Annual Review of Plant Biology

The Genomics of *Cannabis* and Its Close Relatives

I. Kovalchuk,[1] M. Pellino,[2] P. Rigault,[3,4] R. van Velzen,[5,6] J. Ebersbach,[7] J. R. Ashnest,[2] M. Mau,[2] M. E. Schranz,[5] J. Alcorn,[8] R. B. Laprairie,[8,9] J. K. McKay,[10] C. Burbridge,[11] D. Schneider,[11] D. Vergara,[12] N. C. Kane,[12] and T. F. Sharbel[2]

[1]Department of Biology, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

[2]College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, Saskatchewan S7N 4J8, Canada; email: tim.sharbel@usask.ca

[3]Gydle Inc., Québec, Québec G1S 1E7, Canada

[4]Center for Organismal Studies (COS), University of Heidelberg, 69120 Heidelberg, Germany

[5]Biosystematics Group, Wageningen University, 6703 BD Wageningen, The Netherlands

[6]Bedrocan International, 9640 CA Veendam, The Netherlands

[7]Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, Saskatchewan S7N 0X2, Canada

[8]College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5E5, Canada

[9]Department of Pharmacology, College of Medicine, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

[10]College of Agricultural Sciences, Colorado State University, Fort Collins, Colorado 80523, USA

[11]School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5E5, Canada

[12]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA

Annu. Rev. Plant Biol. 2020. 71:713–39

First published as a Review in Advance on March 10, 2020

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-081519-040203

Keywords

Cannabaceae, *Cannabis sativa* L., genomics, hemp, hops, *Humulus lupulus*, Y chromosome, biosynthesis pathway evolution, proteomics

Abstract

Cannabis sativa L. is an important yet controversial plant with a long history of recreational, medicinal, industrial, and agricultural use, and together with its sister genus *Humulus*, it represents a group of plants with a myriad of academic, agricultural, pharmaceutical, industrial, and social interests. We have performed a meta-analysis of pooled published genomics data, and

we present a comprehensive literature review on the evolutionary history of *Cannabis* and *Humulus*, including medicinal and industrial applications. We demonstrate that current *Cannabis* genome assemblies are incomplete, with ~10% missing, 10–25% unmapped, and 45S and 5S ribosomal DNA clusters as well as centromeres/satellite sequences not represented. These assemblies are also ordered at a low resolution, and their consensus quality clouds the accurate annotation of complete, partial, and pseudogenized gene copies. Considering the importance of genomics in the development of any crop, this analysis underlines the need for a coordinated effort to quantify the genetic and biochemical diversity of this species.

Contents

1. INTRODUCTION

Drug-type *Cannabis*: *Cannabis* varieties characterized by high levels of THCA, also referred to as marijuana; under long-standing prohibition in many countries

Cannabis sativa L. (henceforth *Cannabis*) is an important yet controversial plant with a long history of human use (90). Recreational, medicinal, and ritual uses rely on the psychoactive chemical constituents present in the plant (i.e., drug-type *Cannabis*). Industrial and agricultural use involves the production of fibers and, to a lesser extent, seeds (i.e., fiber-type *Cannabis*, hemp). The United Nations Office on Drugs and Crime (UNODC) (118) estimates that *Cannabis* is consumed by 2.5% of the global population, making it the most highly consumed drug. Recent debate has grown concerning the ethical and societal implications of *Cannabis* legalization for recreational and medicinal use globally. Some countries have allowed limited hemp cultivation for agriculture and industry, but *Cannabis* import and export have been greatly restricted. Recently, the prohibition of *Cannabis* cultivation has ended in some parts of the world; arguably, the boldest regulatory change has occurred in Canada, where federal law now allows large-scale agricultural production of both drug- and hemp-type *Cannabis*.

Legalization, and the concomitant ability for scientists to work with *Cannabis*, underlines the growing need for a concerted effort to, for example, characterize core germplasm collections as is being done for virtually all crops (132). As an agricultural and medicinal plant that has, for the most part, been illegally cultivated and traded over the last 80 years, the shift to crop status requires the quantification of this species' genetic and phenotypic variability as a prerequisite for successful agronomic, breeding, and medicinal applications.

With optimism for scientific and economic development due to the resurgence of this crop, researchers have recently published a number of genomics data sets using a variety of cultivars and methods (38, 56, 77, 98, 120, 124) (**Figure 1**). Here, we pool published genomics data for a meta-analysis and present a comprehensive literature review. We frame this information in the context of the evolutionary history of this species and its closest relatives.

Hemp-type *Cannabis*: *Cannabis* varieties characterized by low levels of THCA and high levels of CBDA; used for fibers, seeds, and seed oils

Dioecious: having male and female reproductive organs on different plants and thus having biparental reproduction

Dioecy: characteristic of a species in which males and females are distinct organisms

2. BIOGEOGRAPHY AND DOMESTICATION

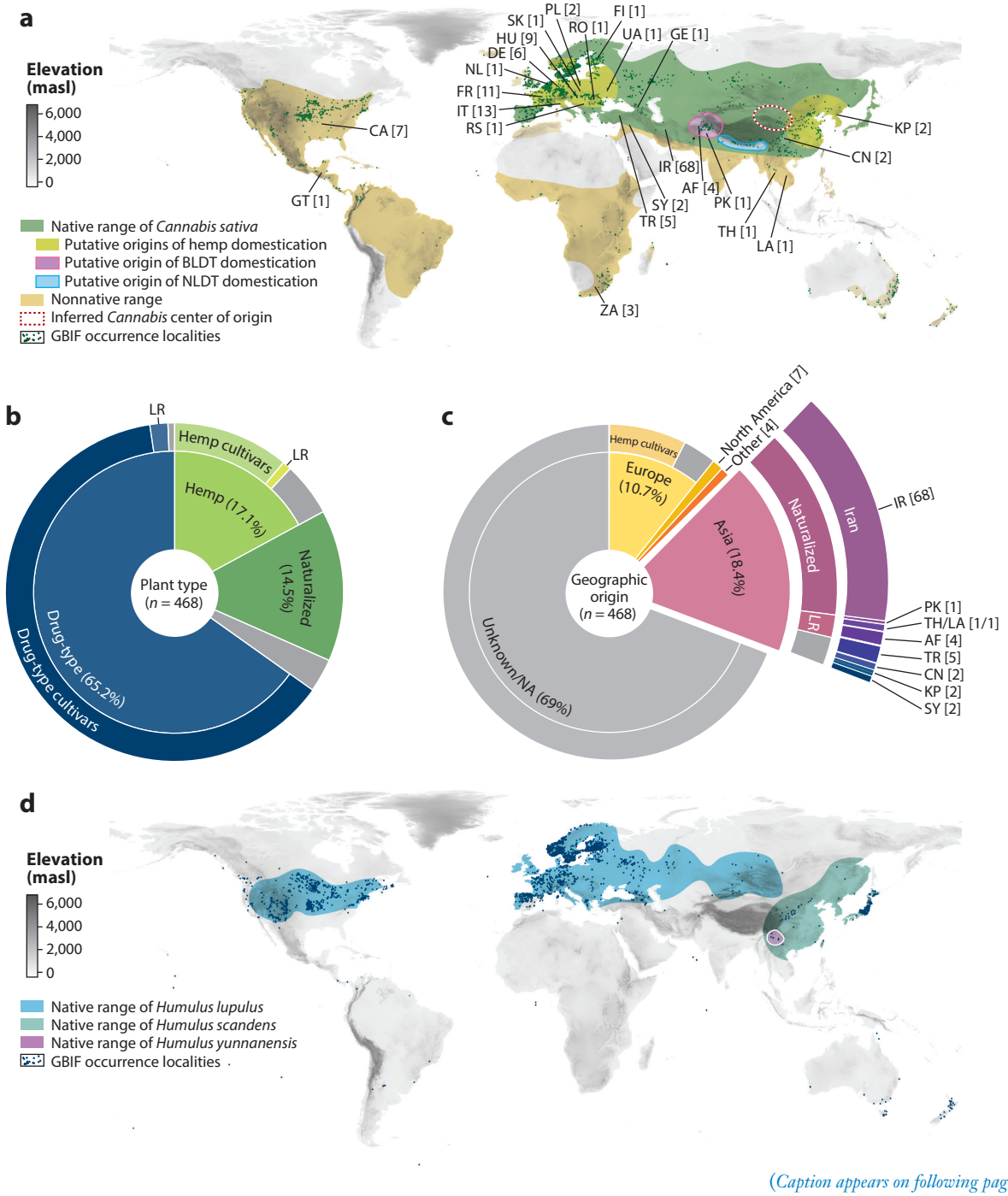
Cannabis comprises a single species of dioecious herb, *Cannabis sativa* L. It is sister to the genus *Humulus* L., which comprises three species of dioecious twining herbs commonly known as hop: the rare *H. yunnanensis* Hu, *H. scandens* (Lour.) (synonym *Humulus japonicus*) Merr., and *H. lupulus* L. with five varieties (10, 23, 104, 105). Together, *Cannabis* and *Humulus* have long been recognized to compose the family Cannabaceae. However, based on recent phylogenetic studies, Cannabaceae is currently considered to include eight additional genera of mostly tropical trees (9, 113) (**Figure 2**).

Based on our current understanding of Cannabaceae phylogenetics, herbal growth and dioecy transitioned from tree growth and monoecy, respectively, in a common ancestor of *Cannabis* and *Humulus*. Subsequently, cannabinoid biosynthesis evolved in an ancestor of *Cannabis* while biosynthesis of bitter acids and xanthohumols evolved in an ancestor of *Humulus*, although the molecular changes underlying the origin and evolution of these pathways are not well known (**Figure 2**).

The majority of Cannabaceae genera are considered to have their origins in East Asia, but the precise areas of origin are often unknown (69). Both *Trema* and *Humulus* have their centers of diversity in China, with a focus on the biodiversity hotspot of the mountains of Southwest China (Hengduan Mountains) in *Humulus* (**Figure 1**; also see the sidebar titled Global *Cannabis* Sampling).

Similarly, C*annabis* is considered to have originated in temperate Asia (reviewed in 23, 43) after having diverged from its common ancestor with *Humulus* (104), likely in the late Oligocene to early Miocene [27.8 million years ago (mya) (68) to 21 mya (142)]. Subfossil pollen analysis suggests a putative center of origin for *Cannabis* on the northeastern fringe of the Qinghai-Tibet Plateau, over 19.6 mya (69). From there, the data suggest that the genus dispersed to Russia and Europe approximately 6 mya and to eastern China approximately 1.2 mya, all likely before the onset of human evolution (68, 69).

Based on achene fossils recovered in East Asia and Europe, early human use of *Cannabis* as a fiber plant is hypothesized to date back to at least 8,000 years before present (BP) in the Neolithic, having become widespread by at least 4,500 years BP (59, 68, 69). Hemp-type *Cannabis* was introduced in the New World as an industrial crop in the sixteenth century. For drug-type *Cannabis*, ancient use dates back to 2,700 years BP, documented through burial gifts containing Δ^9-tetrahydrocannabinol (THC) found in Central Asia, which were likely used in medicinal and/or ritualistic contexts (49, 50, 90, 94). The sites of these fossils did not reveal any evidence of fiber use (94), raising the question of whether *Cannabis* was first used as a multipurpose crop, or whether the domestication of hemp and drug-type *Cannabis* proceeded independently (22, 94). Regardless, cultivated drug-type *Cannabis* was spread from Central Asia and/or India to Africa, the Middle East, and Southeast Asia by Arab and Hindu cultures between 2,000 and 500 years BP (27). In East Africa, *Cannabis* was introduced to cultures that had invented pipe smoking (27).



(Caption appears on following page)

Figure 1 (*Figure appears on preceding page*)

Global distribution ranges and current genotype sampling of *Cannabis* and *Humulus*. (*a*) Putative native and nonnative distribution ranges of *Cannabis sativa* as documented in the literature (23, 106, 107) as well as region of origin as inferred by McPartland (68). Green dots indicate preserved specimen sampling localities stored in the GBIF [GBIF.org (27 September 2019) GBIF Occurrence Download; **https://doi.org/10.15468/dl.qqwywp**]. Lines and abbreviations indicate countries represented in recent genotyping studies (62, 98, 109) with the number of samples given in brackets. (*b*) Number of genotyped *Cannabis* samples achieved in three landmark genotyping-by-sequencing studies (62, 98, 109) by plant type (hemp, drug-type, and naturalized/wild *Cannabis*) and corresponding fractions of cultivars, landraces, and commercial strains. Gray sections indicate accessions of unknown or unspecified plant type. (*c*) Number of genotyped *Cannabis* samples by continent and country of origin and corresponding fractions of naturalized/wild accessions, hemp cultivars, and drug-type LR. Gray sections indicate accessions of unknown or unspecified geographic origin. Lines and abbreviations indicate countries, with the number of samples given in brackets. Other represents South Africa (three samples), Guatemala (one sample). (*d*) Native distribution ranges of *Humulus lupulus*, *Humulus scandens*, and *Humulus yunnanensis* as documented in the literature (104; **http://efloras.org** and **http://www.agroatlas.ru/**). Blue dots indicate preserved specimen sampling localities stored in the GBIF [GBIF.org (13 October 2019) GBIF Occurrence Download; **https://doi.org/10.15468/dl.4ddht**]. Abbreviations: AF, Afghanistan; BLDT, broad leaflet drug-type *Cannabis*; CA, Canada; CN, China; DE, Germany; FI, Finland; FR, France; GBIF, Global Biodiversity Information Facility; GE, Georgia; GT, Guatemala; HU, Hungary; IR, Iran; IT, Italy; KP, Korea; LA, Laos; LR, landrace; masl, meters above sea level; NA, not available; NL, Netherlands; NLDT, narrow leaflet drug-type *Cannabis*; PK, Pakistan; PL, Poland; RO, Romania; RS, Serbia; SK, Slovakia, SY, Syria; TH, Thailand; TR, Turkey; UA, Ukraine; ZA, South Africa.

