



# Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L.

Jacob A. Toth<sup>1</sup>  | George M. Stack<sup>1</sup> | Ali R. Cala<sup>2</sup> | Craig H. Carlson<sup>1</sup> |  
 Rebecca L. Wilk<sup>1</sup> | Jamie L. Crawford<sup>3</sup> | Donald R. Viands<sup>3</sup> | Glenn Philippe<sup>4</sup> |  
 Christine D. Smart<sup>2</sup> | Jocelyn K. C. Rose<sup>4</sup> | Lawrence B. Smart<sup>1</sup> 

<sup>1</sup>Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, USA

<sup>2</sup>Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY, USA

<sup>3</sup>Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

<sup>4</sup>Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

## Correspondence

Lawrence B. Smart, Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, NY 14456, USA.

Email: lbs33@cornell.edu

## Funding information

Empire State Development Corporation, Grant/Award Number: AC477 and AC483

## Abstract

Hemp (*Cannabis sativa* L.) is an emerging dioecious crop grown primarily for grain, fiber, and cannabinoids. There is good evidence for medicinal benefits of the most abundant cannabinoid in hemp, cannabidiol (CBD). For CBD production, female plants producing CBD but not tetrahydrocannabinol (THC) are desired. We developed and validated high-throughput PACE (PCR Allele Competitive Extension) assays for *C. sativa* plant sex and cannabinoid chemotype. The sex assay was validated across a wide range of germplasm and resolved male plants from female and monoecious plants. The cannabinoid chemotype assay revealed segregation in hemp populations, and resolved plants producing predominantly THC, predominantly CBD, and roughly equal amounts of THC and CBD. Cultivar populations that were thought to be stabilized for CBD production were found to be segregating phenotypically and genotypically. Many plants predominantly producing CBD accumulated more than the current US legal limit of 0.3% THC by dry weight. These assays and data provide potentially useful tools for breeding and early selection of hemp.

## KEYWORDS

cannabinoids, cannabis, CBD, molecular marker, PACE, sex, THC

## 1 | INTRODUCTION

Hemp (*Cannabis sativa* L.) is a multiuse crop grown primarily for grain, fiber, and cannabinoids. In recent years, there has been a resurgence in the study of hemp in the United States and across the world in fields including genomics (Grassa et al., 2018; Lavery et al., 2019), agronomics (Campbell, Berrada, Hudalla, Amaducci, & McKay, 2019), and novel end-uses (Turner, Sloan, & Currell, 2019; Wang & Xiong, 2019). The market for hemp (defined as <0.3%

tetrahydrocannabinol; THC by dry weight) and marijuana (>0.3% THC), is expected to surpass \$26 billion in the US by 2025 (Global Cannabis Report: 2019 Industry Outlook, 2019). The major market for *C. sativa* is as a source of cannabinoids, the two most abundant of which are THC and cannabidiol (CBD). There is growing evidence of the different ways that CBD and THC interact with human endocannabinoid signaling pathways (Citti, Braghiroli, Vandelli, & Cannazza, 2018; Iseger & Bossong, 2015). THC is a psychoactive compound and is currently listed

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *GCB Bioenergy* Published by John Wiley & Sons Ltd

as a Schedule 1 controlled substance in the United States, <https://www.dea.diversion.usdoj.gov/schedules/index.html>. CBD is now formulated as an approved prescription medicine (DEA schedule 5) marketed as Epidiolex® by Greenwich Biosciences.

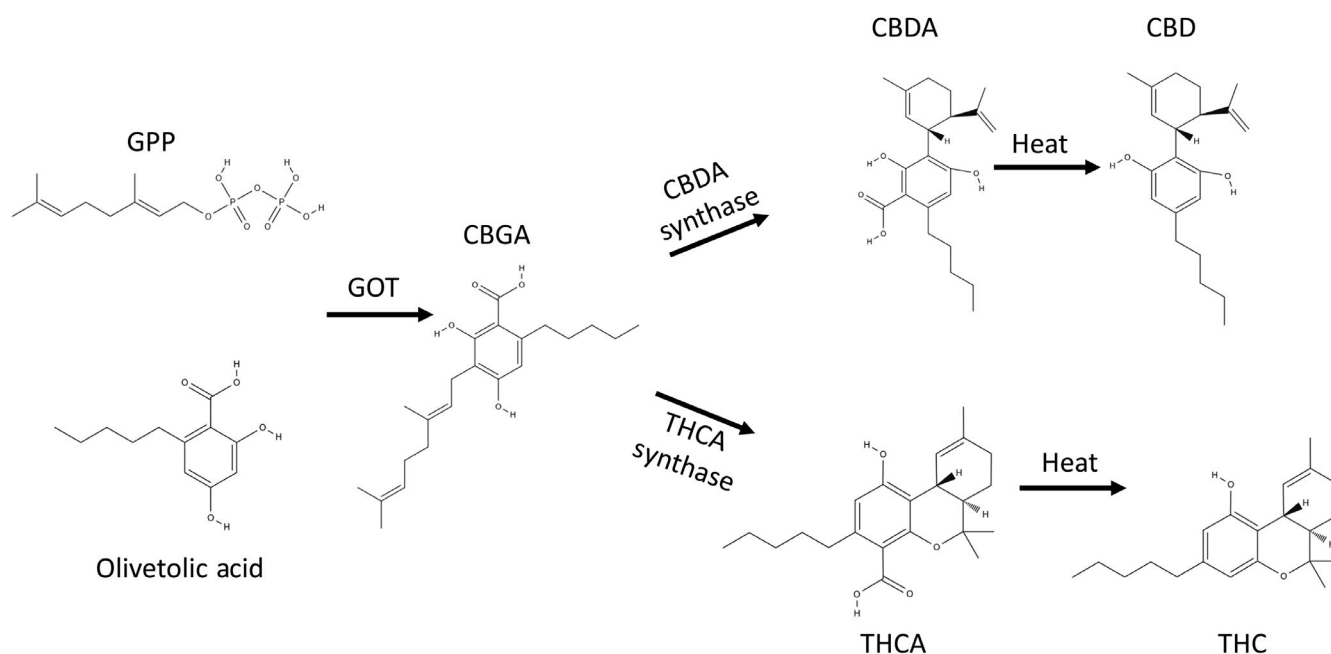
Plant sex and cannabinoid chemotype are essentially qualitative traits that are important for hemp producers and breeders. Maximal production of CBD occurs in unpollinated female hemp plants. To grow and develop legally compliant hemp cultivars (<0.3% THC by dry weight), genetic propensity for THC production must be known. However, sex and cannabinoid chemotype are difficult to phenotype in young plants. Until the onset of flowering, male and female plants are phenotypically indistinguishable, and immature plants produce relatively small quantities of cannabinoids. Cannabinoid chemotype of immature plants may also not reflect the cannabinoid profile of mature plants (de Meijer, Hammond, & Micheler, 2009; Pacifico, Miselli, Carboni, Moschella, & Mandolino, 2008). Molecular markers can address these challenges, since DNA from very young plants can be used in reliable genotype assays.

