

Gene duplication and divergence affecting drug content in *Cannabis sativa*

George D. Weiblen¹, Jonathan P. Wenger¹, Kathleen J. Craft¹, Mahmoud A. ElSohly², Zlatko Mehmedic², Erin L. Treiber¹ and M. David Marks¹

¹Department of Plant Biology and Bell Museum, University of Minnesota, 250 Biological Science Center, 1445 Gortner Ave., St Paul, MN 55108, USA; ²National Center for Natural Product Research, Research Institute of Pharmaceutical Sciences, School of Pharmacy, University of Mississippi, University, MS 38677, USA

Summary

Author for correspondence:

George D. Weiblen

Tel: +1 612 624 3461

Email: gweiblen@umn.edu

Received: 18 May 2015

Accepted: 10 June 2015

New Phytologist (2015) **208**: 1241–1250

doi: 10.1111/nph.13562

Key words: cannabidiol (CBD), cannabinoids, domestication, hemp, marijuana, tetrahydrocannabinol (THC).

- *Cannabis sativa* is an economically important source of durable fibers, nutritious seeds, and psychoactive drugs but few economic plants are so poorly understood genetically.
- Marijuana and hemp were crossed to evaluate competing models of cannabinoid inheritance and to explain the predominance of tetrahydrocannabinolic acid (THCA) in marijuana compared with cannabidiolic acid (CBDA) in hemp. Individuals in the resulting F₂ population were assessed for differential expression of cannabinoid synthase genes and were used in linkage mapping. Genetic markers associated with divergent cannabinoid phenotypes were identified.
- Although phenotypic segregation and a major quantitative trait locus (QTL) for the THCA/CBDA ratio were consistent with a simple model of codominant alleles at a single locus, the diversity of THCA and CBDA synthase sequences observed in the mapping population, the position of enzyme coding loci on the map, and patterns of expression suggest multiple linked loci. Phylogenetic analysis further suggests a history of duplication and divergence affecting drug content.
- Marijuana is distinguished from hemp by a nonfunctional CBDA synthase that appears to have been positively selected to enhance psychoactivity. An unlinked QTL for cannabinoid quantity may also have played a role in the recent escalation of drug potency.

Introduction

Cannabis is an economically important source of durable fibers, nutritious seeds, and psychoactive drugs with a multibillion-dollar annual cost to law enforcement in the United States (Miron, 2008). The commonly cultivated forms of *Cannabis* known as marijuana and hemp are members of the same species that can be readily distinguished by the relative yield of tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA). Recent legislation permitting the cultivation of *Cannabis* for seed oil and fiber in Canada and Europe recognizes differences in cannabinoid content among varieties (Hennink, 1994) as the basis for separating hemp from marijuana (Mignoni, 1997). Although central to *Cannabis* regulatory policy, the hereditary basis of cannabinoid variation has remained uncertain (van Bakel *et al.*, 2011).

The primary agent of marijuana potency is concentrated in glandular trichomes on the bracts of pistillate inflorescences in the form of delta⁹-tetrahydrocannabinolic acid (Kim & Mahlberg, 1997). Cannabidiolic acid is similarly located and also of pharmacological interest (Page & Nagel, 2006). The decarboxylated forms of these molecules (THC and CBD) bind to endocannabinoid receptors in the human body to produce a broad array of effects that range from mild

euphoria in the case of THC to anticonvulsant properties of CBD (Bostwick, 2012). Early reports regarded CBDA as a direct precursor to THCA (Mechoulam *et al.*, 1970; Shoyama *et al.*, 1975) but more recent work showed that CBDA and THCA are alternate products of a common precursor, cannabigerolic acid (CBGA; Taura *et al.*, 1995; Fellermeier *et al.*, 2001). THCA synthase and CBDA synthase convert CBGA to the carboxylated forms of THCA and CBDA, respectively, and share 89% nucleotide sequence identity (Sirikantaramas *et al.*, 2004). *Cannabis* varieties exhibit one of three major cannabinoid profiles depending on the relative abundance of THCA and CBDA. Marijuana cultivars contain high concentrations of THCA relative to CBDA, whereas hemp cultivars have relatively higher concentrations of CBDA and hybrids are intermediate (de Meijer *et al.*, 2003). Although recent work has identified the genes encoding the enzymes of the cannabinoid biosynthetic pathway (Marks *et al.*, 2009a; van Bakel *et al.*, 2011; Gagne *et al.*, 2012), fundamental questions remain about the inheritance of these genes in relation to the diversity of cannabinoid phenotypes (Page & Nagel, 2006).

A simple Mendelian model involving codominant alleles at a single locus was proposed to explain the segregation of drug content among marijuana and hemp plants (de Meijer *et al.*, 2003;

Pacifico *et al.*, 2006). Supposing that the enzymes THCA synthase and CBDA synthase are products of allelic variants at a single locus, de Meijer *et al.* (2003) predicted that marijuana would be homozygous for THCA synthase, hemp would be homozygous for CBDA synthase, and their hybrid would be intermediate. However, the distribution of gene sequences encoding the major cannabinoid biosynthetic enzymes suggests an alternative model involving at least two loci (Kojoma *et al.*, 2006). Linkage of separate loci for THCA and CBDA synthases with variable expression or allelic variation in enzymatic efficiency can explain the observed pattern of phenotypic segregation (van Bakel *et al.*, 2011). We evaluated these alternatives based on analyses of an F₂ population derived from a cross between inbred marijuana and hemp cultivars. Molecular cloning and phylogenetic analysis of THCA synthase and CBDA synthase homologs established a minimum number of homologs per inbred line. Genotyping, quantitative trait locus (QTL) mapping, and gene expression assays associated genomic regions and loci with phenotypes to demonstrate that despite the linkage of separate loci for THCA synthase and CBDA synthase, it is variation at the CBDA synthase locus alone that accounts for the main difference between the cannabinoid profiles of marijuana and hemp.

Materials and Methods

Breeding program

Cannabis sativa L. seeds, including marijuana (Skunk #1) and hemp (Carmen) cultivars, were imported from Hortapharm BV (Amsterdam, the Netherlands) and Kenex Ltd (Ontario, Canada), respectively, under permit from the US Drug Enforcement Administration (DEA). Plants were grown in hydroponic media (Grodan) with an automated water exchange system, commercial nutrient solutions (General Hydroponics, Sebastopol, CA, USA) and high-pressure sodium vapor lighting. An initial 4 wk of vegetative growth conditions employed an 18 : 6 h, light : dark regime and high-N nutrient solution. During vegetative growth, the sex of individual plants was determined using a male-linked sequence characterized amplified region (SCAR) marker (Mandolino *et al.*, 1999). Staminate and pistillate plants were subsequently maintained in separate growth chambers during 10–12 wk of flowering conditions using a 12 : 12 h, light : dark regime and high phosphorus (P) nutrient solution.

