than the NLDTs. The *hemp*, NLDT, and BLDT create distinctive cannabinoid and terpenoid content profiles. Although they were able to identify three distinct groups, they found many patterns of relatedness between these groups supported by evidence such as their distribution of Weir-Cockerham FST for each population (Lynch et al. 2016). Overall, with the lack of existing evidence for reproductive barriers and the generally small genetic differences between *hemp* and the *drug-type* samples examined in their study, the authors recommend continuing to treat all members

of this genus as a monotypic *C. sativa* L. They have also suggested that a thorough assessment of current worldwide genetic diversity and experimental evaluation of reproductive compatibility across all major genetic groupings, along with morphological evaluation, is necessary for the definitive resolution of the taxonomy of *Cannabis* (Lynch et al. 2016). In a similar study, Barcaccia et al. (2020) used DNA barcoding technique to explore the *Cannabis* taxonomic classification. They collected 112 sequences from which they reviewed the DNA barcoding data (ITS1, ITS2, matK and rbcL sequences) available for *Cannabis* in BOLD and GenBank databases through a keyword search and BLASTn analysis for "*Cannabis*" (taxid: 3482). Overall, they confirmed that methods based on nuclear DNA haplotyping based on the ITS1 and ITS2 markers as well as chloroplast DNA barcoding based on the standard genes (matK and rbcL) is complex and does not provide a clear answer. They concluded by confirming McPartland's (2018) assertion that botanical varieties rather than subspecies should be used to classify the *Cannabis* genus (C. *sativa* subsp. *sativa* and C. *sativa* subsp. *indica*) (Barcaccia et al. 2020), further illustrating that the contemporary debate on the taxonomic classification of cannabis is not entirely resolved.

6 Conclusion

In recent years, the field of taxonomic classification has made significant strides, particularly in the use of genomic data to infer phylogenetic relationships. While traditional morphological and phenotypic methods have been successful for many species, they have proved inadequate for resolving the classification of certain taxa, such as *Cannabis*. However, the use of genomic tools and techniques has enabled researchers to identify genetic variations and differentiate between species and subspecies. The increasing accessibility and affordability of these techniques have helped to shed light on the challenging taxonomy of *Cannabis*. The conventional taxonomy of *Cannabis* is challenging due to wind-pollination, the heterozygous nature of the plants, current growing practices and incomplete domestication paired with poorly tracked breading pedigrees. The taxonomy of *Cannabis* is a divisive matter, with various hypotheses as have been discussed in this review. Precise taxonomic information is crucial for the study of *Cannabis* and in its absence, we risk the loss of genetic diversity as historic landraces may not be identified and preserved. The emerging evidence from contemporary genomic sequencing and other molecular techniques outlined in this review, strongly support the hypothesis of a highly diverse monotypic species, and older hypotheses positing multiple species or subspecies now appear less likely in light of emerging data. Although challenging, we believe that the taxonomy of the species is becoming clearer and will continue to benefit from advances in genomic tools, as we have demonstrated in this review. Ultimately, the confirmation of a highly diverse monotypic species of cannabis will be a significant contribution to the plant's taxonomic history and will enable the preservation of valuable genetic diversity.

25

Author Contributions

EL, ASM and DT contributed to writing the manuscript. All authors read and approved the manuscript.

Funding

This work was funded by Natural Sciences and Engineering Research Council (NSERC) of Canada. A-S. Monthony has been supported by a NSERC Canada Vanier Graduate Scholarship.

Acknowledgments

The authors would like to express their gratitude to Caroline Dowling, Jiaqi Shi and Rainer Melzer at University College Dublin for kindly providing hemp photos used in Figure 1.

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

This manuscript does not report data.

7 References

A comparative genomics multitool for scientific discovery and conservation. 2020. Nature **587**(7833): 240–245. Nature Publishing Group UK London.