The stalked capitate trichomes of unpollinated female inflorescences have the highest concentration of cannabinoids (Livingston et al., 2020; Mahlberg & Kim, 2004). CBDA and THCA, the acidic precursors to CBD and THC (also referred to as  $\Delta^9$ -THC), are produced from cannabigerolic acid (CBGA) by CBDA synthase (CBDAS) and THCA synthase (THCAS), respectively. CBGA is produced from the condensation of olivetolic acid and geranyl pyrophosphate (GPP) by the enzyme geranyl-diphosphate:olivetolate geranyltransferase (also known as aromatic prenyltransferase, AP, or GOT;

Fellermeier, Eisenreich, Bacher, & Zenk, 2001; Fellermeier & Zenk, 1998). CBD and THC are generally derived from their corresponding acids through nonenzymatic decarboxylation, enhanced by heat and light (Figure 1; Hanuš, Meyer, Muñoz, Tagliatalata-Scafati, & Appendino, 2016; Smith & Vaughan, 1977). The decarboxylated forms (CBD and THC) are biologically active for medicinal or recreational use, while the acidic precursors do not share the same activity (Citti et al., 2018).

The genetic structure of CBDAS and THCAS have been recently elucidated (Grassa et al., 2018; Lavery et al., 2019). While the genes are highly alike, sharing 84% amino acid identity (Onofri, de Meijer, & Mandolino, 2015), they are not allelic or at equivalent loci. However, chromosomal scaffolds containing these genes are physically linked in repulsion and not highly homologous, leading to low recombination (Lavery et al., 2019). Consequently, cannabinoid chemotype inheritance can largely be modeled as monogenic, with plants producing predominantly THC (chemotype 1,  $B_T/B_T$ ), about equal amounts of CBD and THC (chemotype 2,  $B_T/B_D$ ), or predominantly CBD (chemotype 3,  $B_D/B_D$ ; de Meijer et al., 2003; Mandolino, Bagatta, Carboni, Ranalli, & de Meijer, 2003; Small & Beckstead, 1973).

*Cannabis sativa* is usually dioecious, having male and female flowers produced on separate plants. Male plants have heteromorphic X and Y sex chromosomes, while female and monoecious plants have two X chromosomes (Divashuk, Alexandrov, Razumova, Kirov, & Karlov, 2014; Razumova, Alexandrov, Divashuk, Sukhorada, & Karlov, 2016). Genetic sex determination is believed to function



**FIGURE 1** Biosynthesis of tetrahydrocannabinol (THC) and cannabidiol (CBD). GOT, geranyl-diphosphate:olivetolate geranyltransferase; GPP, geranyl pyrophosphate

through the X: autosome ratio (Ainsworth, 2000; Vyskot & Hobza, 2004), although this mechanism is not fully understood (Divashuk et al., 2014; Ming, Bendahmane, & Renner, 2011). Despite having well-defined heteromorphic sex chromosomes (Sakamoto, Akiyama, Fukui, Kamada, & Satoh, 1998), environment can play a large role in sex determination in *Cannabis sativa* (Schaffner, 1921). Factors such as altered hormones (Lubell & Brand, 2018; Ram & Jaiswal, 1970), daylength (Schaffner, 1921), and autosomal genes (Faux, Berhin, Dauguet, & Bertin, 2014) have been shown to influence sex expression. Through manipulation of hormones, it is possible to create all-female progeny using pollen produced by a female (XX) plant induced to produce male flowers and pollen (Lubell & Brand, 2018). From a commercial perspective, these so-called 'feminized' seeds are generally more expensive to produce than normal dioecious seeds due to the additional work required for their production and the market demand for all-female seed lots.

The absence of a Y chromosome does not appear to be sufficient to ensure a total lack of production of male flowers (Faux et al., 2014; Menzel, 1964; Razumova et al., 2016). Some *Cannabis* plants are monoecious, producing both male and female flowers on the same plant (Menzel, 1964). While monoecious plants are commonly referred to as hermaphrodites, the botanical definition of hermaphrodite requires staminate and carpellate parts on the same flower (Lebel-Hardenack & Grant, 1997), a phenomenon rarely seen in *C. sativa*.

There are several published marker assays for sex in *C. sativa*, but to date none are sufficiently high throughput to be used efficiently in a breeding program. For sex determination, the male-associated sequences MADC1, MADC2, MADC3, MADC4, MADC5, and MADC6 have been published (Mandolino, Carboni, Forapani, Faeti, & Ranalli, 1999; Sakamoto et al., 2005; Sakamoto, Shimomura, Komeda, Kamada, & Satoh, 1995; Törjék et al., 2002). MADC1 is a hybridization-based probe, which is not tractable for either breeders or producers (Sakamoto et al., 1995). MADC3 and MADC4 are Random amplified polymorphic DNA (RAPD) markers, which have issues with reproducibility and interpretation (Sakamoto et al., 2005). MADC5 and MADC6 were reported to be nondiagnostic and are gel-based, which significantly reduces throughput (Törjék et al., 2002). MADC2 is a marker that had been shown to be diagnostic, but is also gel-based (Divashuk et al., 2014; Faux et al., 2014; Mandolino et al., 1999). Male-associated amplified fragment length polymorphism (AFLP) markers have also been reported (Flachowsky, Schumann, Weber, & Peil, 2001; Peil, Flachowsky, Schumann, & Weber, 2003). Another report described female-specific markers, but given the XY sex-determination system and low number of samples tested, this was likely a false positive (Shao, Song, & Clarke, 2003).