Finally, drug-type *Cannabis* became established throughout the Americas after 1800 AD (23). In the postwar twentieth century, targeted breeding and extensive hybridization within and between hemp- and drug-type *Cannabis* marked their development (23), leading to numerous hybrid cultivars that remain to be disentangled.

GLOBAL *CANNABIS* SAMPLING

A comprehensive understanding of *Cannabis* and *Humulus* phylogeographic and domestication history, genome evolution, and genetic and phenotypic diversity is important for future breeding and conservation endeavors. However, there remain several key gaps in our evolutionary knowledge of these species, in particular with regard to *Cannabis*. This is largely due to its long-standing prohibition and the lack of curated, georeferenced germplasm collections, especially from the putative centers of origin and introduction.

For example, the vast majority (~61%) of accessions genotyped by the most recent diversity studies in *Cannabis* are commercial drug strains for which no passport data are available, while hemp and naturalized accessions have mainly been sampled from Europe and the Middle East, respectively. Assembly of a global data set comprising various types of *Cannabis* materials (e.g., undomesticated, naturalized, and landraces) as well as a comprehensive sampling of all putative infrageneric subgroups should be prioritized in future evolutionary studies, agricultural improvements, and medicinal applications.

Specifically, focus needs to be placed on Asian accessions, as we could only trace 3% (15) of the 470 accessions (landraces and naturalized populations) included in the three most recent genotyping studies to this region (62, 98, 109). Considering the remarkable genetic diversity harbored by regions of Asia (45, 143) and its hypothesized importance for the origin and domestication of the genus (23, 69, 105), this needs to be addressed urgently. This could be achieved not only by making use of large germplasm bank collections (73) but also by sourcing from vast herbarium collections. For example, the Global Biodiversity Information Facility (**https://www.gbif.org/**) lists 553 preserved specimens from China, 262 from Pakistan, and 331 from the Russian Federation that are deposited in herbaria worldwide. Even though these collections may not be directly useful for breeding purposes, sampling landraces and wild/naturalized lines from herbaria might help to elucidate patterns of genetic diversity and evolutionary history, which can in turn help to identify important regions that need to be targeted for germplasm sampling for breeding and/or evolutionary studies.

Young shoots of *Humulus lupulus* were collected from supposedly wild hops (cf. *lupum salictarium*, willow wolf) and cooked, as the Roman natural philosopher Pliny the Elder (23–79 AD) mentioned, possibly for the first time, in his *Naturalis Historia*. The collection of wild hops (female flowers and fruit clusters) (25) and the first cultivation of the plants in Bavaria in 859 AD were conducted mainly for the purpose of beer-brewing (8), an industry that has seen a recent surge due to the revival of craft beer production (136). Besides being a beer-flavoring agent, *Humulus lupulus* has a wide variety of uses due to its digestive, antibacterial, and antifungal effects, and it may have anticarcinogenic potential (33, 140). *Humulus scandens* is not used for brewing beer but is used in traditional Chinese medicine to treat pneumonia, diarrhea, hypertension, leprosy, and tuberculosis (112). Medicinal uses of *Humulus yunnanensis* are unknown.

Within Cannabaceae, *Cannabis* and *Humulus* are members of a larger clade that comprises plants with a derived base chromosome number $x = 10$ (113, 137) (**Figure 2**).

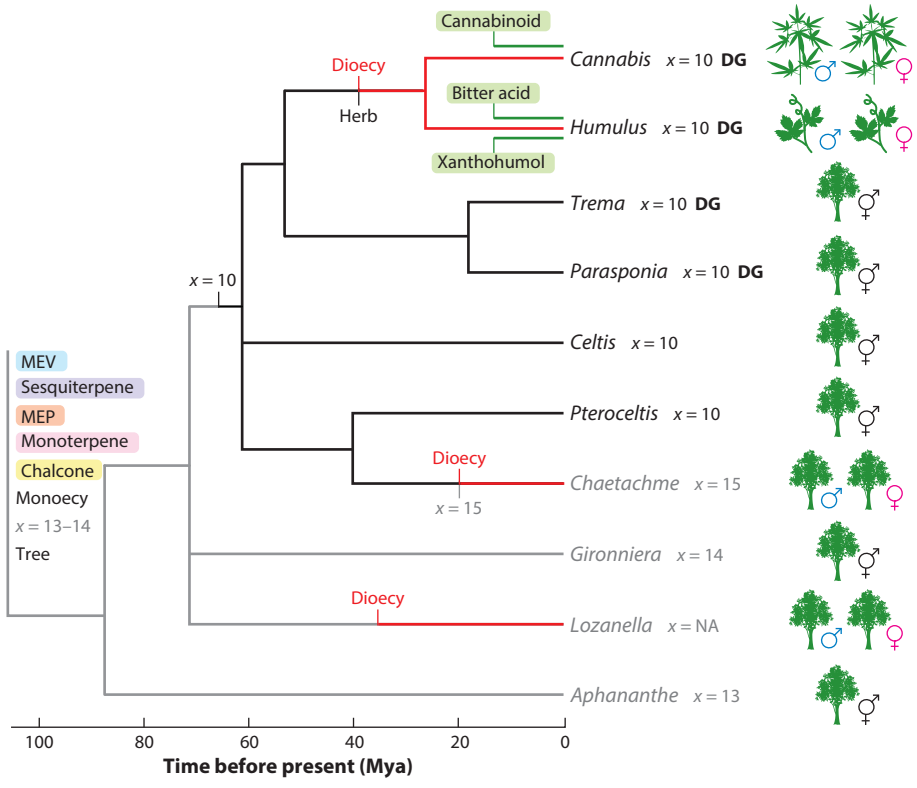


Figure 2

Phylogenetic tree of Cannabaceae family. Time-calibrated genus-level divergences based on five plastid and three nuclear gene regions (51) showing evolutionary transitions of key traits (137) and putative origins of biosynthetic pathways (*colored boxes*). All nodes shown have a posterior probability of 1 based on plastome phylogenetics (144). Estimated ancestral states and pathways are given at the root of the tree. Monoecious plants are indicated by a bisexual sign, dioecious plants are indicated by separate male and female signs, and corresponding branches are in red. Branches representing early-diverging genera with $x = 13$ or 14 are in gray, and those representing genera with $x = 10$ are in black. Genera for which draft genome assemblies are available are marked with DG. We note the genus *Parasponia* is nested within a paraphyletic *Trema* (123, 137, 144). Abbreviations: DG, draft genome; MEP, methylerythritol phosphate pathway; MEV, mevalonate pathway; Mya, million years ago; NA, not available.

Recent phylogenetic analyses based on plastid genomes suggest that the genera *Parasponia* and *Trema* together are sisters to *Cannabis* and *Humulus* (144). *Parasponia* species are of scientific interest due to their ability to form root nodule symbioses with nitrogen-fixing rhizobia, a trait that is only shared with legumes (123). Phylogenetically, *Parasponia* is nested within *Trema*, which otherwise comprises species that do not make root nodules (122). Currently, contig-level genome assemblies are available for *Humulus*, *Parasponia*, and *Trema*, which can be used as outgroups in comparative analyses of *Cannabis* (38, 56, 123).

Outgroups: closely related taxa that are used to establish ancestral states of traits (genetic, phenotypic) in the taxonomic group of interest (e.g., *Cannabis*)

3. THE *CANNABIS* GENOME

Cannabis has a diploid nuclear genome ($2n = 20$) composed of nine autosomal chromosomes and a pair of sex chromosomes. The haploid nuclear genome size is estimated to be 818 mega base pairs (Mbp) for females (karyotype XX) and 843 Mbp for males (karyotype XY) (95). The *Cannabis* plastid and mitochondrial genomes are 153,871 bp (126) and 415,545 bp (133), respectively.

Initial *Cannabis* genome assemblies (120) used Illumina shotgun and mate-pair sequencing in drug-type cultivar Purple Kush (PK) to produce a genome survey of 786 Mbp in 136,290 unmapped scaffolds, with 534 Mbp called in 363,760 contigs and 252 Mbp uncalled (N) in 228,430 gaps. A gene catalog of PK was built using Illumina RNA-sequencing (RNA-Seq) sequences assembled in 30,074 representative transcripts (33 Mbp). A companion genome survey of hemp-type cultivar Finola (FN; 221 Mbp in 224,195 scaffolds) showed genome-wide single nucleotide polymorphism (SNP) differences between these representative drug-type *Cannabis* and hemp-type genomes but relatively few gene copy number changes.

Recent assemblies of PK, FN (56), and a high–cannabidiolic acid (CBDA) cultivar (CBDRx) (38) used long-read sequencing and genetic mapping to produce draft chromosome pseudomolecule sequences (PK 639 Mbp in 6,193 scaffolds, 739 Mbp called, 0.6 Mbp N; FN 784 Mbp in 2,951 scaffolds, 784 Mbp called, 0.3 Mbp N; CBDRx 854 Mbp in 773 scaffolds, 714 Mbp called, 140 Mbp N) complemented by unmapped scaffolds (PK 252 Mbp in 6,643 scaffolds, FN 224 Mbp in 2,352 scaffolds, and CBDRx 21 Mbp in 220 scaffolds). Additional genome survey (unmapped scaffolds) assemblies are available for five other cultivars (**Table 1**).

Additional scaffold-level assemblies (not in peer-reviewed journals) of several *Cannabis* cultivars are publicly available, and their statistics are summarized in **Table 1**. Comparison of these

Table 1 Overview of the statistics of available *Cannabis* genome assemblies

Cultivar	BioSample	BioProject	Size (Mb)	NCBI assembly	Whole genome sequencing	Scaffolds
Finola	SAMN09375800	PRJNA73819	1009.67	GCA_003417725.2	QKVJ02	5,303
Purple Kush	SAMN02981385	PRJNA73819	891.965	GCA_000230575.4	AGQN03	12,836
CBDRx	SAMEA5040675	PRJEB29284	875.722	GCA_900626175.1	UZAU01	220
LA Confidential	SAMN04145444	PRJNA297710	595.358	GCA_001510005.1	LKUA01	311,039
Cannatonic	SAMN05941636	PRJNA350523	585.824	GCA_001865755.1	MNPR01	11,110
Pineapple Banana Bubba Kush	SAMN06546749	PRJNA378470	512.174	GCA_002090435.1	MXBD01	18,355
Jamaican Lion DASH	SAMN09851581	PRJNA486541	1333.38	GCA_003660325.2	QVPT02	3,372
Chemdog91	SAMN04145446	PRJNA297710	285.933	GCA_001509995.1	LKUB01	175,088

The table includes publicly available *Cannabis* genome assemblies representing eight cultivars showing the total assembly size and number of scaffolds. Finola, Purple Kush, and CBDRx are draft assemblies with chromosome pseudomolecule sequences; other assemblies are genome surveys consisting of unmapped scaffolds.

genomes shows large-scale differences in assembled genome size and order (**Table 2**). However, the degree to which this represents genomic differences versus misassembly artifacts remains unclear.

Genome assembly: the computational representation of an organism's genome; it is produced by sequencing fragmented chromosomes and informatically re-assembling them

4. THE *HUMULUS* GENOME

For *Humulus*, several draft genomes are published and available in a single resource (**http://hopbase.org**) (42). Compared to *Cannabis* [*C. sativa*: 1.67pg/2C (95)], diploid genome size varies between 3.21pg/2C for female *H. japonicus* and 6.10pg/2C for certain *H. lupulus* varieties, allocated across nine autosomal bivalents in addition to the sex chromosomes (**https://cvalues.science.kew.org**) (37, 75, 147). The first long-read sequencing run with R6 and R7 MinION Flow Cells on *H. lupulus* var. *lupuloides* resulted in 6.3 and 3.6 Mbp with an average two-dimensional (2D) read length of 1,652 bp and 1,628 bp, respectively (12), but which only aligned to 10% of the reference *C. sativa* (56, 120). These data are superseded by the latest PacBio FALCON-Unzip diploid genome assembly of an *H. lupulus* var. *lupulus* cultivar (cv.) cascade, which covers 3.79 Gbp of the estimated 5.4 Gbp with an average contig NG50 of 866 kbp (**http://hopbase.cgrb.oregonstate.edu**) (42). In addition, this result is superior to the latest Teamaker short-read assembly of 1.77 Gbp with NG50 contigs of 1.4 kbp due to the large, highly repetitive genome size (34.68%), which favors long-read sequencing approaches (Teamaker v1.1; **http://hopbase.cgrb.oregonstate.edu**) (42). In contrast, *C. sativa* exhibits an even higher degree of genome repetition (~70%) (74, 86). Overall, the number of genes recovered in hop genome assemblies is highly variable, for example, 35,482 in *H. lupulus* cv. SW and 24,919 in *H. lupulus* cv. Teamaker, compared to 27,819–34,589 genes in *Cannabis* (**Table 2**). Although the CBDRx genome assembly now includes gene annotations (47) and has been selected as a representative genome by the National Center for Biotechnology Information, detailed gene characterization and localization are still in their infancy for both *Cannabis* and *Humulus* (42, 74, 120).