Full-sib inbreeding was performed in each line for multiple generations to reduce heterozygosity. Each generation, the male cohort was thinned to a single plant before anthesis. Pollen was applied by hand to flowers of a single female plant over several weeks of successive flowering and subsequent generations were established from the resulting seed. This process was repeated five times to produce near-isogenic lines that served as parents in a hybrid cross between the drug and fiber cultivars.

A hybrid cross between a single staminate hemp plant and single pistillate marijuana plant was made and segregation among F₂ progeny was examined by selfing a single F₁ plant. *Cannabis* is predominantly dioecious but monoecious variants are common (Hirata, 1927) and male sex expression may be induced in

genetic females by means of hormonal treatment (Ram & Sett, 1982). Immature pistillate inflorescences of a clonally propagated F₁ plant were treated with AgNO₃ to induce the development of fertile staminate flowers. Pollen collected at anthesis was stained with 2,5-diphenyl monotetrazolium bromide in 5% sucrose to test for viability (Rodriguez-Riano & Dafni, 2000) and applied to pistillate inflorescences of numerous F₁ clones to produce an F₂ population.

Plants were grown for 4 wk under vegetative conditions and subjected to flowering conditions for an additional 8 wk. After 12 wk of growth, plants were harvested and dried for quantitative analysis of cannabinoids in female inflorescences. Residual plant material was disposed of according to a DEA-approved procedure.

Cannabinoid analysis

Dried samples of female flowers from mature plants were analyzed for content (% dry weight (DW)) of the cannabinoid compounds cannabigerol (CBG), CBDA, cannabichromene (CBC), cannabiol (CBN), tetrahydrocannabivarin (THCV) and THCA by gas chromatography (ElSohly *et al.*, 2000). THCA content expressed as a percentage of inflorescence DW was plotted against CBDA content and plants were categorized according to drug, fiber and intermediate phenotypes. The distribution of phenotypes was compared with a Mendelian expectation using a *G*-test for goodness of fit.

Amplified fragment length polymorphisms

Approximately 100 mg of plant material was collected from dried pistillate inflorescences of parents, F₁ plants, and F₂ plants. Plant material was ground to a fine powder in liquid N₂ and genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen). Amplification and scoring of amplified fragment length polymorphism (AFLP) markers were performed according to Datwyler & Weiblen (2006). AFLP products were analyzed on an ABI377 DNA sequencer (Applied Biosystems Inc., Grand Island, NY, USA). Markers differing between marijuana and hemp parents at a threshold of 100 relative fluorescence units were scored for 10 selective primer pairs: *EcoRI*-AAC/*MseI*-CTT (E1M8), *EcoRI*-ACA/*MseI*-CAC (E3M2), *EcoRI*-ACT/*MseI*-CAT (E6M4), *EcoRI*-ACT/*MseI*-CTT (E6M8), *EcoRI*-AGC/*MseI*-CAA (E7M1), *EcoRI*-AGC/*MseI*-CTG (E7M7), *EcoRI*-AGG/*MseI*-CAA (E8M1), *EcoRI*-AGG/*MseI*-CAT (E8M4), *EcoRI*-AGG/*MseI*-CTA (E8M5) and *EcoRI*-AGG/*MseI*-CTT (E8M8).

Microsatellite DNA analysis

Sixteen microsatellite primers previously developed for *C. sativa* were applied to the mapping population, including ANUCS201, ANUCS202, ANUCS204, ANUCS205, ANUCS206, ANUCS301, ANUCS302, ANUCS303, ANUCS304, ANUCS305 and ANUCS501 (Gilmore & Peakall, 2003) and B02, C08, C11, H09 and H11 (Alghanim & Almirall, 2003). PCRs were performed using 0.2–0.4 µg genomic DNA, 250 µM dNTP, 0.2 µM

primer, 2.0 mM MgCl₂, 2.0 µg µl⁻¹ BSA, 1× PCR buffer, and 0.3 units Taq polymerase (Takara exTaq, Clontech Laboratories, Mountain View, CA, USA). One primer of each pair was end-labeled with a fluorescent tag (Fam, Vic, or Ned; Integrated DNA Technologies, Coralville, IA, USA). PCRs consisted of a 5 min preheat at 95°C followed by two rounds of cycles. Touchdown PCR consisted of 10 cycles of 95°C denaturation for 15 s, 63°C touchdown to 53°C annealing for 30 s, and 72°C extension for 30 s. This was followed by 19 cycles of 95°C denaturation for 15 s, 52°C annealing for 30 s, 72°C extension for 1 min and one 72°C final extension for 10 min. Primers were multiplexed in combinations of ANUCS303/ANUCS204/ANUCS205, ANUCS304/ANUCS305 and ANUCS501/ANUCS201 or otherwise screened individually. Pooled PCR products and an ROX 500 fluorescent size standard were separated on an ABI377 automated sequencer (Applied Biosystems). Alleles were scored using GENOTYPER software (Applied Biosystems).

Cannabinoid synthase genotyping

Tetrahydrocannabinolic acid and CBDA synthase genes were amplified from genomic DNA of F₂ plants in 25 µl reactions containing 0.625 U PrimeSTAR Hot Start DNA polymerase (TaKaRa), 1× PrimeSTAR reaction buffer, 1 mM Mg²⁺, 1 µM each primer, 0.2 mM per dNTP, and *c.* 100 ng template. Amplification conditions for the THCA synthase gene were 94°C for 1 min, followed by 35 cycles of 94°C for 10 s, 55°C for 10 s, and 72°C for 1 min 45 s, then one cycle of 72°C for 5 min followed by a 4°C indefinite hold. Procedures for amplification of the CBDA synthase gene were identical except that cycles were reduced from 35 to 30 cycles.

In the case of CBDA synthase, *c.* 960 bp at the 5' end of the gene was obtained using primers CBDAsynFor (forward; ATG AAG TGC TCA ACA TTC) and CBDA961Rev (reverse; CCA CTC CAC CAA GGA AAA C). PCR products were sequenced using the CBDAsynFor primer by the Sanger method. CBDA synthase genotypes were determined by comparison with reference sequences from marijuana (van Bakel *et al.*, 2011) and hemp (Kojoma *et al.*, 2006) in GENEIOUS 6.1.8 software. A 4 bp deletion present in marijuana (van Bakel *et al.*, 2011) was the basis for scoring F₂ plants as homozygous for either marijuana-type or hemp-type CBDA synthase. Heterozygous individuals were identified when chromatograms showed multiple peaks immediately downstream of the four bp deletion.

In the case of THCA synthase, genotypes were generally determined using an agarose gel assay to visualize presence or absence of 1676 bp product amplified using THCA synthase-specific primers derived from GenBank accession AB057805 (THCAsynF: GGA CTG AAG AAA AAT GAA TTG CTC AG and THCAsynR: GGG AAA TAT ATC TAT TTA AAG ATA ATT AAT GAT). The THCAsynR primer is selective for marijuana-type THCA synthase and generally failed to yield products from hemp genomic DNA (J. P. Wenger, pers. obs.). Faint products observed from a subset of F₂ individuals were sequenced in order to determine the presence of

marijuana-type THCA synthase or cross-amplification of hemp-type THCA synthase.