Multiple marker assays to determine cannabinoid chemotype have been described involving a range of assay technologies (Borna, Salami, & Shokrpour, 2017; Kojoma, Seki, Yoshida, & Muranaka, 2006; Pacifico et al., 2006; Rotherham & Harbison, 2011). However, these published assays are low throughput or require expensive instrumentation. There have been extensive studies on sequence and copy number variation of THCAS and CBDAS, but sequence–function relationships are not always clear (McKernan et al., 2015; Onofri et al., 2015). An inactive form of CBDAS appears to be conserved among chemotype 1 plants (Van Bakel et al., 2011; Weiblen et al., 2015); however, the published assay conditions poorly resolve chemotype 2 plants, limiting the diagnostic use of this protocol (McKernan et al., 2015).

PACE (PCR Allele Competitive Extension, 3CR Bioscience Ltd) is a high-throughput, fluorescence-based marker system that can interrogate SNPs, indels, or other polymorphic DNA features. Fluorescence-based marker systems such as PACE, KASP, TaqMan, and the recently developed RhAMP are estimated to be 45 times faster than gel-based systems (Rasheed et al., 2016; Toth, Pandurangan, Burt, Mitchell Fetch, & Kumar, 2018). PACE has the further advantage of lower cost while retaining simple codominance, unlike loop-mediated isothermal amplification (LAMP; Notomi et al., 2000) or direct sequencing (McKernan et al., 2015; Weiblen et al., 2015). PACE assays, unlike some other marker systems, require a fluorescent plate reader or qPCR system to score, limiting field-based testing. However, the ease and speed of PACE assays is well suited for breeding or advanced production systems. Here we used publicly available sequence information to develop reliable, high-throughput PACE assays that are highly predictive of sex and cannabinoid chemotype phenotypes in *C. sativa*.

## 2 | MATERIALS AND METHODS

For sex testing, dioecious seeds from CBD cultivars of *C. sativa* were started in a greenhouse in plugs of soilless mix in 2019 and cultivated under a 18L:6D light regime. DNA was extracted from leaves harvested from 2-week old plants. Genotyped female plants were planted in outdoor field trials, while males were transplanted to two-gallon pots and kept in the greenhouse. Plant sex was noted at the onset of flowering.

For cannabinoid chemotype marker testing, cannabinoid data and tissue from the 2018 Cornell CBD Hemp Cultivar Trials was used. Hemp seeds were obtained from multiple sources (Table S3). Plants for CBD production were started in the greenhouse under a 18L:6D light regime with males or monoecious plants removed based on phenotype. The trials were located on Cornell University

farms in upstate New York: one at Bluegrass Lane Turf and Ornamental Research Farm and the other at Cornell AgriTech Gates West Farm. Late season rainfall lead to saturated field conditions in the Geneva location during flowering.

The top 10 cm of mature female plants were harvested by hand at maturity and dried in a greenhouse. The inflorescence was then milled using a Magic Bullet food grinder (Homeland Housewares) and stored at 4°C until analysis. For each sample, 50 mg of dried, milled tissue was mixed with 1.5 ml ethanol by high-speed shaking at room temperature with a TissueLyser (Qiagen), and filtered through a SINGLE STEP PTFE Filter Vial (Thomson). The resultant liquid was directly subjected to HPLC analysis (Dionex UltiMate 3000; Thermo Fisher) with biphenyl-4-carboxylic acid (BCPA) as an internal standard, using a Phenomenex Kinetex 2.6 µm Polar 100 Å column 150 × 4.6 mm heated at 35°C. Samples were injected and eluted at 1.2 ml/min over a 6 min gradient, from 65% acetonitrile, 0.1% formic acid, to 80% acetonitrile, 0.1% formic acid, followed by a 4 min isocratic step. Absorbance was measured at 214 nm. The following standards were used as calibrants: THCA, Δ<sup>9</sup>-THC, CBDA,

CBD, cannabichromenic acid (CBCA), cannabichromene (CBC), CBGA, CBG, cannabinol (CBN), and Δ<sup>8</sup>-THC (Sigma Aldrich).

DNA was isolated using a high-throughput modified CTAB method utilizing PALL DNA binding plates (Doyle & Doyle, 1987). PACE reactions were run according to the manufacturer's (3CR Bioscience Ltd) instructions with five extra final cycles on a Bio-Rad C1000 Touch thermocycler. A Bio-Rad CFX96 qPCR machine was used as a fluorescent plate reader, and the data were analyzed using Bio-Rad CFX Maestro software. The primers (Table 1) were designed so that samples from plants with a Y chromosome and THC-dominant plants produce HEX fluorescence.

Multiple linear regression analysis was performed in RStudio version 1.1.463 running R version 3.5.1 (R Core Team, 2018). Cannabinoid chemotype allele score was numerically coded as [-1,0,1] while all other variables were coded as factors. Total potential CBD and THC were calculated by summing the concentration of the decarboxylated form with the concentration of CBDA or THCA multiplied by 0.877.