Alqudah, A.M., Sallam, A., Baenziger, P.S., and Börner, A. 2020. GWAS: fast-forwarding gene identification and characterization in temperate cereals: lessons from barley–a review. J Adv Res **22**: 119–135. Elsevier.

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., and Page, J.E. 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biol **12**(10): R102. doi:10.1186/gb-2011-12-10-r102.

Barcaccia, G., Palumbo, F., Scariolo, F., Vannozzi, A., Borin, M., and Bona, S. 2020. Potentials and challenges of genomics for breeding cannabis cultivars. Front Plant Sci **11**: 1472. Frontiers.

Barrus, D.G., Capogrossi, K.L., Cates, S.C., Gourdet, C.K., Peiper, N.C., Novak, S.P., Lefever, T.W., and Wiley, J.L. 2016. Tasty THC: Promises and Challenges of Cannabis Edibles. Methods Rep RTI Press **2016**. Methods Rep RTI Press. doi:10.3768/RTIPRESS.2016.OP.0035.1611.

Berenji, J., Sikora, V., Fournier, G., and Beherec, O. 2013. Genetics and selection of hemp. Hemp: Industrial Production and Uses: 48–71. doi:10.1079/9781845937935.0048.

de Candolle, A. 1867. Lois de la nomenclature botanique adoptées par le Congrès international de botanique tenu à Paris en août 1867: suivies d'une 2e édition de l'introduction historique et du commentaire qui accompagnaient la rédaction préparatoire présentée au Congrès. H. Georg.

27

Chase, M.W., Christenhusz, M.J.M., Fay, M.F., Byng, J.W., Judd, W.S., Soltis, D.E., Mabberley, D.J., Sennikov, A.N., Soltis, P.S., Stevens, P.F., Briggs, B., Brockington, S., Chautems, A., Clark, J.C., Conran, J., Haston, E., Möller, M., Moore, M., Olmstead, R., Perret, M., Skog, L., Smith, J., Tank, D., Vorontsova, M., and Weber, A. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society **181**(1): 1–20. Oxford Academic. doi:10.1111/BOJ.12385.

Clarke, R.C., and Merlin, M.D. 2013. Cannabis. Evolution and Ethnobotany, University of Cali. fornia Press, Berkeley and Los Angeles.

Cuperus, J.T. 2022. Single-cell genomics in plants: current state, future directions, and hurdles to overcome. Plant Physiol **188**(2): 749–755. doi:10.1093/plphys/kiab478.

Emboden, W.A. 1974. Cannabis: a polytypic genus. Econ Bot **28**(3): 304–310. JSTOR.

Ereshefsky, M. 1997. The evolution of the Linnaean hierarchy. Biol Philos **12**(4): 493–519. Springer.

Erkelens, J.L., and Hazekamp, A. 2014. That which we call Indica, by any other name would smell as sweet. Cannabinoids **9**(1): 9–15.

Espeland, M., Breinholt, J., Willmott, K.R., Warren, A.D., Vila, R., Toussaint, E.F.A., Maunsell, S.C., Aduse-Poku, K., Talavera, G., and Eastwood, R. 2018. A comprehensive and dated phylogenomic analysis of butterflies. Current Biology **28**(5): 770–778. Elsevier.

Fetterman, P.S., Keith, E.S., Waller, C.W., Guerrero, O., Doorenbos, N.J., and Quimby, M.W. 1971. Mississippi-grown Cannabis sativa L.: Preliminary observation on

chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part. J Pharm Sci **60**(8): 1246–1249. Elsevier.

Gao, S., Wang, B., Xie, S., Xu, X., Zhang, J., Pei, L., Yu, Y., Yang, W., and Zhang, Y. 2020. A high-quality reference genome of wild Cannabis sativa. Hortic Res **7**: 73. doi:10.1038/s41438-020-0295-3.