Linkage mapping and QTL analysis

Amplified fragment length polymorphism and microsatellite genotypes for 62 F₂ individuals and the parents were used to produce a linkage group tree using JOINMAP 4 (Wageningen, the Netherlands). Linkage groups were assembled from independent log-of-odds scores (LOD) based on *G*-tests for independence of two-way contingency tables. Linkage groups with LOD > 3.0 and containing four or more markers were used to construct a linkage map using the function of Kosambi (1944). Cannabinoid profiles of the same 62 individual genotypes were analyzed with respect to the linkage map using Windows QTL CARTOGRAPHER v.2.5 (Wang *et al.*, 2006). Significant associations between traits and linkage groups were identified using an experiment-wise LOD threshold. Simple interval mapping was used to estimate LOD over intervals equal to the mean distance between markers for a given linkage group (i.e. intervals of 4.5 and 7.0 cM for two linkage groups with putative QTL). Significance thresholds for LOD were estimated in Win QTL CARTOGRAPHER 2.5 by permutation. Experiment-wise thresholds were estimated by 1000 permutations for a given trait over all linkage groups to obtain an $\alpha = 0.05$ threshold LOD for the trait with respect to the entire map. Group-wise thresholds were estimated in like manner, except that the permutations were performed only over the corresponding linkage group for a given trait. Results were plotted with MAPCHART 2.1 (Wageningen, the Netherlands).

Molecular cloning and phylogenetic analysis

Tetrahydrocannabinolic acid and CBDA synthase genes were amplified from parental marijuana and hemp plants using genomic DNA in separate 25 µl PCR reactions with specific primers for each gene as described earlier, but instead of using CBDA961Rev, we used a CBDA synthase-specific reverse primer anchored in the 3' end of the gene that was derived from GenBank accession AB292682 (CBDAsynR: TTA ATG ACG ATG CCG TGG). PCR products were isolated from 1% (w/v) agarose gels using the QIAEX II kit (Qiagen) and cloned into the pCR4-TOPO vector (Invitrogen). Plasmid DNA was isolated from clones and subjected to DNA sequencing using M13 forward, M13 reverse, and internal primers. Internal primers used for sequencing THCA synthase clones were THCA583For: GTG GAG GAG GCT ATG GAG C and THCA1034Rev: CCC AAC TCA GGA AAG CTC TTG. Internal primers for CBDA synthase sequencing were CBDA694For: GGT GGT GGA GCA GAA AGC TTC and CBDA961Rev: CCA CTC CAC CAA GGA AAA C. Sequence contigs were assembled using the program SEQMAN PRO (DNASTAR/Lasergene 8).

Unique sequences obtained from the mapping population were aligned with published THCA and CBDA synthase sequences (Sirikantaramas *et al.*, 2004; Kojoma *et al.*, 2006; Taura *et al.*, 2007; van Bakel *et al.*, 2011) and inspected for premature stop codons or frame shifts. Phylogenetic analysis was

performed using MRBAYES 3.2 (Huelsenbeck & Ronquist, 2001). Two analyses ran in parallel, each with six chains of one million generations and the posterior distribution was sampled every 100 generations. The first 25% of the sample was discarded as 'burn-in'. Clade support was assessed by posterior probability. We performed tests for evidence of selection by comparing nonsynonymous and synonymous substitution rates (dN/dS) using HyPhy (Kosakovsky Pond *et al.*, 2005).

Real-time quantitative PCR

Pistillate inflorescences bearing capitate glandular trichomes were harvested from F₂ plants after 21 d of growth under flowering conditions and total RNA was extracted from 100 mg of tissue using a Plant RNeasy Kit (Qiagen), treated with Turbo DNase (Ambion), purified, and concentrated using an RNA MinElute kit (Qiagen). One µg of total RNA from each sample served as template in synthesizing cDNA with SuperScript III First-Strand Synthesis SuperMix (Invitrogen).

Real-time quantitative PCR (RT-qPCR) was performed with SYBR[®] Green JumpStart Taq ReadyMix Capillary Formulation (Sigma-Aldrich) on a Light Cycler (Roche) using replicate samples of the derived cDNA for analysis of expression of THCA and CBDA synthases following previously reported protocols (Marks *et al.*, 2009a). Primer sets were designed to specifically amplify either gene. Primers for THCA synthase were THCAsynF: GGA CTG AAG AAA AAT GAA TTG C and THCAsynR: GGT CGT GTT GAG TGT ATA CG. Primers for CBDA synthase were CBDAsynF: GAA CTA AAG AAA AAT GAA GTG CTC AA and CBDAsynR: GTT GCA TTA TTG GGA ATA TAT TGC. Three serial dilutions of isolated PCR products synthesized using the respective primer pairs were used as templates to generate standard curves for the RT-qPCR analysis. Sample template quantities were inferred from standard curves using LIGHT CYCLER software. Gel-isolated RT-qPCR products were sequenced to verify their identity.

Results

Cannabinoid variation

Three major cannabinoid phenotypes (marijuana, hemp, and intermediate) were evident among the pistillate inflorescences of 540 genetically distinct *Cannabis* plants (Fig. 1). Plants were classified as marijuana when $\log(\%THCA/\%CBDA) \geq 1.0$, hemp when $\log(\%THCA/\%CBDA) \leq -1.0$, and intermediate when $\log(\%THCA/\%CBDA) > -1.0$ and < 1.0 (Table 1). All F₁ plants were intermediate and segregation of cannabinoid phenotypes in the F₂ population was not significantly different from the 1 : 2 : 1 Mendelian expectation of marijuana to intermediate to hemp plants (127 : 264 : 105 plants, $G = 4.16$, $P = 0.125$).

Apart from qualitative differences in cannabinoid content, the total quantity of cannabinoids was significantly different between marijuana and hemp plants (Table 1, $P < 0.05$, Tukey's honest significant difference (HSD) test). Marijuana plants averaged 4.5 times more total cannabinoids per unit inflorescence biomass