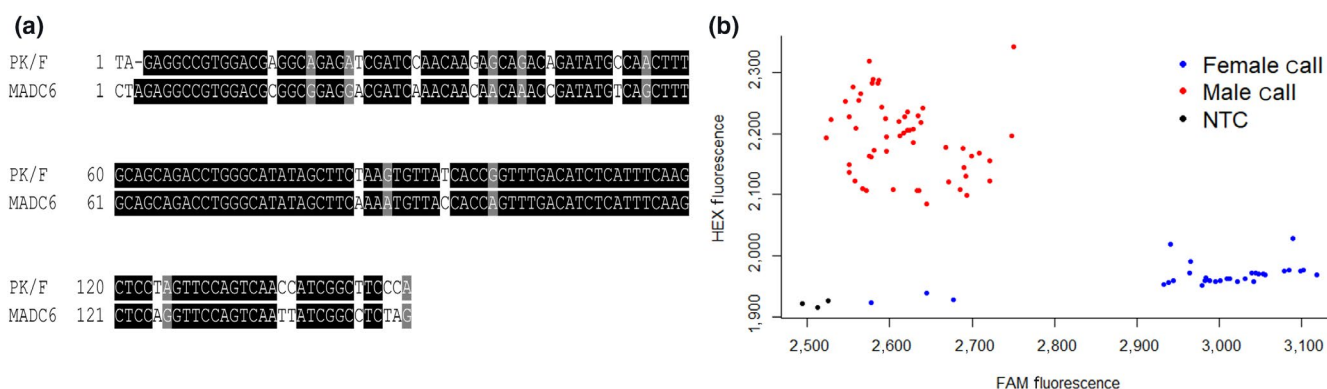
**TABLE 1** Primer sequences used

Name	Sequence	Notes
CSP1-FAM	GAAGGTGACCAAGTTCATGCTA GCTTGAAATGAGATGTCAAACC	Female
CSP1-HEX	GAAGGTCGGAGTCAACGGATTG AGCTTGAAATGAGATGTCAAAC	Male
CSP1-COMMON	GCAGCAGACCTGGGCATATAG	
CCP1-FAM	GAAGGTGACCAAGTTCATGCTAT TAGACTGGTTGCTGTCCCAA	B <sub>D</sub>
CCP1-HEX	GAAGGTCGGAGTCAACGGATTAT TAGACTGGTTGCTGTCCCAAC	B <sub>T</sub>
CCP1-COMMON	ACTTGACAAGCTCATGTATCTCCA	

## 3 | RESULTS

### 3.1 | Sex assay development

A novel high-throughput assay for the Y chromosome was developed based on the previously identified male-specific MADC6 sequence (Genbank AF364955.1). To develop the assay, the MADC6 sequence was compared to the FINOLA (Genbank GCA\_003417725.1) and Purple Kush (Genbank GCA\_000230575.1) genomes (Van Bakel et al., 2011) using BLAT (BLAST-Like Alignment Tool) on the *C. sativa* Genome Browser Gateway (UCSC Genome Browser, University of California). A PACE assay, named CSP-1, was designed based on a SNP between the sequences (Figure 2).



**FIGURE 2** CSP-1 development. (a) Sequence comparison of closest BLAT result of the Purple Kush/FINOLA (PK/F) genomes and MADC6. (b) Representative chromatogram of CSP-1. Some females clustered with the No Template Control (NTC)

### 3.2 | Sex assay validation

The CSP-1 assay was used to test a total of 2,170 plants of 14 cultivars. In all but one population the genetic male:female ratio fit the expected 1:1 model (Chi-square  $p > .05$ , Table S1). The individuals genetically scored as females were planted in field trials and the individuals genetically scored as males were discarded or moved to greenhouse conditions. Approximately 98% of the screened genetic females were phenotypically female. Approximately 1% of the screened genetic females were monoecious, including individuals from three cultivars (Table S1). Two screened plants were phenotypically male, and when retested, shown to be originally mis-called. About 270 plants genetically scored as male from four hemp cultivars were allowed to flower in greenhouse conditions, and all were phenotypically male (Table 2).

Monoecious plants (20 plants each of the cultivars ‘Anka’, ‘Hlesia’, and ‘USO-31’) were also examined with this assay, and all monoecious plants were scored as female.

### 3.3 | Cannabinoid chemotype assay development

A PACE assay to predict cannabinoid chemotype was generated through comparison of marijuana-type CBDAS ( $B_T$ ) and hemp-type CBDAS ( $B_D$ ), which were previously found to correspond to high-THC and high-CBD chemotypes, respectively (Figure 3; Weiblen et al., 2015). While THCAS

**TABLE 2** Validation of the *Cannabis sativa* sex assay CSP-1

	Lithuanian	Nebraska	RN16	RNF
Male plants	14	53	46	157
Female/ monoecious plants	21	54	47	157
Marker accuracy	100%	100%	100%	100%

Note: Lithuanian is a grain type. Nebraska is a grain/fiber type with monoecious individuals. RN16 and RNF are cannabidiol types.

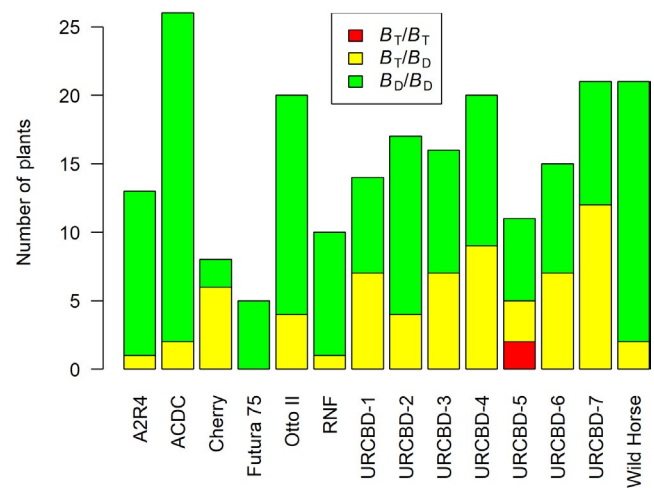
<b>THCAS (Purple Kush)</b>	ATTGCAGCATGGAAAATCAAACCTGGTTGCTGTCCCATCAAAGTCTACTATATTCAGTGT
<b>CBDAS (Finola)</b>	ATTGTAGCATGGAAAATTAGACTGGTTGCTGTCCCAA---GTCTACTATGTTTAGTGT
<b>CBDAS (Purple Kush)</b>	ATTGTAGCATGGAAAATTAGACTGGTTGCTGTCCCAACAAAGTCTACTATGTTTAGTGT
<b>THCAS (Purple Kush)</b>	AAAAAGAACATGGAGATACATGGGCTTGTCAGTTATTTAAACAAATGGCAAAAATATTGCT
<b>CBDAS (Finola)</b>	AAAAAGATCATGGAGATACATGAGCTTGTCAGTTAGTTAAACAAATGGCAAAAATATTGCT
<b>CBDAS (Purple Kush)</b>	AAAAAGATCATGGAGATACATGAGCTTGTCAGTTAGTTAAACAAATGGCAAAAATATTGCT

**FIGURE 3** Alignment of CCP-1 primers to cannabinoid synthase genes. The blue sequence is specific to  $B_D$  alleles, while the red sequence is specific to  $B_T$  alleles. The sequence in purple is common to the Purple Kush THCAS (scaffold 19,603), the FINOLA CBDAS (scaffold 14,546,436), and the Purple Kush CBDAS (scaffold 39,155, Van Bakel et al., 2011)