Genereux, D.P., Serres, A., Armstrong, J., Johnson, J., Marinescu, V.D., Murén, E., Juan, D., Bejerano, G., Casewell, N.R., Chemnick, L.G., Damas, J., Di Palma, F., Diekhans, M., Fiddes, I.T., Garber, M., Gladyshev, V.N., Goodman, L., Haerty, W., Houck, M.L., Hubley, R., Kivioja, T., Koepfli, K.P., Kuderna, L.F.K., Lander, E.S., Meadows, J.R.S., Murphy, W.J., Nash, W., Noh, H.J., Nweeia, M., Pfenning, A.R., Pollard, K.S., Ray, D.A., Shapiro, B., Smit, A.F.A., Springer, M.S., Steiner, C.C., Swofford, R., Taipale, J., Teeling, E.C., Turner-Maier, J., Alfoldi, J., Birren, B., Ryder, O.A., Lewin, H.A., Paten, B., Marques-Bonet, T., Lindblad-Toh, K., and Karlsson, E.K. 2020. A comparative genomics multitool for scientific discovery and conservation. Nature 2020 587:7833 **587**(7833): 240–245. Nature Publishing Group. doi:10.1038/s41586-020-2876-6.

Giani, A.M., Gallo, G.R., Gianfranceschi, L., and Formenti, G. 2020. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J **18**: 9–19. doi:https://doi.org/10.1016/j.csbj.2019.11.002.

Goncharov, N.P. 2011. Genus Triticum L. taxonomy: the present and the future. Plant Systematics and Evolution **295**(1): 1–11. Springer.

Goncharov, N.P., Golovnina, K.A., and Kondratenko, E.Y. 2009. Taxonomy and molecular phylogeny of natural and artificial wheat species. Breed Sci **59**(5): 492–498. Japanese Society of Breeding.

Gould, J. 2015. The cannabis crop. Nature **525**(7570): S2–S3. doi:10.1038/525S2a.

Grassa, C.J., Weiblen, G.D., Wenger, J.P., Dabney, C., Poplawski, S.G., Timothy Motley, S., Michael, T.P., and Schwartz, C.J. 2021. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. New Phytologist **230**(4): 1665–1679. Wiley Online Library.

Hardison, R.C. 2003. Comparative genomics. PLoS Biol **1**(2): e58. Public Library of Science San Francisco, USA.

Hazekamp, A., and Fischedick, J.T. 2012. Cannabis-from cultivar to chemovar. Drug Test Anal **4**(7–8): 660–667. Wiley Online Library.

Hillig, K.W. 2005. Genetic evidence for speciation in Cannabis (Cannabaceae). Genet Resour Crop Evol **52**(2): 161–180. doi:10.1007/s10722-003-4452-y.

Hu, G.-X., Takano, A., Drew, B.T., Liu, E.-D., Soltis, D.E., Soltis, P.S., Peng, H., and Xiang, C.-L. 2018. Phylogeny and staminal evolution of Salvia (Lamiaceae, Nepetoideae) in East Asia. Ann Bot **122**(4): 649–668. Oxford University Press US.

Huang, X., Wang, W., Gong, T., Wickell, D., Kuo, L.-Y., Zhang, X., Wen, J., Kim, H., Lu, F., and Zhao, H. 2022. The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. Nat Plants **8**(5): 500–512. Nature Publishing Group.

Hurgobin, B., Tamiru-Oli, M., Welling, M.T., Doblin, M.S., Bacic, A., Whelan, J., and Lewsey, M.G. 2021. Recent advances in Cannabis sativa genomics research. New

Phytologist **230**(1): 73–89. John Wiley & Sons, Ltd. doi:https://doi.org/10.1111/nph.17140.

Jin, D., Henry, P., Shan, J., and Chen, J. 2021. 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**. Frontiers Media SA.

Kalant, H. 2001. Medicinal Use of Cannabis: History and Current Status. Pain Res Manag **6**: 469629. Pulsus Group Inc. doi:10.1155/2001/469629.

Kang, M.S. 2002. Quantitative Genetics, Genomics, and Plant Breeding. CABI Pub. Available from https://books.google.ca/books?id=bdd0ngEACAAJ.