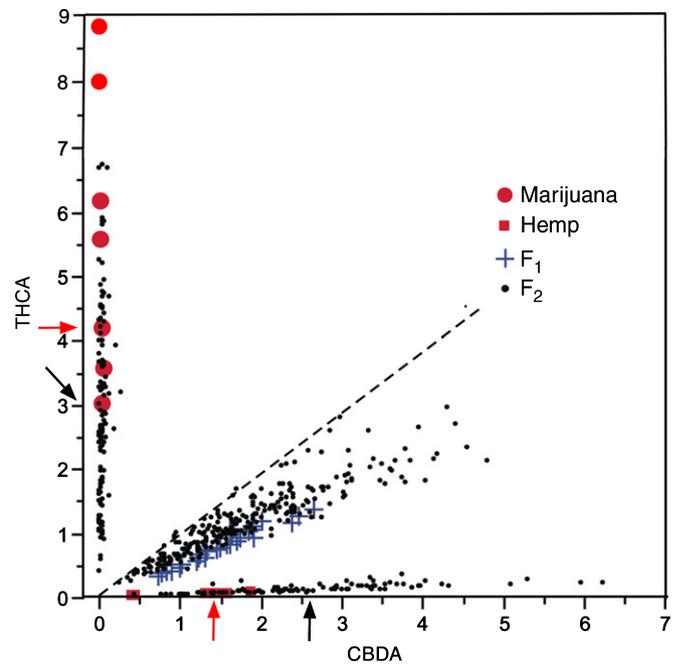


Fig. 1 Delta⁹-tetrahydrocannabinolic acid (THCA) versus cannabidiolic acid (CBDA) content as a % of dry weight (DW) in mature pistillate inflorescences of *Cannabis sativa* derived from a marijuana parent and siblings, a hemp parent and siblings, the F₁ parent and siblings, and the F₂ cohort. Red arrows point to the hemp and marijuana parents; black arrows point to mean THCA content and CBDA content for marijuana-like and hemp-like F₂s, respectively. The dashed line indicates the expected ratio of THCA to CBDA if the two main cannabinoid synthase enzymes were equally competitive for their common precursor, cannabigerolic acid.

than hemp plants and F₁ plants were intermediate. The overall distribution of cannabinoids in the F₂ generation indicated the independent assortment of cannabinoid quantity and quality. In particular, marijuana F₂ plants had significantly lower total cannabinoid content on average than the marijuana parent, whereas F₂ hemp plants had significantly higher content than the hemp parent (Fig. 1).

Linkage mapping and QTL

A subset of 62 plants from the F₂ population was genotyped for 103 AFLP markers, 16 microsatellite markers, CBDA synthase, and THCA synthase sequences to construct a linkage map. Nine linkage groups with LOD > 3.0 and containing at least four markers were identified (Fig. 2). These groups covered a genetic distance of 335.7 cM with an average between-marker distance of 6.10 cM among 11 microsatellites, 51 AFLPs, CBDA synthase and THCA synthase.

Analysis of cannabinoid profiles from the same 62 individuals used in mapping identified putative QTLs for cannabinoid content in association with two of the nine linkage groups. Such a small sample of genotypes would seem to be insufficient for identifying QTLs in general, but discrete cannabinoid phenotypes (Fig. 1) simplified the problem to such a degree that a significant QTL for the THCA/CBDA ratio could be assigned to linkage group six (Fig. 2). Both the group-wise and experiment-wise LOD thresholds ($\alpha = 0.05$) for $\log THCA/CBDA$ of 3.21 and

Table 1 Mean (SD) cannabinoid content in mature pistillate inflorescences from *n* marijuana, hemp, F₁, and F₂ plants as a percentage of total dry weight (DW)

	<i>n</i>	% THCA	% CBDA	% total	log(THCA/CBDA)
Marijuana	7	5.65 (2.24)	0.03 (0.02)	6.25 (2.29)	2.32 (0.41)
Hemp	7	0.07 (0.02)	1.21 (0.57)	1.38 (0.59)	-1.17 (0.18)
F ₁	30	0.82 (0.28)	1.52 (0.49)	2.53 (0.81)	-0.27 (0.04)
F ₂ marijuana-like	127	3.05 (1.67)	0.03 (0.04)	3.75 (1.53)	2.05 (0.41)
F ₂ intermediate	264	1.19 (0.55)	1.19 (0.92)	2.89 (1.46)	-0.18 (0.1)
F ₂ hemp-like	105	0.13 (0.07)	2.61 (1.34)	3.69 (1.87)	-1.29 (0.09)

THCA, tetrahydrocannabinolic acid; CBDA, cannabidiolic acid.

21.52, respectively, were surpassed by a maximum LOD of 40.44. A putative QTL for overall cannabinoid quantity located on linkage group one exceeded a group-wise threshold of 2.18 with an LOD of 3.30 but was not significant given an experiment-wise LOD of 3.42.

Cannabinoid synthase genes

Molecular cloning of THCA and CBDA synthase genes of Skunk #1 (marijuana), Carmen (hemp), as well as F₁ intermediate plants from the mapping population yielded products similar to previously published sequences. We identified nine unique sequences among 139 clones. A phylogenetic analysis including all available cannabinoid synthase sequences was rooted with the most similar sequence from *Humulus* sharing 74–86% nucleotide identity (Fig. 3). Branch lengths in Fig. 3 are proportional to the number of substitutions per site under the general time-reversible model of molecular evolution and circles indicate Bayesian posterior probabilities (>0.95). Skunk #1 yielded three homologs of the THCA synthase gene, including one nearly identical to the THCA synthase sequence first isolated from marijuana (Sirikantaramas *et al.*, 2004), a second that was highly similar to a sequence previously isolated from hemp (Kojoma *et al.*, 2006), and a new sequence with 92–93% nucleotide similarity to the others. Carmen yielded three homologs, including a sequence nearly identical to that previously isolated from hemp (Kojoma *et al.*, 2006), a second identical to the new sequence from Skunk #1, and a third with 94–95% nucleotide similarity to those published previously. The only CBDA synthase gene sequence obtained from Carmen was highly similar to that isolated from another hemp-type *Cannabis* (Taura *et al.*, 2007), whereas two nonfunctional homologs, containing premature stop codons and frame shift mutations, were isolated from Skunk #1. Each of these homologs clustered with a nonfunctional CBDA synthase sequence from the marijuana Purple Kush genome (van Bakel *et al.*, 2011).

Nonsynonymous substitutions (*dN*) were underrepresented relative to synonymous substitutions (*dS*) in the cannabinoid synthase gene family as a whole (*dN/dS* = 0.531 with a 95% CI of 0.474–0.587) but this was not the case for the nonfunctional homologs. The only clade having a significantly elevated *dN/dS* ratio comprised the nonfunctional CBDA synthases (Fig. 3). These homologs shared a 4 bp deletion starting at position 153 of the gene and Skunk #1 homolog 1 also contained a premature stop codon upstream of this position.

Cannabinoid synthase gene expression

Inflorescences of marijuana-type and hemp-type plants yielded significantly different quantities of cannabinoid synthase RNA transcripts as measured by RT-qPCR (Table 2). Expression of THCA synthase in marijuana inflorescences was 102-fold greater than in hemp inflorescences ($P < 0.05$, *t*-test), whereas CBDA synthase expression in hemp inflorescences was 113-fold greater than in marijuana inflorescences. Relative CBDA synthase and THCA synthase expression levels in inflorescences of F₂ plants were examined alongside variation in cannabinoid phenotype (Table 2). As measured by the PCR cycle during which specific products were first detected (i.e. crossing threshold), F₂ inflorescences expressing relatively more CBDA synthase were higher in CBDA than in THCA, whereas F₂ inflorescences expressing relatively more THCA synthase were higher in THCA than in CBDA. F₂ plants expressing similar levels of CBDA and THCA synthases were of intermediate phenotype.