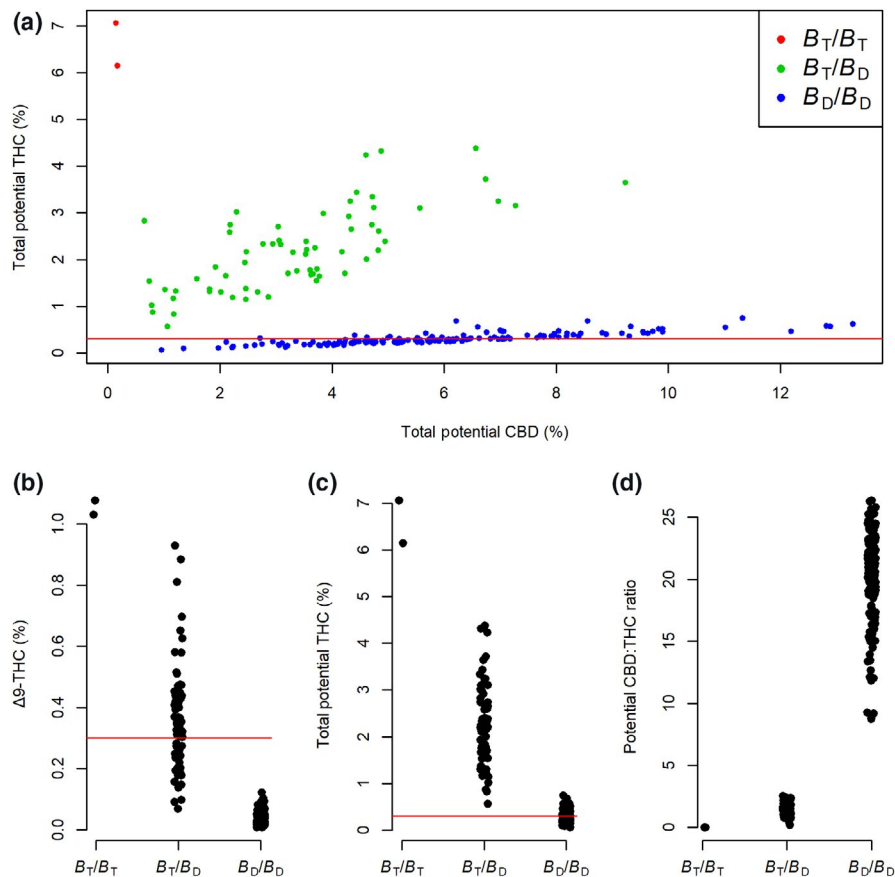
and CBDAS are not the same gene, their close linkage in repulsion suggests that they are inherited monogenically as a cannabinoid chemotype locus (de Meijer et al., 2003; Laverty et al., 2019). This assay was named CCP-1.

### 3.4 | Cannabinoid chemotype assay validation

Two hundred and seventeen plants from 14 hemp cultivars grown for CBD in two locations were tested with the cannabinoid chemotype (CCP-1) assay and phenotyped for cannabinoids using HPLC. Of these, two were homozygous for the marijuana-type allele ( $B_T/B_T$ ), 65 were heterozygous ( $B_T/B_D$ ), and 150 were homozygous for the hemp-type allele ( $B_D/B_D$ ). Most cultivar populations were segregating for this allele, which was consistent with the phenotypic data (Figure 4). The genotypic data corresponded to three apparent chemotypes, in terms of total potential CBD and THC (Figure 5a, ANOVA  $p < 1e-4$ ). This indicates that the CCP-1 assay identifies previously established  $B_T$  and  $B_D$  alleles (de Meijer et al., 2003).



**FIGURE 4** Distribution of cannabinoid chemotype alleles across cultivar populations. Additional allele frequency data can be found in Table S3



**FIGURE 5** Genotype to phenotype relationships. (a) Total potential tetrahydrocannabinol (THC) and cannabidiol (CBD) concentration (% dry mass) in individual plants for which B locus genotype was also determined. The red line indicates 0.3% total potential THC (b)  $\Delta^9$ -THC concentration by genotype. The red line indicates 0.3% dry weight  $\Delta^9$ -THC. (c) Total potential THC concentration by genotype. The red line indicates 0.3% dry weight total potential THC. (d) Total potential CBD:THC concentration ratio. All means differ (ANOVA  $p < 1e-4$ ). Tabular data can be found in Table S2

**TABLE 3** Linear regression  $R^2$  values of various models predicting cannabinoid data

	Trial	Marker Coding	Cultivar	Potential CBD:THC	$\Delta^9$ -THC (%)	CBD (%)	Potential THC (%)	Potential CBD (%)	Total Potential Cannabinoids (%)
Model 1	+	+	+	0.89	0.77	0.21	0.81	0.38	0.19
Model 2	+	+	-	0.86	0.74	0.03	0.78	0.25	0.01
Model 3	+	-	+	0.26	0.23	0.19	0.20	0.18	0.19
Model 4	+	-	-	0.01	0.01	0.03	0.00	0.01	0.01
Model 5	-	+	+	0.89	0.76	0.18	0.81	0.38	0.19
Model 6	-	+	-	0.86	0.73	0.00	0.77	0.25	0.01
Model 7	-	-	+	0.25	0.22	0.16	0.20	0.17	0.18

Note: '+' indicates the variable was included in the model and '-' indicates that the variable was not included in the model. Light gray cells are  $p < .01$ . Dark gray cells are  $p < 1e-4$ .

Abbreviations: CBD, cannabidiol; THC, tetrahydrocannabinol.

Mean  $\Delta^9$ -THC and total potential THC differed across genotypic groups (ANOVA  $p < 1e-4$ ). Within the genotypic groups there was a strong correlation between total potential CBD and total potential THC concentrations (Figure 5a;  $B_T/B_D$   $r = .72$   $p < 1e-4$ ,  $B_D/B_D$   $r = .86$   $p < 1e-4$ ).

The  $\Delta^9$ -THC concentration for  $B_D/B_D$  samples was consistently  $<0.3\%$  (dry weight), while 35% of the  $B_T/B_D$  samples had a  $\Delta^9$ -THC concentration  $<0.3\%$  (Figure 5b). Only 39% of the  $B_D/B_D$  samples had total potential THC

concentration  $<0.3\%$  (Figure 5b). The mean ratio of total potential CBD:THC was 0.02, 1.6, and 20.3 for  $B_T/B_T$ ,  $B_T/B_D$ , and  $B_D/B_D$  lines, respectively (Figure 5d; Table S2).