5. CHROMOSOME-LEVEL COMPARISONS OF PUBLISHED DATA

To assess the completeness and representativity of the PK, FN, and CBDRx assemblies, we used Illumina sequences from PK, FN, and 55 public whole-genome-sequenced (WGS) samples. Excluding contaminant, plastid, and mitochondrial fragments and filtering reads for high quality, we mapped 91%, 89%, and 82% (±5% among individuals) of the fragments to the PK, FN, and CBDRx assemblies and 60%, 67%, and 80% to their chromosomes, respectively.

Chromosomes are numbered and oriented differently in the PK, FN, and CBDRx assemblies. We used the representative transcripts from the PK gene catalog to locate 25,467, 28,987, and 27,811 gene sequences in the PK, FN, and CBDRx chromosomes (27,819, 34,589, and 30,163 in total assemblies). Comparing the chromosome positions of uniquely mapped genes, we identified and oriented equivalent chromosomes between the assemblies (**Table 2**). Between assemblies, 90% of genes mapped in equivalent chromosome quartiles; however, the conservation dropped significantly for equivalent segments below 10% of the chromosome length. The chromosome assemblies are ordered with a genetic map containing a large fraction of recombination-free regions (52% in FN), and the local physical order of scaffolds is largely undetermined. The chromosome assemblies show that both genes and recombination events are concentrated near chromosome ends (**Figure 3**); in this respect, *Cannabis* behaves much more like a grain than like a member of *Rosales* (56).

Mapping gene sequences to the PK and FN chromosome assemblies, we noticed a high number of mismatches in alignments to the same cultivar [PK transcripts showed <97% similarity with the PK chromosome sequences; none of 11 FN GenBank coding (CDS) sequences mapped

Table 2 Chromosome correspondence and statistics among assemblies

CHR-ID		Orientation	Size (Mb)	Scaffolds	Genes
	FIN-ID				
1	1	+	100,649,945	368	2,893
2	2	+	95,692,043	357	2,894
3	3	+	94,587,701	382	3,346
4	4	+	92,111,078	340	2,572
5	5	+	87,047,438	344	5,117
6	6	+	77,135,887	288	2,218
7	7	+	76,634,836	290	2,915
8	8	+	76,024,397	284	2,519
9	9	+	49,536,295	170	2,576
10	10	+	35,338,263	128	1,937
CHR	CHR		784,757,883	2,951	28,987
UN	UN		224,916,856	2,352	5,602
Total	Total		1,009,674,739	5,303	34,589
	PK-ID				
1	1	+	79,255,070	818	2,472
2	2	–	73,430,137	803	2,300
3	3	–	72,195,804	755	2,773
4	4	–	78,152,476	732	2,534
5	6	–	72,519,175	670	4,304
6	7	–	60,968,100	609	1,986
7	9	–	60,626,883	488	2,506
8	5	+	62,039,859	625	2,265
9	8	–	51,081,244	444	2,579
10	10	+	29,404,172	249	1,748
CHR	CHR		639,672,920	6,193	25,467
UN	UN		252,291,743	6,643	2,352
Total	Total		891,964,663	12,836	27,819
	CBDRx-ID				
1	5	+	88,181,582	133	2,215
2	8	+	79,335,105	40	2,365
3	6	+	96,346,938	67	2,849
4	3	–	94,670,641	105	2,516
5	2	+	101,209,240	144	4,414
6	9	–	71,238,074	51	1,898
7	4	+	91,913,879	49	2,923
8	7	+	61,561,104	45	2,238
9	10	–	64,622,176	60	2,862
10	1	–	104,987,320	79	3,531
CHR	CHR		854,066,059	773	27,811
UN	UN		21,666,009	210	2,352
Total	Total		875,732,068	983	30,163

Rows contain equivalent chromosomes in Finola (FIN), Purple Kush (PK), and CBDRx [1–10 individual chromosomes, total sequence in chromosomes (CHR), and total sequence in unmapped scaffolds (UN)]. Each assembly shows the chromosome ID, orientation, total size, called size (without gaps), number of scaffolds, and number of genes.

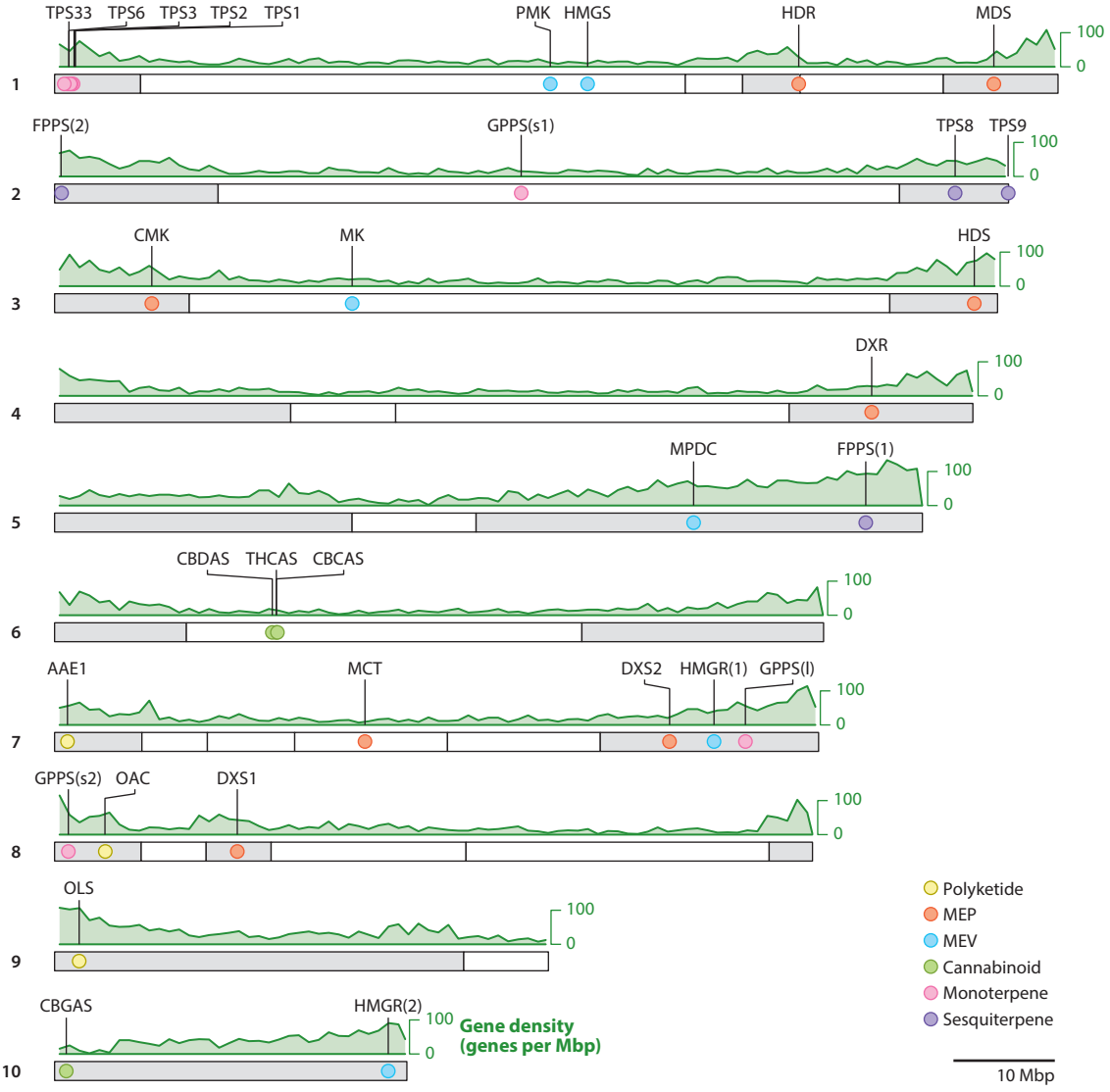


Figure 3

Cannabis chromosomes showing gene and recombination density and position of selected biosynthesis genes. Chromosomes with numbering, sizes, and gene positions corresponding to Finola assembly v2 are represented with rectangular boxes showing regions with the presence (*gray*) or absence (*white*) of recombination (a horizontal 10-Mbp scale is shown). Recombination intervals in Finola were extracted from supplemental data of Laverty et al. (56). Gene density (number of genes per mega base pair) is plotted in green above the chromosomes (a vertical scale of 0–100 is shown). The position of the genes of six selected biosynthetic classes (**Supplemental Table 1**) is indicated by circles, the colors of which correspond to the biochemical pathways shown in **Figures 4** and **5**. Abbreviations: AAE1, hexanoyl-CoA synthetase 1; CBCAS, cannabichromenic acid synthase; CBDAS, cannabidiolic acid synthase; CBGAS, cannabigerolic acid synthase; CMK, CDP-ME kinase; DXR, deoxyxylulose phosphate reductoisomerase; DXS, deoxyxylulose phosphate synthase; FPPS, farnesyl pyrophosphate synthase; GPPS, geranyl-pyrophosphate synthase; HDR, hydroxymethylbutenyl diphosphate reductase; HDS, hydroxymethylbutenyl diphosphate synthase; HMGR, hydroxymethylglutaryl-CoA reductase; HMGS, hydroxymethylglutaryl-CoA synthase; MCT, methylerythritol phosphate cytidylyltransferase; MDS, MECDP-synthase; MEP, methylerythritol phosphate pathway; MEV, mevalonate pathway; MK, mevalonate kinase; MPDC, mevalonate diphosphate decarboxylase; OAC, olivetolic acid cyclase; OLS, olivetol synthase; PMK, phosphomevalonate kinase; THCAS, tetrahydrocannabinolic acid synthase; TPS, terpene synthase.

Supplemental Material >

perfectly to the FN chromosomes] and frequent stop codons in the corresponding chromosome CDS sequences. This suggests that the poor accuracy of long-read sequencing negatively impacts the assembly sequence and prevents accurate genome annotation in these assemblies.

The currently available *Cannabis* genome assemblies are still incomplete, with ~10% missing, 10–25% unmapped, and 45S and 5S rDNA clusters as well as centromeres/satellite sequences not represented. These assemblies are also ordered at a low resolution, and their consensus quality makes accurate annotation and accounting of complete, partial, and pseudogenized gene copies challenging. Elucidating quantitative traits and biosynthesis pathway differences between cultivars requires a higher-quality *Cannabis* reference genome with sequences that are physically ordered (for example, using optical or single-molecule technologies) at high resolution and locally finished, and the resolution of allelic sequences and tandem repeats (53).

Heterogamety: presence of different gametes that lead to sex differentiation in males (XY; e.g., humans) and females (ZW; e.g., some birds, fish and insects)

6. DIOECY AND SEX DETERMINATION

Cannabis is a dioecious plant displaying occasional hermaphroditism, with cases of monoecy in hemp (70). Dioecy is the precondition for the evolution of sex chromosomes (17–19), estimated to have evolved from a monoecy in a common ancestor of *Cannabis* and *Humulus* (see **Figure 2**). Contrary to most other dioecious plants, where sex allocation is environmentally modulated (91), *Cannabis* and *Humulus* have sex chromosomes, with males representing the heterogametic sex (XY system). It is yet unknown if sex allocation in the other dioecious Cannabaceae genera, *Chaetachme* and *Lozanella*, is determined environmentally or via sex chromosomes. *Cannabis* has been described as having either an X chromosome to autosome balance (28) or a dominant Y chromosome system (95). This confusion regarding the role of the Y chromosome has been attributed to the influence of abiotic stress in sexual determination (124). In *Humulus*, sex determination is dependent on the X/autosome balance, and various sex chromosome systems exist, such as Winge, new Winge, Sinoto, and new Sinoto (84). *H. lupulus* ($2n = 20 + XY$), *H. japonicus* ($2n = 14 + XY1\ Y2$), and *H. lupulus* var. *cordifolius* are typical karyotypes ($2n = 16 + X1Y1\ X2Y2$) (75, 79, 134). Due to the close relationship between *Humulus* and *Cannabis*, it is unclear whether sex chromosome diversity represents single or multiple evolutionary origins.

7. Y CHROMOSOME EVOLUTION IN *CANNABIS*

In order for a sex chromosome to evolve, dioecy must be established through the incidence of two closely linked and complementary dominant mutations that disrupt male and female organ development (70). Once this seed occurs on any chromosome, a cascade of events shapes its evolution toward a sex chromosome. Increased linkage of sex-specific genes gives rise to an expanding, nonrecombining region, resulting in the accumulation of different mutations and rearrangements (1, 18, 19). Both male (XY) and female (ZW) heterogamety exist, though the XY system is predominant due to intrinsic evolutionary advantages (13); in the XY system, male sterility in XX is due to a recessive loss of function, while female sterility in XY is induced dominantly (1, 19).