Gel purification and sequencing of qPCR products confirmed that the expressed genes of intermediate plants correspond to THCA and CBDA synthase genes of their respective marijuana and hemp parents (Fig. 3). Furthermore, F₂ plants with elevated CBDA content expressed the functional homolog of CBDA synthase and a relatively low level of hemp-type THCA synthase (Kojoma *et al.*, 2006). Elevated THCA content was associated with expression of marijuana THCA synthase and extremely low levels of a nonfunctional CBDA synthase homolog.

Discussion

Cannabinoid inheritance

Cannabis genotype–phenotype associations, patterns of gene expression, and the evolutionary history of the cannabinoid synthase gene family provide new insights into the genetic basis of drug content. The Mendelian pattern of inheritance observed in our mapping population (Fig. 1) is consistent with the interpretation that cannabinoid quality might be the product of a single enzymatic locus with codominant alleles. de Meijer *et al.* (2003) hypothesized that THCA synthase and CBDA synthase are allelic variants, whereas van Bakel *et al.* (2011) suggested that they could rather be linked loci. The distribution and expression of cannabinoid synthase homologs among descendants of marijuana crossed with hemp favors the alternative view (Table 2). Five generations of full-sib inbreeding make it likely that the parental lines would

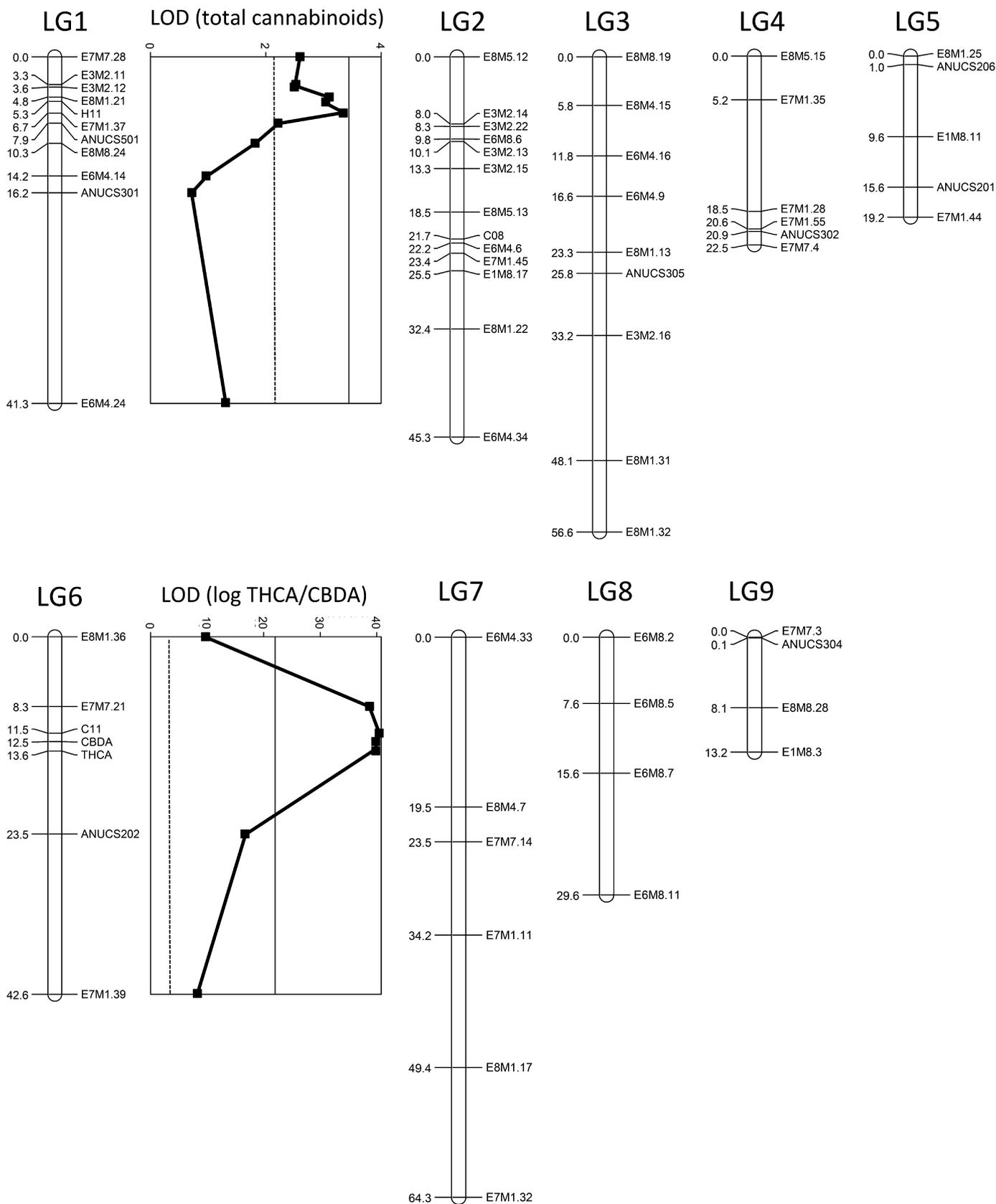


Fig. 2 Nine linkage groups in *Cannabis sativa* based on amplified fragment length polymorphism (AFLP), microsatellite markers and cannabinoid synthases with log-of-odds (LOD) plots of total cannabinoid content (% inflorescence DW) and log THCA/CBDA content shown to the right of linkage groups containing significant group-wise quantitative trait loci (QTL). Dashed and solid vertical lines represent group-wise and experiment-wise $\alpha = 0.05$ thresholds, respectively. THCA, tetrahydrocannabinolic acid; CBDA, cannabidiolic acid.