A total of 1,420 plants from 47 cultivars were tested with the CCP-1 assay (Table S3). These cultivars were from multiple sources and grown for CBD, grain, or grain/fiber. The THC-associated  $B_T$  allele frequency varied by cultivar, from 0% in some clones grown for CBD, up to 98% in a Chinese grain/fiber cultivar (Table S3).

### 3.5 | Other factors affecting cannabinoid production

Genotypic group, cultivar, and trial were used to create models explaining the potential CBD:THC concentration ratio as well as the concentrations of  $\Delta^9$ -THC, CBD, potential THC, potential CBD, and total potential cannabinoids (Table 3). Total potential cannabinoids included CBD, THC, CBC, CBG, and their corresponding acids. Genotypic group explained the most variance in the CBD:THC ratio, as well as  $\Delta^9$ -THC and potential THC levels, but not total potential cannabinoids. Cultivar was an important factor in total potential cannabinoid abundance, as well as the concentration of CBD and  $\Delta^9$ -THC. The cultivar explained ~3% of the variation in the potential CBD:THC ratio when the genotypic group was taken into consideration, and the trial was a poor predictor of all measured variables.

## 4 | DISCUSSION

### 4.1 | CSP-1 sex assay

The CSP-1 assay was found to be a reliable predictor of plant sex. There was a 50:50 segregation ratio in nearly all tested dioecious populations including CBD types, grain types, and grain/fiber types. As expected, monoecious plants were scored as female (Divashuk et al., 2014). Given these data, it is likely that the CSP-1 assay distinguishes a nonrecombining part of the Y chromosome. This is somewhat surprising, given that the original MADC6 marker assay was not a completely accurate predictor of plant sex, with 2/75 reported recombinants (Törjék et al., 2002). It is possible that this was due to PCR failure, monoecious plants with a quantitatively male phenotype, or that the CSP-1 assay in fact examines a different DNA sequence than the original MADC6 assay. Recent *C. sativa* whole-genome sequencing (Lavery et al., 2019) showed six unassembled scaffolds in the male genome with >99% identity to the MADC6 sequence in *C. sativa*, possibly contributing to the empirical success of this assay. As MADC6 shows some sequence relationship to retrotransposons, it is possible that the sequence was subject to copy number increase in the recent past (Sakamoto, Ohmido, Fukui, Kamada, & Satoh, 2000; Törjék et al., 2002). It is well known that in the development of sex-determining regions of plants, an absence of recombination between male- and female-specific sequences can lead to an expansion of retrotransposon copy number repeats, which are not lost through a Muller's ratchet-type mechanism (Sakamoto et al., 2000; Vyskot & Hobza, 2004).

### 4.2 | CCP-1 cannabinoid chemotype assay

The CCP-1 cannabinoid chemotype marker assay detected three genotypic groups that corresponded to three

phenotypic groups, reflecting previously described chemotypes (de Meijer et al., 2003). Since the CCP-1 assay examines CBDAS only and CBDAS and THCAS are not allelic, it is possible that a plant with a recombination between CBDAS and THCAS would not be scored correctly. However, we did not detect this in any of our samples, and the tight linkage in repulsion between CBDAS and THCAS is well established (de Meijer et al., 2003; Grassa et al., 2018; Weiblen et al., 2015).

The mean  $\Delta^9$ -THC and total potential THC concentrations as percent dry matter were significantly different in each chemotype. If  $\Delta^9$ -THC concentration alone as assayed by HPLC was used as the criterion for legal compliance at the level of 0.3%, then all  $B_D/B_D$  samples and 35% of the  $B_T/B_D$  samples would be below the threshold. It is possible that past breeding material chosen for low THC was in fact heterozygous, leading to segregation in released cultivars (Table S3). If total potential THC is used as the legal criterion, then 61% of the  $B_D/B_D$  samples would be above the legal threshold of 0.3%, and therefore be noncompliant. The close correlation between potential CBD and THC concentrations in the  $B_D/B_D$  class suggests that it might be difficult to develop a cultivar that accumulates high CBD concentrations, while maintaining low total potential THC. The average potential CBD:THC ratio was about 20:1, which suggests that accumulation of greater than 6% CBD will result in rise of the total potential THC above 0.3%. A clear target for breeders developing high CBD hemp cultivars is to raise the ratio of total potential CBD:THC.

There are several lines of evidence to suggest that the concomitant increase in THC concentration with that of CBD in the  $B_D/B_D$  group is due to promiscuous activity of the active CBDAS. Despite attempts, no demonstrably active transcribed THCAS has been isolated from a confirmed chemotype 3 plant (Kojoma et al., 2006; Lavery et al., 2019; Onofri et al., 2015). Other research found a *C. sativa* plant with a  $B_D/B_D$  genotype and a catalytically inactive CBDAS that accumulates CBGA and essentially no THCA, although mutations in CBDAS and a putative THCAS would also explain this observation (Onofri et al., 2015). A purported active THCAS from CBD-dominant fiber-type hemp was later shown to be a cannabichromenic acid synthase (CBCAS) (Kojoma et al., 2006; Lavery et al., 2019). Lastly, in vitro expression of wild-type CBDAS leads to production of CBDA:THCA in a ratio very close to 20:1 at optimal pH (Zirpel, Kayser, & Stehle, 2018). Future breeding efforts should be informed by this promiscuous activity.

Some studies have found sequence and copy number variation in CBDAS and THCAS and correlated them to differences in cannabinoid production (McKernan et al., 2015; Weiblen et al., 2015). These differences were not assayed here, but could have conceivably contributed to some of the variation within groups.

### 4.3 | Other factors affecting cannabinoid production

Neither trial nor cultivar per se explained variations in the CBD:THC ratio. Trial did not appear to have much of an effect on any measured parameters, despite flooding stress in one trial location. The explanatory power of models 3 and 7, including cultivar but not marker coding, are likely due to differences in allele frequency in each cultivar population. While cultivar per se poorly explained CBD:THC ratio, cultivar was the best predictor of total potential cannabinoid concentration. It has previously been demonstrated that the factors affecting cannabinoid chemotype are not linked to total cannabinoid content (Grassa et al., 2018).