The *Cannabis* Y chromosome was found to be 47 Mbp larger than the X chromosome, with differentiation in the subtelocentric region of its long arm (26, 95). This size difference was attributed to evidence of long interspersed nuclear elements (LINE)-like retrotransposon accumulation on the terminal region of the Y chromosome (95). These retrotransposons were not detected on either the X chromosome or the autosomes, although this could reflect the sensitivity of the technique [fluorescence in situ hybridization (FISH)], which is unable to detect low copy numbers (96). Male-specific fragments were also detected along the same region, although the extent of variability in these sequences and their copy numbers remains unclear (97).

Ligands: substances that form complexes with other biomolecules (e.g., proteins and DNA), leading to specific biological activities

The search for male-specific genetic markers for breeding purposes and other applications has led to the identification of several loci on the Y chromosome (26, 64). However, several of these markers are only partially reliable (64, 71, 85, 116). To further complicate things, a pseudoautosomal region allows for ongoing recombination between X and Y chromosomes (95).

These observations have led to the assumption that the Y chromosome in *Cannabis* is at an early evolutionary stage (124), wherein male and female sterility have been recently established and the nonrecombining region of sex-linked genes has not wholly formed (70). Recently, Prentout et al. (87) provided an alternative scenario based on RNA-Seq data, followed by mutation accumulation and segregation analyses, and concluded that the *Cannabis* Y chromosome is highly degenerate. Nevertheless, a more comprehensive genomic and transcriptomic overview of the *Cannabis* Y chromosome is required. Given the importance of sex determination for breeding and for our understanding of dioecy and XY chromosomes in plants, a comparative study of the Y chromosome in *Cannabis* and *Humulus* presents a unique opportunity to understand sex chromosome evolution.

8. PHYTOCANNABINOID BIOCHEMISTRY AND EFFECTS ON HUMAN PHYSIOLOGY

The primary biologically active components of *Cannabis* are the (phyto)cannabinoids, which are terpenophenolic ligands synthesized by the plant (141). The two most abundant phytocannabinoids are Δ^9-tetrahydrocannabinolic acid (THCA) and CBDA, although at least 120 other cannabinoid compounds have been identified in *Cannabis* (141). When heated, these compounds are converted into their bioactive forms, THC and cannabidiol (CBD).

In humans and other animals, cannabinoids modulate the activity of the body's endogenous cannabinoid system (ECS) (61). The ECS is composed of several receptors, including the type 1 and 2 cannabinoid receptors (CB1R and CB2R), endogenous cannabinoids [anandamide (AEA) and 2-arachidonoylglycerol (2-AG)], and the anabolic and catabolic enzymes that maintain endogenous cannabinoid tone in the body (61). The ECS participates in the regulation of multiple physiological processes, including lipid homeostasis, appetite, mood, arousal, motor control, inflammation, and pain (110).

Phytocannabinoids modulate different aspects of the ECS depending on their mechanism of action. THC is a partial agonist of CB1R and CB2R; activation of CB1R results in the intoxicating high of *Cannabis* as well as analgesic effects (41) and acute increases in appetite (46). CBD modulates the activity of many different receptors that are part of the ECS, including CB1R and CB2R as well as other receptors such as serotonin and opioid receptors (52, 66). The precise mechanism of the action of CBD in vivo remains unclear. Very little is known regarding the molecular pharmacology of the many other phytocannabinoids and how this pharmacology may influence the effects of the phytocannabinoids in humans (88). Some evidence suggests that THCA is a weak partial agonist of CB1R and CB2R, and far less abundant cannabinoids and terpenes, such as Δ^9-tetrahydrocannabivarin, cannabichromene, and β-caryophyllene, may act on CB1R, CB2R, or the putative cannabinoids receptor GPR55 (4, 34, 117). Such studies have only begun to explore the vast array of compounds present in *Cannabis*, and as data on the pharmacology of cannabinoids accumulate, it is increasingly clear that cannabinoids are promiscuous ligands with multiple receptor targets, including targets outside of the ECS. Research will thus have to consider not only nonspecific pharmacological effects but also the net effect(s) produced through unique combinations of cannabinoids present in the different cultivars of *Cannabis* and products derived from them. The present state of knowledge has very much focused on THC and CBD, as these two compounds represent the majority of cannabinoids present in products by mass.

With respect to pharmacokinetics, the phytocannabinoids appear to share common drug-metabolizing pathways in the human body (cytochromes P450 2D6, 2C9, and 3A4) (141), and their efficacy may differ between individuals, depending on individual ECS and drug-metabolizing genes (44). Much of our current understanding of cannabinoid health effects is associated with THC and CBD (6), and only few data are available for the short- and long-term physiological effects of other cannabinoids or their combined effects. Such combinations of biochemistry are likely to differ in terms of both the relative quantities and presence or absence between cultivars of *Cannabis*, thus underlying the importance of having a polished genome and core germplasm collection as solid foundations upon which pharmaceutical trials can be based.

A wealth of consumer-driven Internet information, limited human clinical trial data, and modest preclinical animal studies purport a plethora of health benefits of cannabinoids. An important caveat to the putative medical effects of *Cannabis* is their lack of substantiation with rigorous human clinical trial evidence. Most trials that are available involve *Cannabis*-derived or *Cannabis*-related drug products that underwent regulatory safety and efficacy evaluation and are now marketed as *Cannabis*-based drugs (119). Other trials have largely involved subjects who smoked *Cannabis* products that are usually high in THC (14, 121). Data from these trials are further complicated by the important concern that study outcomes for any one *Cannabis* product cannot necessarily be compared to others, considering biochemical variability between different *Cannabis* cultivars and even within cultivars that are cloned.

The limited existing data do support that *Cannabis* has a number of potential health benefits. Approved products include a purified form of CBD (i.e., Epidiolex) for seizures associated with Dravet or Lennox-Gastaut syndrome (20); dronabinol (synthetic THC) for the treatment of cachexia in HIV/AIDS patients (7); and nabiximols (Sativex, a botanical extract composed of 2.5-mg CBD and 2.7-mg THC per single 100-mL spray), used as adjunctive treatment for the symptomatic relief of neuropathic pain in multiple sclerosis in adults (35). Until now, other *Cannabis*-based medicines have not yet entered regulatory testing for safety and efficacy, as would be expected with other commercial pharmaceutical agents. However, observational studies, case reports, and anecdotal evidence have been accumulating, and meta-analyses suggest that *Cannabis* has potential therapeutic applications in areas such as pain, spasticity in multiple sclerosis, and opioid sparing (**https://www.cannabis-med.org/studies/study.php**).

Despite the promise of therapeutic benefits in a variety of chronic disease conditions, we need to remain mindful of the possible adverse effects of the cannabinoids. The negative short-term effects of THC have long been understood and include reduced cognitive function, enhanced anxiety, fatigue, and negative cardiovascular effects (128). Long-term negative consequences of chronic high THC product use are also associated with permanent loss in higher cognitive functions such as memory, intelligence, mental focus, and judgment (55). Addiction is also possible, particularly for those individuals with predisposing risk factors, such as a family history for mental illness (60). When smoked, long-term use can result in lung problems, such as chronic bronchitis and cough (92). Moving forward, it is critical to recognize *Cannabis* and cannabinoids as drugs with potential benefits and associated risks, as would be the case for the investigation of any novel drug.

9. CANNABINOID SYNTHESIS

THC accumulates mainly in capitate-stalked glandular trichomes of *Cannabis*. Glandular trichomes have two major parts: the stem and the gland. Glands consist of disk cells with large cavities; these cavities contain cannabinoids and secondary metabolites (103). In addition, the trichome heads also contain RNAs encoding three possible polyketide synthases, members of the methylerythritol 4-phosphate (MEP) pathway, and THCA synthase (THCAS), suggesting that

the storage cavity is a major site of cannabinoid production (65). It is still not understood why THCA is secreted into the cavity or why cannabinoid biosynthesis occurs in the cavity. It is hypothesized that the end products or by-products, including hydrogen peroxide, may be toxic for plant cells (101). Interestingly, antimicrobial properties of cannabinoids and their presence in trichomes imply an association with plant defense (80).

Cannabinoids are synthesized and stored as acids; however, exposure to light or heat during storage, processing, and consumption leads to nonenzymatic decarboxylation to their neutral forms, which have psychoactive properties (54). Most of the biosynthetic pathway of synthesis of cannabinoids has been elucidated, and genes and enzymes for most steps of the pathway have been isolated and characterized (16) (**Figure 4**).

9.1. Precursor Synthesis

Synthesis begins from two distinct metabolic pathways: the polyketide pathway, leading to the formation of olivetolic acid (OA), and the MEP pathway, producing geranyl diphosphate (GPP) (102) (**Figure 4**). The synthesis of OA via the polyketide pathway is considered specific to *Cannabis* and starts with hexanoic acid, which is hypothesized to be synthesized from acetyl-coenzyme A (CoA) and five molecules of malonyl-CoA, although its origin in the trichomes has not been shown (29). Hexanoic acid is converted into hexanoyl-CoA by an acyl-activating enzyme (AAE) called hexanoyl-CoA synthetase (HCS) (65), of which two forms have been identified in *Cannabis*: CsHCS1 and CsHCS2 (81, 111). Whereas CsHCS1 [720 amino acids (AA)] is believed to be trichome-specific, CsHCS2 (543 AA) is found to be expressed in all tissues. CsAAE1 (CsHCS1) is likely the enzyme involved in the synthesis of cannabinoids since it is the most abundant AAE, specific for short-chain fatty acyl-CoA, and localized to the cytosol, similar to olivetol synthase (81, 111).

OA is synthesized by a polyketide synthase (PKS) named olivetol synthase (OLS), which catalyzes the aldol condensation of hexanoyl-CoA with three molecules of malonyl-CoA (89, 115). The malonyl-CoA is derived from acetyl-CoA by carboxylation, catalyzed by an acetyl-CoA carboxylase (EC 6.4.1.2). OLS contains 385 AA with no signal peptide and is found in flowers and rapidly expanding leaves. Interestingly, this enzyme itself does not produce OA, but rather olivetol, triketide pyrone, and tetraketide pyrone, suggesting that another PKS may be involved in OA biosynthesis (138). Similarly, OA has never been detected in a *Cannabis* plant (2). In 2012, a small protein (12 kDa) was isolated with high expression levels in glandular trichomes, which showed similarities to a dimeric $\alpha + \beta$ barrel (DABB)-type polyketide cyclase (31). This molecule has been suggested as a possible olivetolic acid cyclase (OAC).

GPP is a general precursor molecule for various pathways and is produced through condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) by GPP synthase (GPPS). In *Cannabis* and most other plants, GPPS uses IPP and DMAPP produced via the MEP pathway in the plastid (15).

9.2. Cannabigerolic, Δ^9-Tetrahydrocannabinolic, Cannabidiolic, and Cannabichromenic Acid Synthesis

Cannabigerolic acid (CBGA) is synthesized by the aromatic prenyltransferase CBGA synthase (CBGAS), through C-prenylation of OA by GPP. CBGA is assumed to be a central precursor of cannabinoid biosynthesis, since it is used to produce THCA, CBDA, and cannabichromenic acid (CBCA) through the cyclization of a prenyl moiety with the aid of specific synthases. Whereas OA seems to be used as a prenyl acceptor by CBGAS, different prenyl donors like GPP, and to a lesser extent neryl diphosphate, can be used (30).