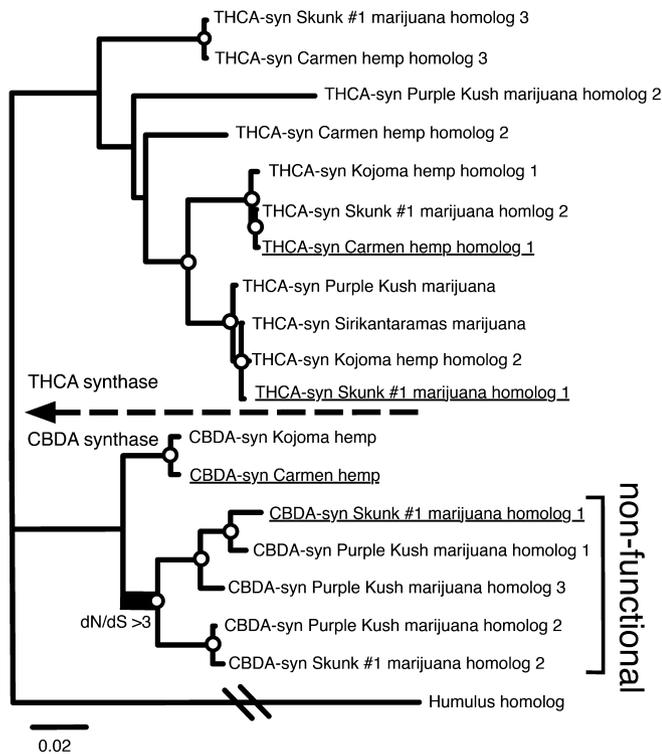


Fig. 3 Bayesian gene tree for tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) synthase gene sequences inferred from a 1614 base-pair alignment of nine unique sequences obtained from the mapping population (Genbank accession numbers KJ469374–KJ469376 and KJ469378–KJ469383), nine previously published *Cannabis* sequences, and Genbank accession number LA634839, the most similar sequence from the *Humulus* (Natsume *et al.*, 2015). [Correction added after online publication 17 July 2015: Genbank accession numbers in the preceding sentence have been updated.] The tree was rooted with *Humulus*, which is the closest extant relative of *Cannabis* (Zerega *et al.*, 2005). Genes expressed in *Cannabis* trichomes are underlined. Branch lengths are proportional to the number of nucleotide substitutions per site, except for the lengthy branch to *Humulus*, which is truncated. Genbank accession numbers for THCA synthase sequences are listed in parentheses as follows: Skunk #1 homolog 3 (KJ469382), Carmen homolog 3 (KJ469383), Purple Kush homolog 2 (JH227480), Carmen homolog 2 (KJ469381), Kojoma homolog 1 (AB212836), Skunk #1 homolog 2 (KJ469379), Carmen homolog 1 (KJ469380), Purple Kush homolog 1 (JH239911), Sirikantaramas (AB057805), Kojoma homolog 2 (AB212829), Skunk #1 homolog 1 (KJ469378). Accession numbers for CBDA synthase are Kojoma (AB292682), Carmen (KJ469374), Skunk #1 homolog 1 (KJ469375), Purple Kush homolog 1 (JH231038), Purple Kush homolog 3 (AGQN01254730), Purple Kush homolog 2 (AGQN01159678) and Skunk #1 homolog 2 (KJ469376).

be homozygous, and yet inbred Carmen and Skunk #1 yielded four and five cannabinoid synthase homologs, respectively. This observation and that of van Bakel *et al.* (2011) based on a draft genome point to the existence of multiple loci.

However, among nine cannabinoid synthase homologs detected in our mapping population (Fig. 3), only four showed evidence of expression in RT-qPCR assays (Table 2). Marijuana-type plants expressed a functional THCA synthase nearly identical to that first reported by Sirikantaramas *et al.* (2004) and a nonfunctional CBDA synthase. Hemp-type plants expressed

functional THCA and CBDA synthases similar to those reported by Kojoma *et al.* (2006). It is possible that despite similar GC content and conserved primer sites among homologs of THCA synthase and among CBDA synthase homologs, qPCR bias could limit our capacity to measure the expression of other homologs. Regardless, the simultaneous presence of all four homologs in F₂ intermediate plants, as indicated by genotyping, is strong evidence of heterozygosity at separate loci for THCA synthase and CBDA synthase.

Examination of the THCA/CBDA ratio in plants of intermediate phenotype suggests that THCA synthase and CBDA synthase might not be equally competitive for CBGA, the precursor that is shared by the two enzymes. All else being equal we would expect a 1 : 1 THCA/CBDA ratio in plants expressing functional copies of both enzymes, but instead we observed a ratio that is significantly skewed toward CBDA (Fig. 1). The fact that intermediate plants produce mostly CBDA despite expressing a functional THCA synthase suggests that CBDA synthase is a superior competitor for cannabigerolic acid.

Colocation on the map of genes for THCA synthase and CBDA synthase in association with a QTL for the THCA/CBDA ratio (Fig. 2) is consistent with the multilocus model for the inheritance of the three qualitatively different cannabinoid phenotypes (de Meijer *et al.*, 2003; van Bakel *et al.*, 2011). Especially relevant to this interpretation is evidence for the expression of a nonfunctional CBDA synthase homolog in our mapping population (Table 2). If this homolog were an unlinked pseudogene, we would not expect it to segregate with cannabinoid phenotype, but the perfect association of this homolog with the marijuana phenotype suggests that it is probably an allele sharing a locus with functional CBDA synthase. Under a linked multilocus model of inheritance, plants of intermediate phenotype are expected to be heterozygous at separate loci for CBDA and THCA synthase and the genotypes of F₂ intermediates are consistent with this prediction. Expression assays of F₂ intermediates suggest that the marijuana-type THCA synthase allele may be dominant over the hemp-type allele, and the functional CBDA synthase allele may be dominant over the nonfunctional allele (Table 2). Thus, plants of intermediate phenotype inherit a highly expressed THCA synthase from one parent and a highly expressed CBDA synthase from the other, whereas marijuana and hemp plants inherit either a highly expressed THCA synthase or a highly expressed CBDA synthase, respectively. Genotyping also identified a single recombinant individual that was homozygous for the hemp-type THCA synthase and homozygous for the marijuana-type nonfunctional CBDA synthase. The marijuana phenotype exhibited by this individual suggests that the absence of functional CBDA is essential for marijuana potency.

Quality vs quantity

The statistical association between cannabinoid phenotype and the location of cannabinoid synthases on the map (Fig. 2) was so strong that even a relatively small sample size was sufficient to detect it. We attribute this result to extreme divergence among parental cultivars in the THCA/CBDA ratio. This was not the

Table 2 Segregation of cannabinoid phenotype and cannabinoid synthase gene expression among *n* plants in an F₂ population derived from the cross of marijuana and hemp

	<i>n</i>	THCA vs CBDA	THCA synthase vs CBDA synthase	THCA synthase gene sequence	CBDA synthase gene sequence
Marijuana-like	7	2.29 (0.67)	12.726 (10.876)	Marijuana-type	Marijuana-type
Intermediate	18	-0.12 (0.09)	0.152 (0.049)	Marijuana-type	Hemp-type
Hemp-like	9	-1.22 (0.12)	0.003 (0.002)	Hemp-type	Hemp-type

Means (and SDs) are reported for relative cannabinoid content (log %THCA/%CBDA of inflorescence DW) and relative expression as measured by RT-qPCR (ng CDA THCA synthase/CBDA synthase). Expressed sequences are underlined in Fig. 3. THCA, tetrahydrocannabinolic acid; CBDA, cannabidiolic acid.

case for total cannabinoid content, where a putative QTL on a different linkage group nearly approached significance in the experiment-wise test (Fig. 2). The observation that marijuana-like F₂ plants had reduced cannabinoid content and hemp-like F₂ plants had elevated content relative to their respective parental lines (Fig. 1) suggests the independent assortment of drug quality and quantity. We speculate that an improved genetic map with more markers and more extensive screening of individuals is likely to provide the statistical power needed to identify QTLs for quantitative differences among cultivars in potency that is not easily explained by variation at cannabinoid synthase loci.