Karpowicz, M., Świsłocka, M., Moroz, J., and Sługocki, Ł. 2021. New insight into the taxonomic resolution of the genus Bythotrephes Leydig (Crustacea: Cladocera) based on molecular data from Central Europe. Sci Rep **11**(1): 1–11. Nature Publishing Group.

Kersey, P.J. 2019. Plant genome sequences: past, present, future. Curr Opin Plant Biol **48**: 1–8. doi:https://doi.org/10.1016/j.pbi.2018.11.001.

Koonin, E. v, Aravind, L., and Kondrashov, A.S. 2000. The impact of comparative genomics on our understanding of evolution. Cell **101**(6): 573–576. Elsevier.

Koren, A., Sikora, V., Kiprovski, B., Brdar-Jokanović, M., Aćimović, M., Konstantinović, B., and Latković, D. 2020. Controversial taxonomy of hemp. Genetika-Belgrade **52**(1): 1–13. Društvo genetičara Srbije, Beograd.

Kovalchuk, I., Pellino, M., Rigault, P., van Velzen, R., Ebersbach, J., Ashnest, J.R., Mau, M., Schranz, M.E., Alcorn, J., Laprairie, R.B., McKay, J.K., Burbridge, C., Schneider, D., Vergara, D., Kane, N.C., and Sharbel, T.F. 2020. The Genomics of Cannabis and

Its Close Relatives. Annu Rev Plant Biol **71**(1): 713–739. Annual Reviews. doi:10.1146/annurev-arplant-081519-040203.

Kreplak, J., Madoui, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P.E., Gali, K.K., Syme, R.A., and Main, D. 2019. A reference genome for pea provides insight into legume genome evolution. Nat Genet **51**(9): 1411–1422. Nature Publishing Group.

Kyalangalilwa, B., Boatwright, J.S., Daru, B.H., Maurin, O., and van der Bank, M. 2013. Phylogenetic position and revised classification of Acacia s.l. (Fabaceae: Mimosoideae) in Africa, including new combinations in Vachellia and Senegalia. Botanical Journal of the Linnean Society **172**(4): 500–523. John Wiley & Sons, Ltd. doi:10.1111/BOJ.12047.

Lamarck, J.-B.P.A. de M. de, and Poiret, J.-L.-M. 1783. Encyclopédie méthodique. Botanique. Panckoucke;Plomteux, Paris,Liège. Available from https://www.biodiversitylibrary.org/item/15259.

Laverty, K.U., Stout, J.M., Sullivan, M.J., Shah, H., Gill, N., Holbrook, L., Deikus, G., Sebra, R., Hughes, T.R., Page, J.E., and van Bakel, H. 2019. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Res **29**(1): 146–156. doi:10.1101/gr.242594.118.

Lessios, H.A., and Hendler, G. 2022. Mitochondrial phylogeny of the brittle star genus Ophioderma. Sci Rep **12**(1): 1–13. Nature Publishing Group.

Lindley, J. 2011. Flora medica: a botanical account of all the more important plants used in medicine, in different parts of the world. Cambridge University Press.

Liu, X., Zhang, S., Cai, Z., Kuang, Z., Wan, N., Wang, Y., Mao, L., An, X., Li, F., and Feng, T. 2022. Genomic insights into zokors' phylogeny and speciation in China. Proceedings of the National Academy of Sciences **119**(19): e2121819119. National Acad Sciences.

Lynch, R.C., Vergara, D., Tittes, S., White, K., Schwartz, C.J., Gibbs, M.J., Ruthenburg, T.C., deCesare, K., Land, D.P., and Kane, N.C. 2016. Genomic and Chemical Diversity in Cannabis. CRC Crit Rev Plant Sci **35**(5–6): 349–363. Taylor & Francis. doi:10.1080/07352689.2016.1265363.

Maoz TY. 2020. Making Cannabis History in 2020.