In testing 14 different cultivars, with 217 plants total across two locations, we found that none of the  $B_D/B_D$  samples had a  $\Delta^9$ -THC concentration >0.3% dry weight. However, we did find that most cultivar populations were segregating for the  $B_T$  allele. It is possible that differences in cannabinoid production ascribed to changes in environment may in fact be due to sampling of individual plants with  $B_T$  alleles. Additional studies of the influence of environment on cannabinoid production coupled with individual plant genotyping may lead to a better understanding of the regulation of cannabinoid production.

#### ACKNOWLEDGEMENTS

We are grateful to the Smart, Viands, and Rose research teams for their support, especially Lauren Carlson, Jane Petzoldt, Dawn Fishback, Ben DeMoras, and Sarah Wright. This work was supported by New York State Department of Agriculture and Markets through grants AC477 and AC483 from Empire State Development Corporation.

#### ORCID

Jacob A. Toth  <https://orcid.org/0000-0002-0454-1619>

Lawrence B. Smart  <https://orcid.org/0000-0002-7812-7736>

#### REFERENCES

- Ainsworth, C. (2000). Boys and girls come out to play: The molecular biology of dioecious plants. *Annals of Botany*, 86(2), 211–221. <https://doi.org/10.1006/anbo.2000.1201>
- Borna, T., Salami, S. A., & Shokrpour, M. (2017). High resolution melting curve analysis revealed SNPs in major cannabinoid genes associated with drug and non-drug types of cannabis. *Biotechnology & Biotechnological Equipment*, 31(4), 839–845. <https://doi.org/10.1080/13102818.2017.1333456>
- Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., & McKay, J. K. (2019). Genotype  $\times$  environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosystems, Geosciences & Environment*, 2(1). <https://doi.org/10.2134/age2018.11.0057>
- Citti, C., Braghiroli, D., Vandelli, M. A., & Cannazza, G. (2018). Pharmaceutical and biomedical analysis of cannabinoids: A critical review. *Journal of Pharmaceutical and Biomedical Analysis*, 147, 565–579. <https://doi.org/10.1016/j.jpba.2017.06.003>
- de Meijer, E. P., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. C., Ranalli, P., & Mandolino, G. J. G. (2003). The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*, 163(1), 335–346.
- De Meijer, E., Hammond, K., & Micheler, M. (2009). The inheritance of chemical phenotype in *Cannabis sativa* L. (III): Variation in cannabichromene proportion. *Euphytica*, 165(2), 293–311.
- Divashuk, M. G., Alexandrov, O. S., Razumova, O. V., Kirov, I. V., & Karlov, G. I. (2014). Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. *PLoS ONE*, 9(1), e85118. <https://doi.org/10.1371/journal.pone.0085118>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19(1), 11–15.
- 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), 183–197. <https://doi.org/10.1007/s10681-013-1023-y>
- Fellermeier, M., Eisenreich, W., Bacher, A., & Zenk, M. H. (2001). Biosynthesis of cannabinoids: Incorporation experiments with  $^{13}\text{C}$ -labeled glucoses. *European Journal of Biochemistry*, 268(6), 1596–1604. <https://doi.org/10.1046/j.1432-1327.2001.02030.x>
- Fellermeier, M., & Zenk, M. H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Letters*, 427(2), 283–285. [https://doi.org/10.1016/S0014-5793\(98\)00450-5](https://doi.org/10.1016/S0014-5793(98)00450-5)
- Flachowsky, H., Schumann, E., Weber, W., & Peil, A. (2001). Application of AFLP for the detection of sex-specific markers in hemp. *Plant Breeding*, 120(4), 305–309. <https://doi.org/10.1046/j.1439-0523.2001.00620.x>
- Global Cannabis Report: 2019 Industry Outlook. (2019). New Frontier Data. Washington, DC. Retrieved from <https://newfrontierdata.com/product/global-cannabis-2019/>
- Grassa, C. J., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., ... Weiblen, G. D. (2018). A complete cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *BioRxiv*, 458083.
- Hanuš, L. O., Meyer, S. M., Muñoz, E., Tagliatalata-Scafati, O., & Appendino, G. (2016). Phytocannabinoids: A unified critical inventory. *Natural Product Reports*, 33(12), 1357–1392. <https://doi.org/10.1039/C6NP00074F>
- Iseger, T. A., & Bossong, M. G. (2015). A systematic review of the antipsychotic properties of cannabidiol in humans. *Schizophrenia Research*, 162(1–3), 153–161. <https://doi.org/10.1016/j.schres.2015.01.033>
- Kojoma, M., Seki, H., Yoshida, S., & Muranaka, T. (2006). DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in “drug-type” and “fiber-type” *Cannabis sativa* L. *Forensic Science International*, 159(2–3), 132–140. <https://doi.org/10.1016/j.forsciint.2005.07.005>
- Lavery, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., ... Page, J. E. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Research*, 29(1), 146–156.
- Lebel-Hardenack, S., & Grant, S. R. (1997). Genetics of sex determination in flowering plants. *Trends in Plant Science*, 2(4), 130–136. [https://doi.org/10.1016/S1360-1385\(97\)01012-1](https://doi.org/10.1016/S1360-1385(97)01012-1)