Cannabinoid quantity varies among cultivars by four orders of magnitude (Vogelmann *et al.*, 1987; de Meijer *et al.*, 1992; Petri *et al.*, 1996; Szendrei, 1997) and it is likely that other mechanisms besides the cannabinoid biosynthetic pathway are, to some extent, responsible for this variation. In particular, the size of glandular trichomes might account for substantial differences in cannabinoid quantity (Small & Naraine, 2015). The identification of genes affecting trichome development in *Cannabis* using homologs from model systems (Marks *et al.*, 2009b) would therefore appear to be a promising avenue for further investigation. The high variance observed in the overall cannabinoid content of the F₁ and the parents despite five generations of full sib-inbreeding also deserves explanation (Fig. 1). We attribute this to the fact that estimates of cannabinoid quantity were based on a single 100 mg tissue sample per plant (ElSohly *et al.*, 2000). Cannabinoid quantity may vary among samples drawn from the same plant according to the extent to which glandular trichomes heads are intact or dislodged and whether the sample consists primarily of pistillate bracts or inflorescence bracts differing in trichome density. Variance among 100 mg samples drawn from the same genotype or even from the same inflorescence could account for much of the scatter observed in Fig. 1.

The possibility of separate QTLs affecting drug quantity and quality suggest that genetic mapping and QTL analysis could aid in identifying complex mechanisms that underlie other economically important traits in *Cannabis* such as nutritious seed oils or properties of industrial fiber. The incomplete map presented in Fig. 2 is one linkage group short of the *Cannabis* haploid chromosome number ($N=10$). With the addition of new markers, we expect six small linkage groups consisting of less than four markers each (not shown in Fig. 2) to associate with others and to resolve a 10th group. Next-generation sequencing and the identification of single nucleotide polymorphisms hold promise

for improving map resolution and for identifying additional QTLs.

Cannabinoid synthase evolution

Phylogenetic analysis suggests that THCA and CBDA synthases arose by the duplication of an ancestral cannabinoid synthase gene (Fig. 3) and the location of QTL suggest that divergent cannabinoid phenotypes are products of subsequent sequence evolution at linked loci (Fig. 2). Gene duplications are known to result from unequal crossing-over and the fate of duplicated regions will depend on the nature of selection (Zhang, 2005). Theoretical population genetics predicts that duplicated gene copies are likely to persist by divergence through subfunctionalization (Lynch & Force, 2000). In the case of *Cannabis*, we speculate that this might have been associated with the evolution of enzymatic activities yielding different secondary metabolites. The simultaneous presence of THCA and CBDA synthase-like genes in diverse *Cannabis* accessions (Kojoma *et al.*, 2006; Taura *et al.*, 2007; van Bakel *et al.*, 2011) suggests that multiple loci may be ancient and pervasive. Global dN/dS further suggests that amino acid replacements appear to have been selected against during the evolution of cannabinoid synthases, which is consistent with the maintenance of function (Kosakovsky Pond *et al.*, 2005).

In the case of marijuana, loss of CBDA synthase function must have occurred after an ancestral duplication because CBDA synthase and nonfunctional homologs are more closely related to each other than to THCA synthase genes (Fig. 3). Loss of function in marijuana-type CBDA synthases can be attributed to various premature stop codons and frame shift mutations. For example, sequences in the clade including CBDA synthase Skunk #1 homolog 1 contain a premature stop codon at the 5' end of the gene and share a 4 bp deletion with CBDA synthase Skunk #1 homolog 2. Local dN/dS is consistent with the interpretation that such mutations were selected to enhance potency (van Bakel *et al.*, 2011) where the only branch of the gene tree exhibiting evidence of positive selection was that leading to nonfunctional CBDA homologs (Fig. 3). If nonfunctional homologs prove to be ubiquitous among drug cultivars, it is possible that selection occurred early in domestication. However, the evolution of cannabinoid synthases alone might not entirely explain recently observed increases in potency (ElSohly *et al.*, 2000) but potentially other traits as well (see quality vs quantity). With respect to genetic diversity, intense selection by breeders for potency might also

account for drug cultivars being generally less polymorphic and heterozygous than hemp cultivars (de Meijer *et al.*, 1992; Faeti *et al.*, 1996; Forapani *et al.*, 2001; Datwyler & Weiblen, 2006).

The available evidence suggests that there are at least two cannabinoid synthase loci and there are likely to be more. For example, we detected a THCA synthase sequence in Skunk #1 (homolog 3) which was nearly identical to that originally isolated from hemp-type *Cannabis* (Kojoma *et al.*, 2006). This homolog was not detected in a draft genome from Purple Kush (van Bakel *et al.*, 2011) and awaits explanation. It is conceivable that additional gene duplications have occurred (Taylor & Raes, 2004). Next-generation sequencing of diverse *Cannabis* cultivars and comparative genomics hold promise for further resolution of the history of duplication, divergence, and selection affecting drug content.

Regulatory implications

Quantitative trait loci for cannabinoid content, patterns of cannabinoid synthase gene expression, and phylogenetic evidence of selection acting on multiple loci suggest an evolutionary genetic basis for the differentiation of hemp and marijuana. The multilocus model of cannabinoid inheritance also has important implications for drug and agricultural regulatory policy. Legislation in Europe, Canada, and the US aimed at isolating hemp cultivation from the marijuana trade defines hemp at <0.3% THC by DW. Now that genes responsible for this phenotype have been identified, there is further advantage in defining hemp cultivars genetically. The distinction in cannabinoid phenotype that is the basis for the legal separation of marijuana from hemp appears to be the product of genetic variation at an expressed CBDA synthase locus. We predict that plants that are homozygous for functional CBDA synthase lack the capacity to yield >0.3% THC and that we need not consider THCA synthase homologs to distinguish hemp from marijuana. For example, intermediate plants with a single nonfunctional CBDA synthase allele have the potential to exceed the legal threshold of 0.3% THC and, as such, could be considered marijuana. Screening for the presence of nonfunctional CBDA could be employed to verify seed sources before planting and address the drug enforcement concern that hemp might conceal marijuana in the field. [Correction added after online publication 17 July 2015: values of THC in this paragraph have been corrected.]

We anticipate that our genetic explanation for the three major cannabinoid phenotypes may also be of pharmacological interest. Under recent medical *Cannabis* legislation in nearly half of the United States, users demand diverse phenotypes to treat conditions ranging from neuropathic pain to epilepsy (Bostwick, 2012). The functions of enzymes involved in cannabinoid biosynthesis that apparently evolved by gene duplication are therefore obvious candidates for selective breeding efforts and other kinds of genetic engineering. Although drug enforcement has limited such research for decades, we have at last identified a gene to dissociate hemp from marijuana for the benefit of drug policy as well as agriculture.

Acknowledgements

This work was supported by the Minnesota Agricultural Experiment Station under project MIN-71-038 and by a David and Lucille Packard Fellowship to G.W. We thank Sarah Anderson and Linnea Peterson-Bunker for assistance with gene expression assays, Rebecca Sims for assistance with CBDA and THCA synthase genotyping, Evan Johnson for proofreading, Peter Tiffin for thoughtful discussion, and two anonymous reviewers for constructive criticism. We are grateful to the offices of the General Counsel and the Vice President for Research at the University of Minnesota, US Senators for Minnesota, the Minnesota Board of Pharmacy, and the US Drug Enforcement Administration for the opportunity to perform this study.