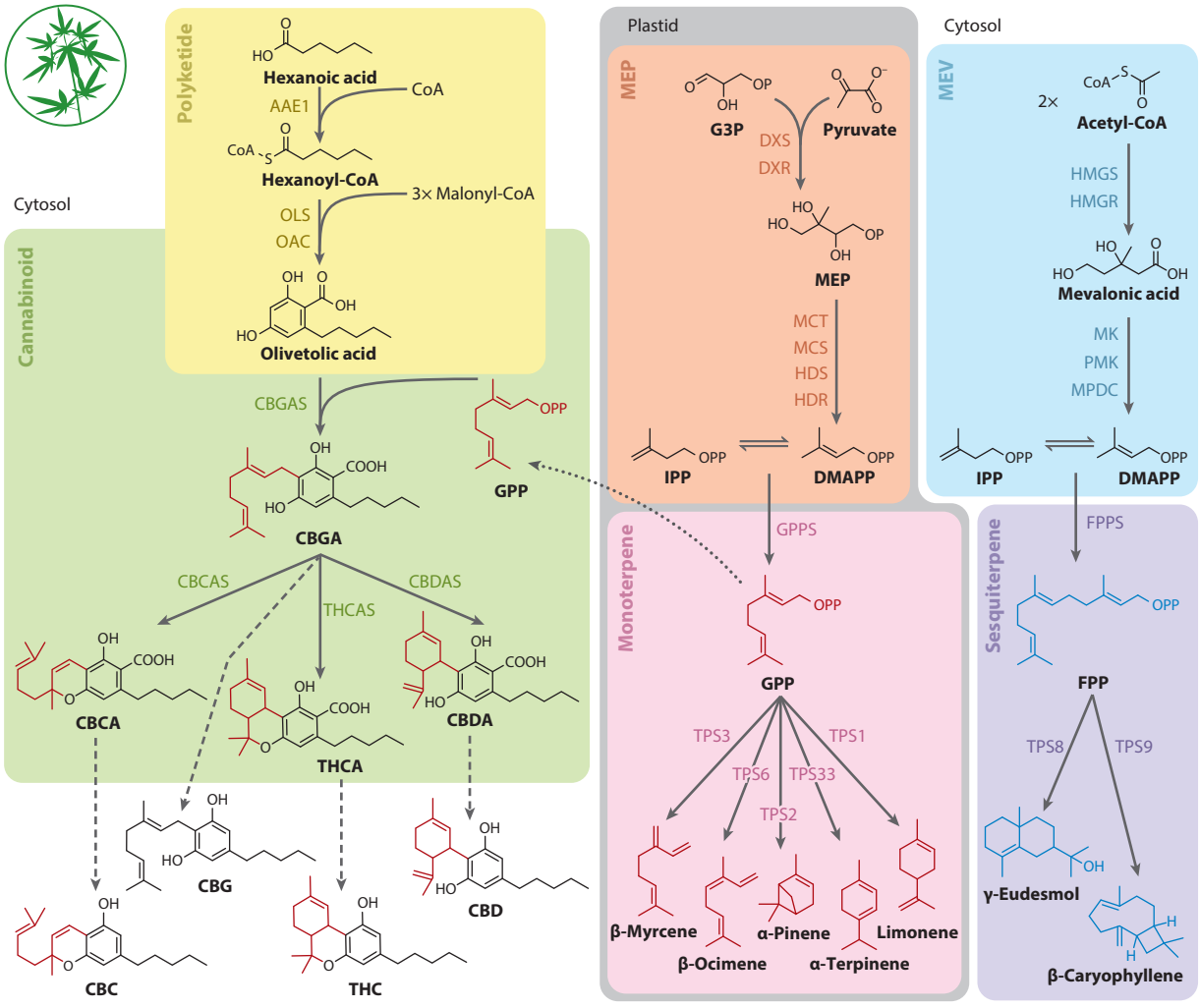


Figure 4

Cannabis sativa terpene and cannabinoid pathways. Colored regions indicate different pathways. The gray background indicates processes located in the plastid; other processes are located in the cytosol. The dotted arrow indicates the transfer of GPP from MEP to the cannabinoid pathway; dashed arrows indicate nonenzymatic decarboxylations. Enzyme names are shown in color, while resulting compounds are shown in black. Compound (sub)structures depicted in red signify those that represent a single unit of GPP; those depicted in blue signify a single unit of FPP. Abbreviations: AAE1, acyl-activating enzyme 1; CBC, cannabichromene; CBCA(S), cannabichromenic acid (synthase); CBD, cannabidiol; CBDA(S), cannabidiolic acid (synthase); CBG, cannabigerol; CBGA(S), cannabigerolic acid (synthase); DMAPP, dimethylallyl pyrophosphate; DXR, deoxyxylulose phosphate reductoisomerase; DXS, deoxyxylulose phosphate synthase; FPP(S), farnesyl pyrophosphate (synthase); G3P, glyceraldehyde 3-phosphate; GPP(S), geranyl-pyrophosphate (synthase); HDR, hydroxymethylbutenyl diphosphate reductase; HDS, hydroxymethylbutenyl diphosphate synthase; HMGR, hydroxymethylglutaryl-CoA reductase; HMGS, hydroxymethylglutaryl-CoA synthase; IPP, isopentenyl pyrophosphate; MCS, methylerythritol cyclodiphosphate synthase; MCT, methylerythritol phosphate cytidylyltransferase; MEP, methylerythritol phosphate; MEV, mevalonate; MK, mevalonate kinase; MPDC, mevalonate diphosphate decarboxylase; OAC, olivetolic acid cyclase; OLS, olivetol synthase; PMK, phosphomevalonate kinase; THC, tetrahydrocannabinol; THCA(S), tetrahydrocannabinolic acid (synthase); TPS, terpene synthase.

THCAS and CBDA synthase (CBDAS) have been well characterized in recent studies. They appear highly similar in their biochemical properties, such as mass (both are 74-kDa monomeric proteins) and K_m for CBGA, and they share 84% similarity at the amino acid level (103). Both enzymes have a 28-AA signal peptide, which dissociates to a secreted mature protein and a flavin adenine dinucleotide (FAD)-binding domain (103, 114). THCAS and CBDAS show high domain homology with the berberine bridge enzyme (BBE) involved in the alkaloid biosynthesis of *Eschscholzia californica*. Both *Cannabis* synthases and BBE activity require molecular oxygen and produce hydrogen peroxide as a side product (100). The sequence similarity of THCAS and CBDAS with the BBE has been confirmed (80).

Synthases for all three major cannabinoids are coded by single exon genes (56). All three genes have high levels of homology—84% similarity between THCAS and CBDAS and 96% between THCAS and CBCAS (56, 114). Recently, copy number variation of the synthase genes in this pathway was found to account for some of the variation in cannabinoid content amongst Cannabis cultivars (67, 125). Instead, a recent, extensive characterization of THCAS and CBDAS variants indicated that the amino acid sequence may affect in vitro activity (146). Although most mutations around catalytic, glycosylation, and disulfide bridge sites resulted in decreased activity or even full inactivation, some caused an increase in activity compared to the wild-type enzyme. Cannabinoid-free *Cannabis* plants have been reported, which may comprise nonfunctional synthase genes in either the polyketide pathway or CBGAS.

Heterologous expression of recombinant THCA and CBDA synthases in the yeast *Komagataella phaffii* followed by whole-cell bioconversion of CBGA showed that both synthases formed up to eight different products. Surprisingly, each enzyme showed the ability to form all three of the major cannabinoids (THCA, CBDA, and CBCA) in different ratios (146). Due to their similarity in structure and supposed catalytic mechanism, it has been suggested that the product specificity of THCAS and CBDAS would be determined by a small number of amino acid residues (114). Microcapillary sampling of stalked trichomes found in mature flowers and immature sessile trichomes found on vegetative leaves revealed similar cannabinoid profiles between the trichome types (58), although it is unclear whether this is true across the genetic spectrum of *Cannabis*.

The sequence similarity between THCAS and CBDAS implies a common ancestry (114), and sequence analysis of THCAS and CBDAS from different cultivars suggests that CBDAS is the ancestral gene and that its duplication led to higher CBDAS variation and the origin of THCA synthase (80). Moreover, additional synthase gene variants have been described (38, 114, 131), suggesting that the structural and functional evolution of these genes is more complex than initially thought. Recent research has shown that CBDAS and THCAS are not two codominant alleles from the same locus as previously suggested but rather two different genes found in very close proximity within a largely nonrecombining region (38, 131) (**Figure 4**).

Commonly known cannabinoids, such as CBGA, THCA, CBDA, and CBCA, are C5-phytocannabinoids that contain an *n*-pentyl side chain. At the same time, *Cannabis* plants also produce minor cannabinoids with an *n*-propyl side chain, called C3-phytocannabinoids. Cannabigerovarinic acid is synthesized by the C-prenylation of divarinic acid instead of OA by GPP (24). It is believed THCAS, CBDAS, and CBCAS are not selective for the substrate and can use CBGVA to produce tetrahydrocannabivarinic acid, cannabidivarinic acid, and cannabichrovarinic acid, respectively (99). This hypothesis requires in vitro confirmation.

10. BIOSYNTHESIS OF TERPENES

Terpenes are responsible for the characteristic aromas of both *Cannabis* and hop. They are one of the largest groups of phytochemicals, and they are classified based on the number of five-carbon building blocks, including monoterpenes (10 carbons), sesquiterpenes (15 carbons),

and triterpenes (30 carbons). In plants, terpenes are synthesized via two different pathways: the cytosolic mevalonic acid pathway contributing to the biosynthesis of sesquiterpenes and triterpenes and the plastid MEP pathway involved in the synthesis of mono-, di-, and tetraterpenes (13). Two molecules of acetyl-coenzyme A are used in the mevalonic acid pathway, whereas pyruvate and D-glyceraldehyde-3-phosphate are used in the MEP pathway. These molecules are converted to IPP and then further isomerized to DMAPP. Monoterpenes are made from the GPP precursor formed by MEP in the plastid, where GPP synthase catalyzes the condensation of one molecule of IPP and one molecule of DMAPP into GPP (129). Sesquiterpenes and triterpenes are produced in the cytosol from a farnesyl diphosphate (FPP) precursor; two molecules of IPP and one molecule of DMAPP are condensed with the aid of FPP synthase to produce FPP. Various terpene synthases convert FPP into sesquiterpenes, whereas triterpenes are produced by condensing two FPP molecules into squalene by squalene synthase. Squalene is then used to produce various triterpenes and sterols (5).

The characterization of terpene synthase (TPS) genes (including additional synthases) in *Cannabis* (11, 39, 58, 139) raises the question as to how they vary (e.g., polymorphism levels and copy number variation) between cultivars. Recently, a gene expression analysis of cannabinoids and terpenoids in eight distinct commercial *Cannabis* varieties confirmed the presence of synthase transcripts and metabolites for most terpenoids, and in addition, TPS genes encoding linalool and nerolidol were identified and characterized from two different cultivars (139).

Many TPS genes are upregulated during the process of calyx maturation (e.g., the development of stalked trichomes from sessile trichomes); this is especially evident for genes involved in the synthesis of monoterpenes (58). Whereas the ratio of cannabinoids stays relatively stable throughout the development of vegetative tissues and maturation of calyces, the ratio of monoterpenes to sesquiterpenes increases with calyx maturation, reaching 12:1 in stalked trichomes. Sequence polymorphisms in TPS genes are not well documented. Zager et al. (139) identified a gene coding for an enzyme (CsTPS19BL) that generates a mixture of (1)-linalool and (2)-linalool from GPP and (E)-nerolidol from tFPP in the cultivar black lime. Curiously, an ortholog cloned from the cultivar Valley Fire that had three mismatched nucleotides produced exclusively (1)-linalool from tFPP instead of nerolidol (139). To date, 18 TPS belonging to two subfamilies, TPS-a and TPS-b, were identified and described (11, 39, 58, 139) (see **Supplemental Table 1**). No sequence polymorphisms or changes in regulation of TPS expression between hemp- and drug-type cultivars have been reported. Nevertheless, the availability of improved *Cannabis* genome assemblies of FN, PK, and CBDRx should allow for the quick identification of the variants in *Cannabis* cultivars (38, 56).

Supplemental Material >

The hop cones of *H. lupulus* generally include relatively large fractions of the monoterpenes β-myrcene and β-pinene and the sesquiterpenes humulene and β-caryophyllene, and genes encoding the enzymes synthesizing β-myrcene, β-caryophyllene, and humulene have been described (130). However, it is unclear if functionally equivalent terpene synthases in *Cannabis* and *Humulus* have the same (orthologous) origin, or resulted from parallel (paralogous) evolution. No terpene-free *Cannabis* or *Humulus* plants are known (11, 24), probably because terpenes are precursors for various vital pathways.

11. BITTER ACID AND XANTHOHUMOL BIOSYNTHESIS IN *HUMULUS LUPULUS*

Similar to the cannabinoid pathway, the production of bitter acids and xanthohumols in *H. lupulus* relies on the prenylation of polyketide precursors (**Figure 5**). The polyketide precursor for bitter acid biosynthesis is phlorisovalerophenone (PIVP), which is produced by first converting ketoisocaproic acid into isovaleryl-CoA by two carboxyl CoA ligases (CCLs) named CCL1 and