Mariette Manktelow. 2010. History of taxonomy.

McPartland, J.M. 2018. Cannabis Systematics at the Levels of Family, Genus, and Species. Cannabis Cannabinoid Res **3**(1): 203–212. Mary Ann Liebert, Inc., publishers. doi:10.1089/can.2018.0039.

McPartland, J.M., and Guy, G.W. 2017. Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names. The Botanical Review **83**(4): 327–381. doi:10.1007/s12229-017-9187-0.

McPartland, J.M., and Small, E. 2020. A classification of endangered high-THC cannabis (Cannabis sativa subsp. indica) domesticates and their wild relatives. PhytoKeys **144**: 81. Pensoft Publishers.

de Meijer, E.P.M., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V.M.C., Ranalli, P., and Mandolino, G. 2003. The inheritance of chemical phenotype in Cannabis sativa L. Genetics **163**(1): 335–346. Oxford University Press.

Meiri, S., and Mace, G.M. 2007. New taxonomy and the origin of species. Public Library of Science San Francisco, USA.

Mo, Y., and Jiao, Y. 2022. Advances and applications of single-cell omics technologies in plant research. The Plant Journal. Wiley Online Library.

Monthony, A., Page, S., Hesami, M., and Jones, A. 2020. The Past, Present and Future of Cannabis sativa Tissue Culture. doi:10.20944/preprints202012.0766.v1.

Monthony, A.S., Page, S.R., Hesami, M., Maxwell, A., and Jones, P. 2021. plants The Past, Present and Future of Cannabis sativa Tissue Culture. doi:10.3390/plants.

Mudge, E.M., Murch, S.J., and Brown, P.N. 2018. Chemometric Analysis of Cannabinoids: Chemotaxonomy and Domestication Syndrome. Sci Rep **8**(1): 13090. doi:10.1038/s41598-018-31120-2.

Nadeem, M.A., Nawaz, M.A., Shahid, M.Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., and Labhane, N. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology & Biotechnological Equipment **32**(2): 261–285. Taylor & Francis.

Peng, H., and Shahidi, F. 2021. Cannabis and Cannabis Edibles: A Review. J Agric Food Chem **69**(6): 1751–1774. American Chemical Society. doi:10.1021/acs.jafc.0c07472.

Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M.L., Yang, Y., Liu, A., Ravikanth, G., Nawaz, M.A., Mumtaz, A.S., Salamin, N., and Fumagalli, L. 2023. Large-scale whole-genome resequencing unravels the domestication history of Cannabis sativa. Sci Adv **7**(29): eabg2286. American Association for the Advancement of Science. doi:10.1126/sciadv.abg2286.

Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E., and Myles, S. 2015. The genetic structure of marijuana and hemp. PLoS One **10**(8): e0133292. Public Library of Science San Francisco, CA USA.

Schultes, R.E., Klein, W.M., Plowman, T., and Lockwood, T.E. 1975. Cannabis: an example of taxonomic neglect. Cannabis and culture: 21–38. Mouton Publishers, The Hague/Paris.

Schwabe, A.L., Hansen, C.J., Hyslop, R.M., and McGlaughlin, M.E. 2021. Comparative genetic structure of Cannabis sativa including federally produced, wild collected, and cultivated samples. Front Plant Sci: 2098. Frontiers.

Schwabe, A.L., and McGlaughlin, M.E. 2019. Genetic tools weed out misconceptions of strain reliability in Cannabis sativa: implications for a budding industry. J Cannabis Res **1**(1): 1–16. BioMed Central.

Small, E. 1977. Nomenclatural nonsense and legal marijuana plants [Cannabis]. Bulletin Pacific Tropical Botanical Garden.

Small, E. 2015. Evolution and Classification of Cannabis sativa (Marijuana, Hemp) in Relation to Human Utilization. The Botanical Review **81**(3): 189–294. doi:10.1007/s12229-015-9157-3.