References

- Alghanim HJ, Almirall JR. 2003. Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. *Analytical and Bioanalytical Chemistry* 376: 1225–1233.
- van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. 2011. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* 12: R102.
- Bostwick JM. 2012. Blurred boundaries: the therapeutics and politics of medical marijuana. *Mayo Clinic Proceedings* 87: 172–186.
- Datwyler SL, Weiblen GD. 2006. Genetic variation in and (*Cannabis sativa* L.) according to amplified fragment length polymorphisms. *Journal of Forensic Sciences* 51: 371–375.
- ElSohly MA, Ross SA, Mehmedic Z, Arafat R, Yi B, Banahan BF. 2000. Potency trends of delta⁹-THC and other cannabinoids in confiscated from 1980–1997. *Journal of Forensic Sciences* 45: 24–30.
- Faeti V, Mandolino G, Ranalli P. 1996. Genetic diversity of *Cannabis sativa* germplasm based on RAPD markers. *Plant Breeding* 115: 367–370.
- Fellermeier M, Eisenreich W, Bacherr A, Zenk MH. 2001. Biosynthesis of cannabinoids: incorporation experiments with ¹³C-labeled glucoses. *European Journal of Biochemistry* 268: 1596–1604.
- Forapani S, Carboni A, Pioletti C, Moliterni VMC, Ranalli P, Mandolino G. 2001. Comparison of varieties using random amplified polymorphic DNA markers. *Crop Science* 41: 1682–1689.
- 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. *Proceedings of the National Academy of Sciences, USA* 109: 12811–12816.
- Gilmore S, Peakall R. 2003. Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana). *Molecular Ecology Notes* 3: 105–107.
- Hennink S. 1994. Optimisation of breeding for agronomic traits in fibre (*Cannabis sativa* L.) by study of parent-offspring relationships. *Euphytica* 78: 69–76.
- Hirata K. 1927. Sex determination in (*Cannabis sativa* L.). *Journal of Genetics* 19: 65–79.
- Huelsbeck JP, Ronquist FR. 2001. MRBAYES: Bayesian inference of phylogeny. *Biometrics* 17: 754–755.
- Kim ES, Mahlberg PG. 1997. Immunochemical localization of tetrahydrocannabinol (THC) in cryofixed glandular trichomes *Cannabis* (Cannabaceae). *American Journal of Botany* 84: 336–342.
- 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: 132–140.
- Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
- Kosambi DD. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* 12: 172–175.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.

- Mandolino G, Carboni A, Forapani S, Faeti V, Ranalli P. 1999. Identification of DNA markers linked to the male sex in dioecious (*Cannabis sativa* L.). *TAG. Theoretical and Applied Genetics* 98: 86–92.
- Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, Ji H, Gang D, Weiblen GD, Dixon RA. 2009a. Identification of candidate genes affecting delta-9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *Journal of Experimental Botany* 13: 3175–3726.
- Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA. 2009b. Transcriptome analysis of *Arabidopsis* wild-type and *gl3-sst sim* trichomes identifies four additional genes required for trichome development. *Molecular Plant* 2: 803–822.
- Mechoulam R, Shani A, Yagnitinsky B, Ben-Zvi Z, Braun P, Gaoni Y. 1970. Some aspects of cannabinoid chemistry. In: Joyce CRB, Curry SH, eds. *The botany and chemistry of Cannabis*. London, UK: Churchill, 93–115.
- de Meijer EPM, Bagatta M, Carboni A, Crucitti P, Moliterni VMC, Ranalli P, Mandolino G. 2003. The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics* 163: 335–346.
- de Meijer EPM, van der Kamp HJ, van Eeuwijk FA. 1992. Characterisation of *Cannabis* accessions with regard to cannabinoid content in relation to other plant characters. *Euphytica* 62: 187–200.
- Mignoni G. 1997. Cannabis as a licit crop: recent developments in Europe. *Bulletin on Narcotics* 49: 23–44.
- Miron JA. 2008. *The budgetary implications of drug prohibition*. Report to the Criminal Justice Policy Foundation. Silver Spring, MD, USA.
- Natsume S, Takagi H, Shiraiishi A, Murata H, Toyonaga J, Patzak J, Takagi M, Yaegashi H, Uemura A, Mitsuoka C *et al.* 2015. The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant and Cell Physiology* 56: 428–441.
- 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: 257–268.
- Page JE, Nagel J. 2006. Biosynthesis of terpenophenolic metabolites in hop and cannabis. In: Romeo JT, ed. *Recent advances in phytochemistry*. Oxford, UK: Elsevier Ltd, 179–210.
- Petri G, Veszki P, Oroszlan P, Turiak G. 1996. Cannabinoid content of foreign and domestic fibre varieties cultivated in Hungary. *Acta Agronomica Hungarica* 44: 133–140.
- Ram HYM, Sett R. 1982. Induction of fertile male flowers in genetically female *Cannabis sativa* plants by silver nitrate and silver thiosulfate anionic complex. *Theoretical and Applied Genetics* 62: 369–375.
- Rodriguez-Riano T, Dafni A. 2000. A new procedure to assess pollen viability. *Sexual Plant Reproduction* 12: 241–244.
- Shoyama Y, Yagi M, Nishioka I, Yamauchi T. 1975. Biosynthesis of cannabinoid acids. *Phytochemistry* 14: 2189–2192.
- Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F. 2004. The gene controlling psychoactivity: molecular cloning and heterologous expression of δ^1 -tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *Journal of Biological Chemistry* 279: 39767–39774.
- Small E, Naraine SGU. 2015. Size matters: evolution of large drug-secreting resin glands in elite pharmaceutical strains of *Cannabis sativa* (marijuana). *Genetic Resources and Crop Evolution* 254: doi: 10.1007/s10722-015-0254-2.
- Szendrei K. 1997. Cannabis as an illicit crop: recent developments in cultivation and product quality. *Bulletin on Narcotics* 49: 1–22.
- Taura F, Morimoto S, Shoyama Y. 1995. Cannabinerolic acid, a cannabinoid from *Cannabis sativa*. *Phytochemistry* 39: 457–458.
- Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S. 2007. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS Letters* 581: 2929–2934.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics* 38: 615–643.
- Vogelmann AF, Turner JC, Mahlberg PG. 1987. Cannabinoid occurrence in seedlings of *Cannabis sativa* L.: quantitation in seedlings of known age and primary leaf length. *Botanical Gazette* 148: 468–474.
- Wang S, Basten CJ, Zeng ZB. 2006. *Windows QTL Cartographer 2.5*. North Carolina State University, Raleigh, NC, USA: Department of Statistics.
- Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. 2005. Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution* 37: 402–416.
- Zhang J. 2005. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 16: 292–298.



About New Phytologist

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <27 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**