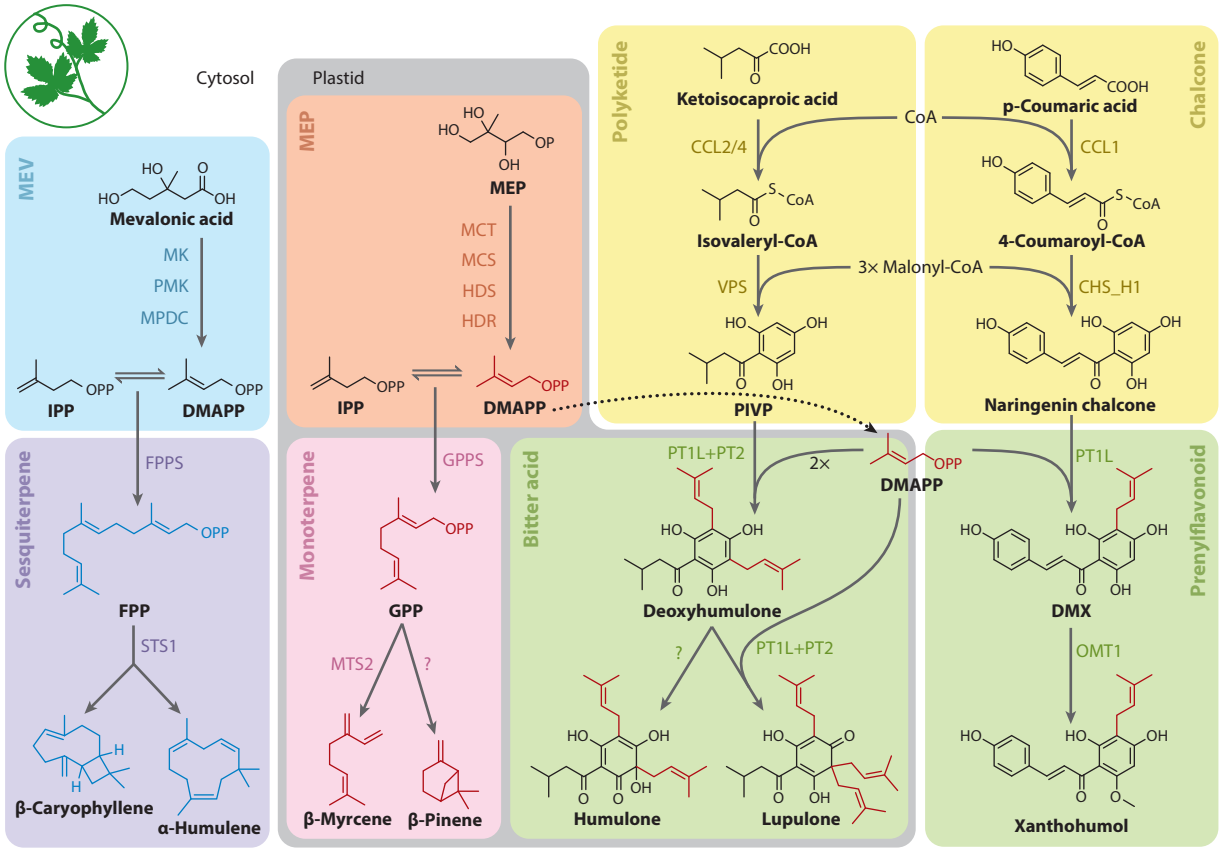


Figure 5

Humulus lupulus terpene, bitter acid, and xanthohumol pathways. Colored regions indicate different pathways. The gray background indicates processes located in the plastid; other processes are located in the cytosol. The dotted arrow indicates the transfer of DMAPP from MEP to bitter acid and prenylflavonoid pathways. Enzyme names are shown in color, while compounds are shown in black. Compound (sub)structures depicted in red signify those that represent unit(s) of DMAPP from the MEP pathway; those depicted in blue signify a single unit of FPP. Question marks indicate an unknown synthase producing the monoterpene α-pinene and an unknown monooxygenase producing the α-bitter acid humulone. CoA, malonyl-CoA, and DMAPP span across pathways and are therefore positioned between colored regions. Abbreviations: CCL, carboxyl CoA ligase; CHS_H1, chalcone synthase homolog 1; DMAPP, dimethylallyl pyrophosphate; DMX, desmethylxanthohumol; FPP(S), farnesyl pyrophosphate (synthase); GPP(S), geranyl-pyrophosphate (synthase); HDR, hydroxymethylbutenyl diphosphate reductase; HDS, hydroxymethylbutenyl diphosphate synthase; IPP, isopentenyl pyrophosphate; MCS, methylerythritol cyclodiphosphate synthase; MCT, methylerythritol phosphate cytidylyltransferase; MEP, methylerythritol phosphate; MEV, mevalonate; MK, mevalonate kinase; MPDC, mevalonate diphosphate decarboxylase; MTS, monoterpene synthase; OMT, *O*-methyltransferase; PIVP, phlorisovalerophenone; PMK, phosphomevalonate kinase; PT, prenyltransferase; STS, sesquiterpene synthase; VPS, phlorisovalerophenone synthase.

CCL2 (21, 135) and then adding three molecules of malonyl-CoA by PIVP synthase (78). PIVP is then prenylated with two molecules of DMAPP by two prenyltransferases PT1L and PT2, forming the general bitter acid precursor deoxyhumulone, which is either further prenylated by the same two PTs into the β-bitter acid lupulone or monoxygenated by an unknown enzyme into the α-bitter acid humulone (57). The polyketide precursor for xanthohumol is the common chalconoid naringenin chalcone. It is produced by first converting *p*-coumaric acid into 4-coumaroyl-CoA by CCL1, followed by adding three molecules of malonyl-CoA by the chalcone

synthase homolog CHS_H1 (76). Naringenin chalcone is then prenylated with one molecule of DMAPP by PT1L to form desmethylxanthohumol, which is further converted into xanthohumol by *O*-methyltransferase OMT1 (72). Despite the obvious analogies between cannabinoid, bitter acid, and xanthohumol biosynthesis, it remains unknown to what extent these pathways share a common evolutionary origin.

12. *CANNABIS* DISCOVERY PROTEOMICS

Proteomic studies of *Cannabis* are rare due to global legislative restrictions. However, the increasing availability of comprehensive genomic and transcriptomic resources, as well as a relaxation of the legislative framework for *Cannabis* research, is likely to spark an increase in *Cannabis* discovery proteomics studies.

Proteomic studies of hemp have investigated hempseed as a highly digestible, protein-rich food source. Two comprehensive proteomic studies of hempseed have been conducted, both descriptive in nature, providing a foundation for the agricultural development of hemp as a functional food (3, 83). More recently, Mamone and colleagues (63) experimentally verified hemp-type *Cannabis* as a hypoallergenic protein source, reporting low abundance of allergenic peptides after gastrointestinal digestion.

Protein studies of drug-type *Cannabis* have predominantly focused on elucidating the biosynthetic pathways for a small number of highly abundant cannabinoids and terpenes, as described above. The first published proteomics study of drug-type *Cannabis* compared the leaf and flower proteomes using 2D gel electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF-MS) to identify the proteins specifically expressed in floral glandular trichomes (89). Over 300 proteins were identified as unique to flowers; fewer than 100 proteins were found in gland extracts. However, this method did not identify any of the known cannabinoid biosynthesis enzymes. The authors suggested that the concentrations of these enzymes may be below the detection limit; however, more recent data indicate that THCAS is one of the most highly abundant proteins in glandular trichomes (40, 127). Since we now know that cannabinoid synthesis enzymes are also expressed in vegetative tissue (93), it seems plausible that these spots were present in both gels and thus not selected as differential or else that extraction methods were insufficient to extract these enzymes.

Improvements in gel-free proteomics technology enabled two further reports characterizing the proteome of *Cannabis* inflorescences and trichomes (40, 127). These studies conducted peptide fingerprinting analyses of complex protein extracts from isolated trichomes or apical buds using nanoflow liquid chromatography tandem mass spectrometry (nanoLC-MS/MS). Additionally, Vincent and colleagues (127) conducted ultra-performance liquid chromatography (UPLC) top-down proteomics of intact protein mixtures. Two of these studies each identified over 400 unique proteins from isolated floral trichomes, and of those, 6.9% (40) and 27% (127) were identified as being involved in the synthesis of secondary metabolites, including most known enzymes of the cannabinoid, terpenoid, and flavonoid biosynthetic pathways. Indeed, both studies identified THCAS and CBDAS with good coverage (up to 63.9%) (40). The difference in the abundance of biosynthesis enzymes in the two studies may be due to the different *Cannabis* cultivars analyzed; however, it seems more likely to be a result of different protein extraction methods. Additionally, 11.0% (40) and 12.3% (127) of identified proteins were of unknown function. As improved genomic resources become available, these uncharacterized proteins may prove to be of interest as biosynthetic enzymes.

An alternate approach conducted by Jenkins and Orsburn (48) using simultaneous next-generation sequencing and shotgun proteomics has been successful in generating an annotated

proteome map without using the standard annotated genome approach. This novel method enabled the identification of over 17,000 putative protein sequences from *Cannabis* leaf, stem, and flower tissue and is the first comprehensive annotated protein FASTA database for this genus. Even so, the identification rate of peptides was <6%, compared with 22% for *Arabidopsis*, indicating that improved genomic resources will be critical for further analyses of this type. Unsurprisingly, a comparison of identified protein networks from leaf and mature female flower tissue revealed substantial differences and corroborated data from the previous gel-free studies. Proteins identified in leaves belonged primarily to the photosynthetic pathway, while those from flowers were predominantly associated with secondary metabolite biosynthesis, including synthesis of cannabinoids, fatty acids, and terpenes (48).

Haplogroup: individuals that share patterns of genetic variation at tightly linked loci

To date, no studies have been published that compare the proteome of different *Cannabis* cultivars. The compilation of large collections of *Cannabis* cultivars and varieties will enable multidimensional comparative proteomic analyses. Combining these data with genomic and metabolomic information as well as cultivar-specific therapeutic or psychedelic user experiences may reveal bioactive low-abundance metabolites, their synthesis enzymes and metabolic pathways.

13. *CANNABIS* GENETIC DIVERSITY

It has been reported that the distinction between hemp- and drug-type *Cannabis* cultivars is not merely based on cannabinoid content but can be traced across the entire genome (38, 43, 62, 98, 109), suggesting independent domestication from already differentiated germplasm stocks. However, these findings are based on limited sampling and lack detailed biogeographic and evolutionary analyses. Furthermore, drug-type *Cannabis* can be roughly divided into narrow-leaflet (commonly referred to as sativa) and broad-leaflet (commonly referred to as indica) cultivars, suggesting distinct genetic pools that may have arisen through breeding (36, 62, 98).

There is an indication that most genetic diversity and substructure in the species are found within hemp (32, 45, 145), where the existence of distinct gene pools in central Asia and Europe has long been recognized (23, 32). A chloroplast haplotype study (145) suggests that the major genetic substructure within hemp-type *Cannabis* might follow a latitudinal gradient. Importantly, major hemp-type cultivars included in this study were part of different haplogroups, suggesting that geographically differentiated gene pools were used for domestication. Such a multiregional origin of domestication would also be supported by fossil data (59, 68). However, a globally sampled genomic data set is required to validate these results and fully understand *Cannabis* phylogeography.

Heterozygosity is another important measure of genetic diversity and can help to elucidate some aspects of crop evolutionary history. As a wind-pollinated crop, *Cannabis* generally exhibits high levels of heterozygosity, but several authors have reported differences in heterozygosity between infrageneric groups. Sawler et al. (98) found significantly higher heterozygosity levels in hemp (~16%) than in other *Cannabis* groups (~12.5%) and interpreted this as indicative of a broader genetic base of hemp and/or less hybridization between closely related accessions. Soler et al. (108) found a similar trend of higher levels of heterozygosity in hemp (40.5%) than in drug-type *Cannabis* (28.2%).

In contrast, Lynch et al. (62) reported significantly lower levels of heterozygosity for European hemp accessions than for drug-type accessions (22% versus 31%), which they interpret as evidence of recent breeding efforts in the North American drug-type *Cannabis*, compared to the long breeding history of hemp. However, this study had only limited sampling of Asian hemp accessions, and in contrast, Gao et al. (32) showed significantly higher levels of heterozygosity in Chinese hemp (35.5–37%) compared to European hemp (18.2%).

The assembly of a global data set comprising various types of *Cannabis* materials (i.e., undomesticated, naturalized, and landraces) as well as comprehensive sampling of all putative intrageneric subgroups should therefore be a priority for future evolutionary studies. Additionally, important sampling gaps, likely due to the difficulty of securing germplasm from certain regions, need to be filled. Specifically, focus needs to be placed on central Asian accessions, as only 3% (i.e., 15) of 470 accessions (landraces and naturalized populations) included in the three most recent genotyping studies were reliably known to originate from this region (62, 98, 109). Considering the remarkable genetic diversity harbored by regions of Asia (45, 143) and its hypothesized importance for the origin and domestication of the genus (23, 69, 105), this needs to be urgently addressed. Emphasis should be placed on sampling landraces and wild/naturalized lines by making use of germplasm bank (73) and herbarium collections. Finally, voucher information and passport data (including putative material type and origin) need to be properly recorded, as conventionally done in other phylogeographic studies.

Landrace: traditionally cultivated variety that is adapted to local growing conditions and farming practices

14. CONCLUSION

Clearly there is a wealth of both data and active research in the *Cannabis* area, much of which is being spurred on by an industry growing into a novel niche created by recent societal and governmental acceptance. Nonetheless, there are large gaps in our knowledge of this high-demand crop on many levels, and as such, a more coordinated effort to quantify, document, and protect this species' phenotypic, genomic, and biochemical diversity is needed. Here we show a number of gaps in our acquired knowledge of *Cannabis* genomics and biochemistry as a guide forward in the development of this multipurpose crop.

DISCLOSURE STATEMENT

J.K.M. is the Director of Genetics for New West Genetics, Inc. and holder of US Patent 10,499,584 B2. R.v.V. is affiliated with Bedrocan International, a private company producing medicinal *Cannabis*. T.F.S. and M.P. are affiliated with Kanata Earth, a private company producing medicinal *Cannabis*. D.V. is the founder and president of the nonprofit organization Agricultural Genomics Foundation, and the sole owner of CGRI, LLC.

AUTHOR CONTRIBUTIONS

I.K., J.A., R.B.L., and R.v.V. wrote Sections 8–10. M.P. wrote Sections 5–7. P.R., C.B., and D.S. performed the genomics meta-analyses and wrote Sections 3 and 5. J.E., R.v.V., M.E.S., J.K.M., D.V., and N.C.K. wrote Section 2. J.E. wrote Section 12. M.M. and R.v.V. wrote Section 4. T.F.S. and M.P. developed the article's context, coordinated the different authors, and wrote and edited the article.