- Livingston, S., Quilichini, T., Booth, J., Wong, D., Rensing, K., Laflamme-Yonkman, J., ... Samuels, A. (2020). Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *The Plant Journal*, *101*, 37–56. <https://doi.org/10.1111/tbj.14516>
- Lubell, J. D., & Brand, M. H. (2018). Foliar sprays of silver thiosulfate produce male flowers on female hemp plants. *HortTechnology*, *28*(6), 743–747. <https://doi.org/10.21273/HORTTECH04188-18>
- Mahlberg, P. G., & Kim, E. S. (2004). Accumulation of cannabinoids in glandular trichomes of *Cannabis* (Cannabaceae). *Journal of Industrial Hemp*, *9*(1), 15–36.
- Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., & de Meijer, E. (2003). Qualitative and quantitative aspects of the inheritance of chemical phenotype in *Cannabis*. *Journal of Industrial Hemp*, *8*(2), 51–72.
- 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.). *Theoretical and Applied Genetics*, *98*(1), 86–92. <https://doi.org/10.1007/s001220051043>
- McKernan, K. J., Helbert, Y., Tadigotla, V., McLaughlin, S., Spangler, J., Zhang, L., & Smith, D. (2015). Single molecule sequencing of THCA synthase reveals copy number variation in modern drug-type *Cannabis sativa* L., *BioRxiv*, 028654.
- Menzel, M. Y. (1964). Meiotic chromosomes of monoecious Kentucky hemp (*Cannabis sativa*). *Bulletin of the Torrey Botanical Club*, *91*, 193–205. <https://doi.org/10.2307/2483524>
- Ming, R., Bendahmane, A., & Renner, S. S. (2011). Sex chromosomes in land plants. *Annual Review of Plant Biology*, *62*, 485–514. <https://doi.org/10.1146/annurev-arplant-042110-103914>
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, *28*(12), e63–e63. <https://doi.org/10.1093/nar/28.12.e63>
- Onofri, C., de Meijer, E. P., & Mandolino, G. (2015). Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry*, *116*, 57–68. <https://doi.org/10.1016/j.phytochem.2015.03.006>
- Pacifico, D., Miselli, F., Carboni, A., Moschella, A., & Mandolino, G. (2008). Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica*, *160*(2), 231–240. <https://doi.org/10.1007/s10681-007-9543-y>
- Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., & Mandolino, G. (2006). Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. *Molecular Breeding*, *17*(3), 257–268. <https://doi.org/10.1007/s11032-005-5681-x>
- Peil, A., Flachowsky, H., Schumann, E., & Weber, W. (2003). Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). *Theoretical and Applied Genetics*, *107*(1), 102–109. <https://doi.org/10.1007/s00122-003-1212-5>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Austria: R Foundation for Statistical Computing Vienna. Retrieved from <https://www.R-project.org/>
- Ram, H. M., & Jaiswal, V. S. (1970). Induction of female flowers on male plants of *Cannabis sativa* L. by 2-chloroethanephosphonic acid. *Cellular and Molecular Life Sciences*, *26*(2), 214–216. <https://doi.org/10.1007/BF01895593>
- Rasheed, A., Wen, W., Gao, F., Zhai, S., Jin, H., Liu, J., ... He, Z. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theoretical and Applied Genetics*, *129*(10), 1843–1860. <https://doi.org/10.1007/s00122-016-2743-x>
- Razumova, O. V., Alexandrov, O. S., Divashuk, M. G., Sukhorada, T. I., & Karlov, G. I. (2016). Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution. *Protoplasma*, *253*(3), 895–901.
- Rotherham, D., & Harbison, S. (2011). Differentiation of drug and non-drug *Cannabis* using a single nucleotide polymorphism (SNP) assay. *Forensic Science International*, *207*(1–3), 193–197. <https://doi.org/10.1016/j.forsciint.2010.10.006>
- Sakamoto, K., Abe, T., Matsuyama, T., Yoshida, S., Ohmido, N., Fukui, K., & Satoh, S. (2005). RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. *Genome*, *48*(5), 931–936.
- Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H., & Satoh, S. (1998). Characterization; genome sizes and morphology of sex chromosomes in hemp (*Cannabis sativa* L.). *Cytologia*, *63*(4), 459–464. <https://doi.org/10.1508/cytologia.63.459>
- 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 Molecular Biology*, *44*(6), 723–732.
- Sakamoto, K., Shimomura, K., Komeda, Y., Kamada, H., & Satoh, S. (1995). A male-associated DNA sequence in a dioecious plant, *Cannabis sativa* L. *Plant and Cell Physiology*, *36*(8), 1549–1554.
- Schaffner, J. H. (1921). Influence of environment on sexual expression in hemp. *Botanical Gazette*, *71*(3), 197–219. <https://doi.org/10.1086/332818>
- Shao, H., Song, S.-J., & Clarke, R. C. (2003). Female-associated DNA polymorphisms of hemp (*Cannabis sativa* L.). *Journal of Industrial Hemp*, *8*(1), 5–9.
- Small, E., & Beckstead, H. (1973). Common cannabinoid phenotypes in 350 stocks of *Cannabis*. *Lloydia*, *36*(2), 144–165.
- Smith, R., & Vaughan, C. (1977). The decomposition of acidic and neutral cannabinoids in organic solvents. *Journal of Pharmacy and Pharmacology*, *29*(1), 286–290. <https://doi.org/10.1111/j.2042-7158.1977.tb11313.x>
- Törjék, O., Bucherna, N., Kiss, E., Homoki, H., Finta-Korpelová, Z., Bócsa, I., ... Heszky, L. E. (2002). Novel male-specific molecular markers (MADC5, MAD6) in hemp. *Euphytica*, *127*(2), 209–218.
- Toth, J., Pandurangan, S., Burt, A., Mitchell Fetch, J., & Kumar, S. (2018). Marker-assisted breeding of hexaploid spring wheat in the Canadian prairies. *Canadian Journal of Plant Science*, *99*(2), 111–127. <https://doi.org/10.1139/cjps-2018-0183>
- Turner, B. D., Sloan, S. W., & Currell, G. R. (2019). Novel remediation of per- and polyfluoroalkyl substances (PFASs) from contaminated groundwater using *Cannabis sativa* L. (hemp) protein powder. *Chemosphere*, *229*, 22–31.
- United States Department of Justice, Drug Enforcement Administration, Diversion Control Division. Retrieved from <https://www.deadiversi.on.usdoj.gov/schedules/>
- Van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., & Page, J. E. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology*, *12*(10), R102. <https://doi.org/10.1186/gb-2011-12-10-r102>
- Vyskot, B., & Hobza, R. (2004). Gender in plants: Sex chromosomes are emerging from the fog. *TRENDS in Genetics*, *20*(9), 432–438. <https://doi.org/10.1016/j.tig.2004.06.006>

- Wang, Q., & Xiong, Y. L. (2019). Processing, nutrition, and functionality of hempseed protein: A review. *Comprehensive Reviews in Food Science and Food Safety*, 18(4), 936–952. <https://doi.org/10.1111/1541-4337.12450>
- Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., Marks, M. D. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytologist*, 208(4), 1241–1250.
- Zirpel, B., Kayser, O., & Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. *Journal of Biotechnology*, 284, 17–26. <https://doi.org/10.1016/j.jbiotec.2018.07.031>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Toth JA, Stack GM, Cala AR, et al. Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy*. 2020;00:1–10. <https://doi.org/10.1111/gcbb.12667>