Small, E., and Cronquist, A. 1976. A PRACTICAL AND NATURAL TAXONOMY FOR CANNABIS. Taxon **25**(4): 405–435. John Wiley & Sons, Ltd. doi:https://doi.org/10.2307/1220524.

Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., and Guo, L. 2021. Twenty years of plant genome sequencing: Achievements and challenges. Trends Plant Sci. Elsevier.

Systematics, M. 1982. Taxonomy. Microbial Ecology, Diversity.

35

Tibbs Cortes, L., Zhang, Z., and Yu, J. 2021. Status and prospects of genome-wide association studies in plants. Plant Genome **14**(1): e20077. Wiley Online Library.

Torkamaneh, D., and Jones, A.M.P. 2021. Cannabis, the multibillion dollar plant that no genebank wanted. Genome **65**(1): 1–5. NRC Research Press. doi:10.1139/gen-2021-0016.

Turner, C.E., Elsohly, M.A., Cheng, P.C., and Lewis, G. 1979. Constituents of Cannabis sativa L., XIV: Intrinsic problems in classifying cannabis based on a single cannabinoid analysis. J Nat Prod **42**(3): 317–319. ACS Publications.

Vergara, D., Baker, H., Clancy, K., Keepers, K.G., Mendieta, J.P., Pauli, C.S., Tittes, S.B., White, K.H., and Kane, N.C. 2016. Genetic and genomic tools for Cannabis sativa. CRC Crit Rev Plant Sci **35**(5–6): 364–377. Taylor & Francis.

Vergara, D., Bidwell, L., Gaudino, R., Torres, A., Du, G., Ruthenburg, T.C., DeCesare, K., Land, D.P., Hutchison, K.E., and Kane, N.C. 2017. Compromised external validity: federally produced cannabis does not reflect legal markets. Sci Rep **7**(1): 1–8. Nature Publishing Group.

Vergara, D., Feathers, C., Huscher, E.L., Holmes, B., Haas, J.A., and Kane, N.C. 2021a. Widely assumed phenotypic associations in Cannabis sativa lack a shared genetic basis. PeerJ **9**: e10672. PeerJ Inc.

Vergara, D., Huscher, E.L., Keepers, K.G., Pisupati, R., Schwabe, A.L., McGlaughlin, M.E., and Kane, N.C. 2021b. Genomic evidence that governmentally produced Cannabis sativa poorly represents genetic variation available in state markets. Front Plant Sci: 1502. Frontiers.

Watts, S., McElroy, M., Migicovsky, Z., Maassen, H., van Velzen, R., and Myles, S. 2021. Cannabis labelling is associated with genetic variation in terpene synthase genes. Nat Plants **7**(10): 1330–1334. Nature Publishing Group.

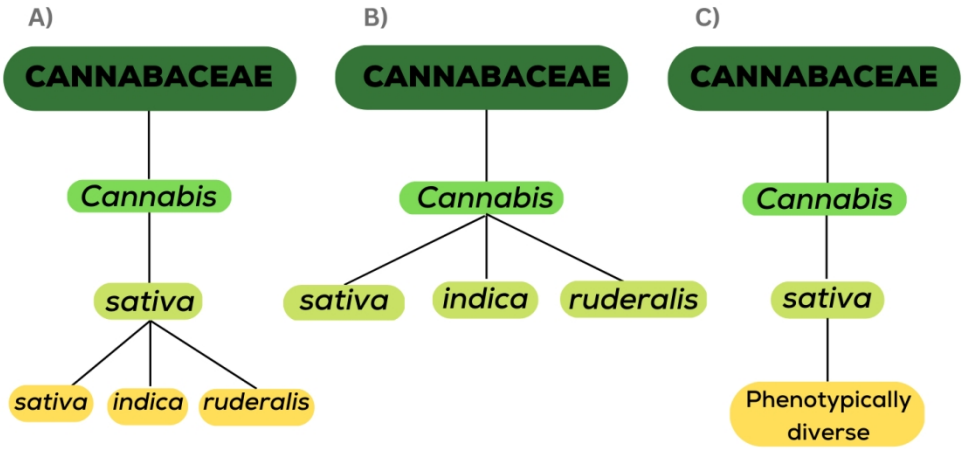
Xiang, K.-L., Mao, W., Peng, H.-W., Erst, A.S., Yang, Y.-X., He, W.-C., and Wu, Z.-Q. 2022. Organization, Phylogenetic Marker Exploitation, and Gene Evolution in the Plastome of Thalictrum (Ranunculaceae). Front Plant Sci **13**: 897843. Frontiers Media SA.