ACKNOWLEDGMENTS

This work was partially funded by the grant CannaNeXT: A Next-Generation *Cannabis* Breeding and Innovation Platform (UofS 420852-1019).

LITERATURE CITED

1. Abbott JK, Nordén AK, Hansson B. 2017. Sex chromosome evolution: historical insights and future perspectives. *Proc. R. Soc. B* 284:20162806
2. Aghazadeh Tabrizi M, Baraldi PG. 2017. Chemistry of cannabinoid receptor agonists. In *Handbook of Cannabis and Related Pathologies*, ed. VR Preedy, pp. 592–605. London: Academic

3. Aiello G, Fasoli E, Boschin G, Lammi C, Zanoni C, et al. 2016. Proteomic characterization of hempseed (*Cannabis sativa* L.). *J. Proteom.* 147:187–96
4. Anavi-Goffer S, Baillie G, Irving AJ, Gertsch J, Greig IR, et al. 2012. Modulation of L-α-lysophosphatidylinositol/GPR55 mitogen-activated protein kinase (MAPK) signaling by cannabinoids. *J. Biol. Chem.* 287(1):91–104
5. Andre CM, Hausman J-F, Guerriero G. 2016. *Cannabis sativa*: the plant of the thousand and one molecules. *Front. Plant Sci.* 7:19
6. Atakan Z. 2012. Cannabis, a complex plant: different compounds and different effects on individuals. *Ther. Adv. Psychopharmacol.* 2(6):241–54
7. Badowski M, Yanful PK. 2018. Dronabinol oral solution in the management of anorexia and weight loss in AIDS and cancer. *Ther. Clin. Risk Manag.* 14:643–51
8. Behre K-E. 1999. The history of beer additives in Europe—a review. *Veg. Hist. Archaeobot.* 8(1–2):35–48
9. Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* 97(8):1296–303
10. Bocquet L, Sahpaz S, Hilbert JL, Rambaud C, Rivière C. 2018. *Humulus lupulus* L., a very popular beer ingredient and medicinal plant: overview of its phytochemistry, its bioactivity, and its biotechnology. *Phytochem. Rev.* 17(5):1047–90
11. Booth JK, Page JE, Bohlmann J. 2017. Terpene synthases from *Cannabis sativa*. *PLOS ONE* 12(3):e0173911
12. Boutain JR, Boutain MR. 2015. *Nanopore DNA sequencing a native North American hop (*Humulus lupulus *var*. lupuloides*) with implications for research on Cannabaceae collections*. Presented at IV International Humulus Symposium, Yakima, WA
13. Bouvier F, Rahier A, Camara B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* 44(6):357–429
14. Bowen LL, McRae-Clark AL. 2018. Therapeutic benefit of smoked cannabis in randomized placebo-controlled studies. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* 38(1):80–85
15. Burke CC, Wildung MR, Croteau R. 1999. Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer. *PNAS* 96(23):13062–67
16. Carvalho Â, EH Hansen, Kayser O, Carlsen S, Stehle F. 2017. Designing microorganisms for heterologous biosynthesis of cannabinoids. *FEMS Yeast Res.* 17(4):fox037
17. Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 251:1030–33
18. Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* 112(988):975–97
19. Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95(2):118–28
20. Chen JW, Borgelt LM, Blackmer AB. 2019. Cannabidiol: a new hope for patients with Dravet or Lennox-Gastaut syndromes. *Ann. Pharmacother.* 53(6):603–11
21. Clark SM, Vaitheeswaran V, Ambrose SJ, Purves RW, Page JE. 2013. Transcriptome analysis of bitter acid biosynthesis and precursor pathways in hop (*Humulus lupulus*). *BMC Plant Biol.* 13:12
22. Clarke RC. 2007. Traditional *Cannabis* cultivation in Darchula District, Nepal—seed, resin and textiles. *J. Ind. Hemp* 12(2):19–42
23. Clarke RC, Merlin MD. 2013. *Cannabis: Evolution and Ethnobotany*. Berkeley: Univ. Calif. Press
24. de Meijer EPM, Hammond AKM, Sutton AA, Hammond KM, Sutton AA. 2009. The inheritance of chemical phenotype in *Cannabis sativa* L. IV. Cannabinoid-free plants. *Euphytica* 168(1):95–112
25. Delyser DY, Kasper WJ. 1994. Hopped beer: the case for cultivation. *Econ. Bot.* 48(2):166–70
26. Divashuk MG, Alexandrov OS, Razumova OV, Kirov IV, Karlov GI. 2014. Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. *PLOS ONE* 9(1):e85118
27. Duvall CS. 2017. Cannabis and tobacco in precolonial and colonial Africa. In *Oxford Research Encyclopedia of African History*, ed. T Spear. New York: Oxford Univ. Press. **https://doi.org/10.1093/acrefore/9780190277734.013.44**
28. Faux A-M, Berhin A, Dauguet N, Bertin P. 2014. Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). *Euphytica* 196(2):183–97

29. Fellermeier M, Eisenreich W, Bacher A, Zenk MH. 2001. Biosynthesis of cannabinoids. *Eur. J. Biochem.* 268(6):1596–604
30. Fellermeier M, Zenk MH. 1998. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Lett.* 427(2):283–85
31. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE. 2012. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *PNAS* 109(31):12811–16
32. Gao C, Xin P, Cheng C, Tang Q, Chen P, et al. 2014. Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. *PLOS ONE* 9(10):e110638
33. Gerhäuser C. 2009. Phenolic beer compounds to prevent cancer. In *Beer in Health and Disease Prevention*, ed. VR Preedy, pp. 669–84. Burlington, MA: Academic
34. Gertsch J, Leonti M, Raduner S, Racz I, Chen J-Z, et al. 2008. Beta-caryophyllene is a dietary cannabinoid. *PNAS* 105(26):9099–104
35. Giacoppo S, Bramanti P, Mazzon E. 2017. Sativex in the management of multiple sclerosis-related spasticity: an overview of the last decade of clinical evaluation. *Mult. Scler. Relat. Disord.* 17:22–31
36. Gilmore S, Peakall R, Robertson J. 2007. Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*. *Forensic Sci. Int.* 172(2–3):179–90
37. Grabowska-Joachimiak A, Śliwińska E, Pigula M, Skomra U, Joachimiak AJ. 2006. Genome size in *Humulus lupulus* L. and *H. japonicus* Siebold and Zucc. (Cannabaceae) 2006. *Acta Soc. Bot. Pol.* 75(3):207–14
38. Grassa CJ, Wenger JP, Dabney C, Poplawski SG, Motley ST, et al. 2018. A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. bioRxiv 458083. **https://doi.org/10.1101/458083**
39. Günnewich N, Page JE, Köllner TG, Degenhardt J, Kutchan TM. 2007. Functional expression and characterization of trichome-specific (-)-limonene synthase and (+)-α-pinene synthase from *Cannabis sativa*. *Nat. Prod. Commun.* 2(3):223–32
40. Happyana N. 2014. *Metabolomics, proteomics, and transcriptomics of* Cannabis sativa L. *trichomes*. PhD Thesis, TU Dortmund Univ., Dortmund, Ger.
41. Häuser W, Welsch P, Klose P, Radbruch L, Fitzcharles M-A. 2019. Efficacy, tolerability and safety of cannabis-based medicines for cancer pain. *Schmerz* 33(5):424–36
42. Hill ST, Sudarsanam R, Henning J, Hendrix D. 2017. HopBase: a unified resource for *Humulus* genomics. *Database* 2017:bax009
43. Hillig KW. 2005. Genetic evidence for speciation in *Cannabis* (Cannabaceae). *Genet. Resour. Crop Evol.* 52(2):161–80
44. Hryhorowicz S, Walczak M, Zakerska-Banaszak O, Słomski R, Skrzypczak-Zielińska M. 2018. Pharmacogenetics of cannabinoids. *Eur. J. Drug Metab. Pharmacokinet.* 43(1):1–12
45. Hu Z-G, Guo H-Y, Hu X-L, Chen X, Liu X-Y, et al. 2012. Genetic diversity research of hemp (*Cannabis sativa* L) cultivar based on AFLP analysis. *J. Plant Genet. Resour.* 13(4):555–61
46. Izzo AA, Coutts AA. 2005. Cannabinoids and the digestive tract. *Handb. Exp. Pharmacol.* 168:573–98
47. Jenkins C, Orsburn B. 2019. The first publicly available annotated genome for Cannabis plants. bioRxiv 786186. **https://doi.org/10.1101/786186**
48. Jenkins C, Orsburn B. 2020. The Cannabis proteome draft map project. *Int. J. Mol. Sci.* 21(3):965
49. Jiang H-E, Li X, Zhao Y-X, Ferguson DK, Hueber F, et al. 2006. A new insight into *Cannabis sativa* (Cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China. *J. Ethnopharmacol.* 108(3):414–22
50. Jiang H-E, Merlin M. 2016. Ancient *Cannabis* burial shroud in a central Eurasian cemetery. *Econ. Bot.* 70(3): 213–21
51. Jin J-J, Yang M-Q, Fritsch PW, van Velzen R, Li D-Z, Yi T-S. 2019. Born migrators: historical biogeography of the cosmopolitan family Cannabaceae. *J. System. Evol.* In press. **https://doi.org/10.1111/jse.12552**
52. Kathmann M, Flau K, Redmer A, Tränkle C, Schlicker E. 2006. Cannabidiol is an allosteric modulator at mu- and delta-opioid receptors. *Naunyn Schmiedebergs Arch. Pharmacol.* 372(5):354–61

53. Keeble-Gagnère G, Rigault P, Tibbits J, Pasam R, Hayden M, et al. 2018. Optical and physical mapping with local finishing enables megabase-scale resolution of agronomically important regions in the wheat genome. *Genome Biol*. 19(1):112
54. Kimura M, Okamoto K. 1970. Distribution of tetrahydrocannabinolic acid in fresh wild cannabis. *Experientia* 26(8):819–20
55. Kowal MA, Hazekamp A, Colzato LS, van Steenbergen H, van der Wee NJA, et al. 2015. Cannabis and creativity: Highly potent cannabis impairs divergent thinking in regular cannabis users. *Psychopharmacology* 232(6):1123–34
56. Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, et al. 2019. A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the *THC/CBD acid synthase* loci. *Genome Res*. 29(1):146–56
57. Li H, Ban Z, Qin H, Ma L, King AJ, Wang G. 2015. A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. *Plant Physiol*. 167(3):650–59
58. Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, et al. 2019. Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *Plant J*. 101:37–56
59. Long T, Wagner M, Demske D, Leipe C, Tarasov PE. 2017. Cannabis in Eurasia: origin of human use and Bronze Age trans-continental connections. *Veg. Hist. Archaeobot*. 26(2):245–58
60. Lowe DJE, Sasiadek JD, Coles AS, George TP. 2019. Cannabis and mental illness: a review. *Eur. Arch. Psychiatry Clin. Neurosci*. 269(1):107–20
61. Lu H-C, Mackie K. 2016. An introduction to the endogenous cannabinoid system. *Biol. Psychiatr.* 79(7):516–25
62. Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, et al. 2016. Genomic and chemical diversity in *Cannabis*. *Crit. Rev. Plant Sci*. 35(5–6):349–63
63. Mamone G, Picariello G, Ramondo A, Nicolai MA, Ferranti P. 2019. Production, digestibility and allergenicity of hemp (*Cannabis sativa* L.) protein isolates. *Food Res. Int.* 115:562–71
64. Mandolino G, Carboni A, Forapani S, Faeti V, Ranalli P. 1999. Identification of DNA markers linked to the male sex in dioecious hemp (*Cannabis sativa* L.). *Theor. Appl. Genet.* 98(1):86–92
65. Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, et al. 2009. Identification of candidate genes affecting Δ^9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J. Exp. Bot*. 60(13):3715–26
66. Maroon J, Bost J. 2018. Review of the neurological benefits of phytocannabinoids. *Surg. Neurol. Int.* 9(1):91
67. McKernan KJ, Helbert Y, Kane LT, Ebling H, Zhang L, et al. 2020. Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. bioRxiv 894428. **https://doi.org/10.1101/2020.01.03.894428**
68. McPartland JM. 2018. *Cannabis* systematics at the levels of family, genus, and species. *Cannabis Cannabinoid Res*. 3(1):203–12
69. McPartland JM, Hegman W, Long T. 2019. *Cannabis* in Asia: its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Veg. Hist. Archaeobot*. 28(6):691–702
70. Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* 62:485–514
71. Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G. 2004. The sexual differentiation of *Cannabis sativa* L.: a morphological and molecular study. *Euphytica* 140(1):95–106
72. Nagel J, Culley LK, Lu Y, Liu E, Matthews PD, et al. 2008. EST analysis of hop glandular trichomes identifies an *O*-methyltransferase that catalyzes the biosynthesis of xanthohumol. *Plant Cell* 20(1):186–200
73. Naraine SGU, Small E. 2017. Germplasm sources of protective glandular leaf trichomes of hop (*Humulus lupulus*). *Genet. Resour. Crop Evol.* 64(7):1491–97
74. Natsume S, Takagi H, Shiraishi A, Murata J, Toyonaga H, et al. 2015. The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant Cell Physiol*. 56(3):428–41
75. Neve RA. 1961. *Sex determination in the cultivated hop* Humulus lupulus. PhD Thesis, Univ. London
76. Novak P, Krofta K, Matousek J. 2006. Chalcone synthase homologues from *Humulus lupulus*: some enzymatic properties and expression. *Biol. Plant.* 50(1):48–54

77. Oh H, Seo B, Lee S, Ahn D-H, Jo E, et al. 2015. Two complete chloroplast genome sequences of *Cannabis sativa* varieties. *Mitochondrial DNA A* 27(4):2835–37
78. Okada Y, Ito K. 2001. Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.). *Biosci. Biotechnol. Biochem.* 65(1):150–55
79. Ono T, Suzuki H. 1962. Cytological studies. In *The Wild Hop Native to Japan*, ed. T Ono, pp. 71–110. Sendai, Jpn.: Sasaki Print.
80. Onofri C, de Meijer EPM, Mandolino G. 2015. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry* 116:57–68
81. Page JE, Stout JM. 2015. *Cannabichromenic acid synthase from* Cannabis sativa. US Patent 10364416B2
82. Palmer JD. 1983. Chloroplast DNA exists in two orientations. *Nature* 301(5895):92–93
83. Park S-K, Seo J-B, Lee M-Y. 2012. Proteomic profiling of hempseed proteins from Cheungsam. *Biochim. Biophys. Acta* 1824(2):374–82
84. Parker JS, Clark MS. 1991. Dosage sex-chromosome system in plants. *Plant Sci.* 80(1–2):79–92
85. Peil A, Flachowsky H, Schumann E, Weber WE. 2003. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). *Theor. Appl. Genet.* 107(1):102–9
86. Pisupati R, Vergara D, Kane NC. 2018. Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. *BMC Genom.* 19(1):156
87. Prentout D, Razumova O, Rhoné B, Badouin H, Henri H, et al. 2020. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. *Genome Res.* 30(2):164–72
88. Radwan MM, ElSohly MA, El-Alfy AT, Ahmed SA, Slade D, et al. 2015. Isolation and pharmacological evaluation of minor cannabinoids from high-potency *Cannabis sativa*. *J. Nat. Prod.* 78(6):1271–76
89. Raharjo TJ, Chang W-T, Choi YH, Peltenburg-Looman AM, Verpoorte R. 2004. Olivetol as product of a polyketide synthase in *Cannabis sativa* L. *Plant Sci.* 166(2):381–85
90. Ren M, Tang Z, Wu X, Spengler R, Jiang H, et al. 2019. The origins of cannabis smoking: chemical residue evidence from the first millennium BCE in the Pamirs. *Sci. Adv.* 5(6):eaaw1391
91. Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* 101(10):1588–96
92. Ribeiro LI, Ind PW. 2016. Effect of cannabis smoking on lung function and respiratory symptoms: a structured literature review. *NPJ Prim. Care Respir. Med.* 26(1):16071
93. Richins RD, Rodriguez-Uribe L, Lowe K, Ferral R, O'Connell MA. 2018. Accumulation of bioactive metabolites in cultivated medical *Cannabis*. *PLOS ONE* 13(7):e0201119
94. Russo EB, Jiang H-E, Li X, Sutton A, Carboni A, et al. 2008. Phytochemical and genetic analyses of ancient cannabis from Central Asia. *J. Exp. Bot.* 59(15):4171–82
95. Sakamoto K, Akiyama Y, Fukui K, Kamada H, Satoh S. 1998. Characterization; genome sizes and morphology of sex chromosomes in hemp (*Cannabis sativa* L.). *Jpn. Mendel Soc.* 63(4):459–64
96. Sakamoto K, Ohmido N, Fukui K, Kamada H, Satoh S. 2000. Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Mol. Biol.* 44(6):723–32
97. Sakamoto K, Shimomura K, Komeda Y, Kamada H, Satoh S. 1995. A male-associated DNA sequence in a dioecious plant, *Cannabis sativa* L. *Plant Cell Physiol.* 36(8):1549–54
98. Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, et al. 2015. The genetic structure of marijuana and hemp. *PLOS ONE* 10(8):e0133292
99. Shoyama Y, Hirano H, Nishioka I. 1984. Biosynthesis of propyl cannabinoid acid and its biosynthetic relationship with pentyl and methyl cannabinoid acids. *Phytochemistry* 23(9):1909–12
100. Sirikantaramas S, Morimoto S, Shoyama YY, Ishikawa Y, Wada Y, et al. 2004. The gene controlling marijuana psychoactivity. *J. Biol. Chem.* 279(38):39767–74
101. Sirikantaramas S, Taura F. 2017. Cannabinoids: biosynthesis and biotechnological applications. In *Cannabis sativa L. - Botany and Biotechnology*, ed. S Chandra, H Lata, MA ElSohly, pp. 183–206. Cham, Switz.: Springer
102. Sirikantaramas S, Taura F, Morimoto S, Shoyama Y. 2007. Recent advances in Cannabis sativa research: biosynthetic studies and its potential in biotechnology. *Curr. Pharm. Biotechnol.* 8(4):237–43

103. Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, Shoyama Y. 2005. Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant Cell Physiol*. 46(9):1578–82
104. Small E. 1978. A numerical and nomenclatural analysis of morpho-geographic taxa of *Humulus*. *Syst. Bot*. 3(1):37–76
105. Small E. 2016. *Cannabis: A Complete Guide*. Boca Raton, FL: CRC Press
106. Small E. 2018. Dwarf germplasm: the key to giant Cannabis hempseed and cannabinoid crops. *Gen. Res. Crop Evol.* 65(4):1071–107
107. Small E, Pocock T, Cavers PB. 2003. The biology of Canadian weeds. 119. *Cannabis sativa* L. *Can. J. Plant Sci*. 83(1):217–37
108. Soler S, Gramazio P, Figàs MR, Vilanova S, Rosa E, et al. 2017. Genetic structure of *Cannabis sativa* var. *indica* cultivars based on genomic SSR (gSSR) markers: implications for breeding and germplasm management. *Ind. Crop. Prod.* 104:171–78
109. Soorni A, Fatahi R, Haak DC, Salami SA, Bombarely A. 2017. Assessment of genetic diversity and population structure in Iranian Cannabis germplasm. *Sci. Rep.* 7(1):15668
110. Stone NL, Millar SA, Herrod PJJ, Barrett DA, Ortori CA, et al. 2018. An analysis of endocannabinoid concentrations and mood following singing and exercise in healthy volunteers. *Front. Behav. Neurosci.* 12:269
111. Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. 2012. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. *Plant J.* 71(3):353–65
112. Sung B, Chung JW, Bae HR, Choi JS, Kim CM, Kim ND. 2015. *Humulus japonicus* extract exhibits antioxidative and anti-aging effects via modulation of the AMPK-SIRT1 pathway. *Exp. Ther. Med.* 9(5):1819–26
113. Sytsma KJ, Morawetz J, Pires JC, Nepokroeff M, Conti E, et al. 2002. Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on *rbcL*, *trnL-F*, and *ndhF* sequences. *Am. J. Bot.* 89(9):1531–46
114. Taura F, Sirikantaramas S, Shoyama YY, Yoshikai K, Shoyama YY, Morimoto S. 2007. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS Lett*. 581(16):2929–34
115. Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, et al. 2009. Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett*. 583 (12):2061–66
116. Törjék O, Bucherna N, Kiss E, Homoki H, Finta-Korpelová Z, et al. 2002. Novel male-specific molecular markers (MADC5, MADC6) in hemp. *Euphytica* 127(2):209–18
117. Udoh M, Santiago M, Devenish S, McGregor IS, Connor M. 2019. Cannabichromene is a cannabinoid CB2 receptor agonist. *Br. J. Pharmacol.* 176(23):4537–47
118. U. N. Off. Drugs Crime (UNODC). 2019. *World drug report 2019*. Rep., U.N., New York. **https://doi.org/10.18356/a4dd519a-en**
119. Urits I, Borchart M, Hasegawa M, Kochanski J, Orhurhu V, Viswanath O. 2019. An update of current Cannabis-based pharmaceuticals in pain medicine. *Pain Ther*. 8(1):41–51
120. van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, et al. 2011. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol*. 12(10):R102
121. van de Donk T, Niesters M, Kowal MA, Olofsen E, Dahan A, van Velzen M. 2019. An experimental randomized study on the analgesic effects of pharmaceutical-grade cannabis in chronic pain patients with fibromyalgia. *Pain* 160(4):860–69
122. van Velzen R, Doyle JJ, Geurts R. 2019. A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. *Trends Plant Sci*. 24(1):49–57
123. van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, et al. 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *PNAS* 115(20):E4700–9
124. Vergara D, Baker H, Clancy K, Keepers KG, Mendieta JP, et al. 2016. Genetic and genomic tools for *Cannabis sativa*. *Crit. Rev. Plant Sci*. 35(5–6):364–77

125. Vergara D, Huscher EL, Keepers KG, Givens RM, Cizek CG, et al. 2019. Gene copy number is associated with phytochemistry in *Cannabis sativa*. *AoB Plants* 11(6):plz074
126. Vergara D, White KH, Keepers KG, Kane NC. 2016. The complete chloroplast genomes of *Cannabis sativa* and *Humulus lupulus*. *Mitochondrial DNA A* 27(5):3793–94
127. Vincent D, Rochfort S, Spangenberg G. 2019. Optimisation of protein extraction from medicinal Cannabis mature buds for bottom-up proteomics. *Molecules* 24(4):659
128. Volkow ND, Baler RD, Compton WM, Weiss SRB. 2014. Adverse health effects of marijuana use. *N. Engl. J. Med.* 370(23):2219–27
129. Wang G, Dixon RA. 2009. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *PNAS* 106(24):9914–19
130. Wang G, Tian L, Aziz N, Broun P, Dai X, et al. 2008. Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol.* 148(3):1254–66
131. Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, et al. 2015. Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytol.* 208(4):1241–50
132. Welling MT, Liu L, Raymond CA, Ansari O, King GJ. 2018. Developmental plasticity of the major alkyl cannabinoid chemotypes in a diverse *Cannabis* genetic resource collection. *Front. Plant Sci.* 9:1510
133. White KH, Vergara D, Keepers KG, Kane NC. 2016. The complete mitochondrial genome for *Cannabis sativa*. *Mitochondrial DNA B* 1(1):715–16
134. Winge O. 1923. On sex chromosomes, sex determination and preponderance of female in some dioecious plants. *C. R. Trav. Lab. Carlsb.* 15(5):1–25
135. Xu H, Zhang F, Liu B, Huhman DV, Sumner LW, et al. 2013. Characterization of the formation of branched short-chain fatty acid: CoAs for bitter acid biosynthesis in hop glandular trichomes. *Mol. Plant* 6(4):1301–17
136. Yan D, Wong YF, Shellie RA, Marriott PJ, Whittock SP, Koutoulis A. 2019. Assessment of the phytochemical profiles of novel hop (*Humulus lupulus* L.) cultivars: a potential route to beer crafting. *Food Chem.* 275:15–23
137. Yang M-Q, van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. 2013. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* 62(3):473–85
138. Yang X, Matsui T, Kodama T, Mori T, Zhou X, et al. 2016. Structural basis for olivetolic acid formation by a polyketide cyclase from *Cannabis sativa*. *FEBS J.* 283(6):1088–106
139. Zager JJ, Lange I, Srividya N, Smith A, Lange BM. 2019. Gene networks underlying cannabinoid and terpenoid accumulation in cannabis. *Plant Physiol.* 180(4):1877–97
140. Zanoli P, Zavatti M. 2008. Pharmacognostic and pharmacological profile of *Humulus lupulus* L. *J. Ethnopharmacol.* 116(3):383–96
141. Zendulka O, Dovrtělová G, Nosková K, Turjap M, Šulcová A, et al. 2016. Cannabinoids and cytochrome P450 interactions. *Curr. Drug Metab.* 17(3):206–26
142. Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. 2005. Biogeography and divergence times in the mulberry family (Moraceae). *Mol. Phylogenet. Evol.* 37(2):402–16
143. Zhang D, Easterling KA, Pitra NJ, Coles MC, Buckler ES, et al. 2017. Non-Mendelian single-nucleotide polymorphism inheritance and atypical meiotic configurations are prevalent in hop. *Plant Genome* 10(3):1–14
144. Zhang H, Jin J, Moore MJ, Yi T, Li D. 2018. Plastome characteristics of Cannabaceae. *Plant Divers.* 40(3):127–37
145. Zhang Q, Chen X, Guo H, Trindade LM, Salentijn EMJ, et al. 2018. Latitudinal adaptation and genetic insights into the origins of *Cannabis sativa* L. *Front. Plant Sci.* 9:1876
146. Zirpel B, Kayser O, Stehle F. 2018. Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *J. Biotechnol.* 284:17–26
147. Zonneveld BJM, Leitch IJ, Bennett MD. 2005. First nuclear DNA amounts in more than 300 angiosperms. *Ann. Bot.* 96(2):229–44