Zhao, H., Yang, L., Peng, Z., Sun, H., Yue, X., Lou, Y., Dong, L., Wang, L., and Gao, Z. 2015. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus Phyllostachys. Sci Rep **5**(1): 1–10. Nature Publishing Group.

Zhukovskii, P.M. 1971. Cultivated plants and their wild relatives. Cultivated plants and their wild relatives. Ed. 3, Leningrad, USSR, Kolos.

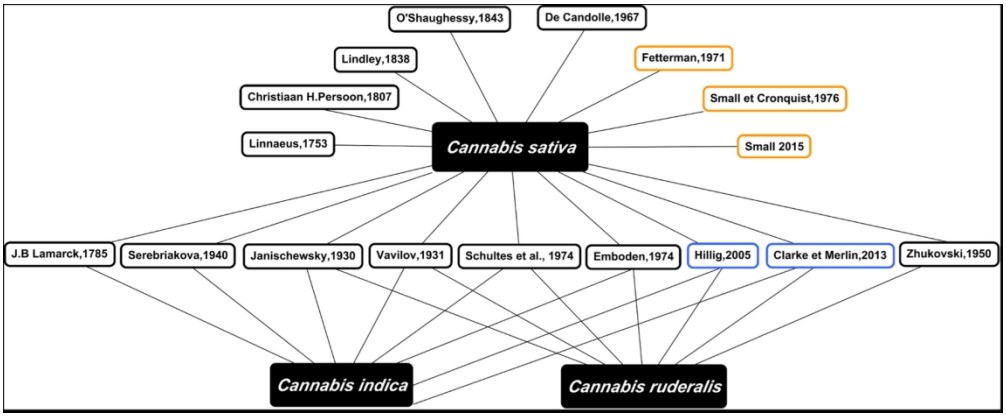
Visual representation of the phenotypic diversity in mature inflorescences of 10 Cannabis sativa L. drug-type genotype varieties (A-J), and 2 hemp-type (K-L). Varieties are as follows: A) Hyperion 7, B) Trop2683, C) Mandarin Zkittlez, D) Grape Pupil, E) Purple Majik 5, F) Funky Cherries, G) White Wedding, H) Cres 805, I) End Game R2, J) White Truffle Pupil, K) Finola, L) Felina32. A-J varieties we're grown under controlled environment in the high-performance greenhouses of Laval University by Davoud Torkamaneh lab. K and L were grown at University College Dublin and images were generously provided by C. Dowling, J. Shi and R. Melzer at the UCD Flower Power Lab.

579x806mm (63 x 63 DPI)



Three theories of classification for the Cannabaceae based on the synthesis and summary of the available articles analyzed in this global literature review. From left to right, monotypic with three subspecies (A), polytypic consisting of up to three species (B) and single phenotypically diverse species (C).

1002x721mm (38 x 38 DPI)



The three primary taxonomic classifications of the genus Cannabis: C. sativa, C. indica and C. ruderalis proposed since the study of the plant began and the authors that have endorsed them. Black text boxes distinguish authors whose classification it was carried out using a morphological/anatomic approach, the orange outlined text boxes distinguish authors who used a chemical approach and in the blue boxes the authors used a molecular approach to classify the Cannabis genus.

152x62mm (300 x 300 DPI)