Cannabinoid Inheritance Relies on Complex Genetic Architecture

Lesley G. Campbell,* Jaimie Dufresne, and Sarah A. Sabatinos

Abstract

Introduction: Understanding the inheritance of cannabinoid compounds in *Cannabis sativa* will facilitate effective crop breeding and careful regulation of controlled substances. The production of two key cannabinoids, Δ -9-tetrahydrocannabinol (THC) and cannabidiol (CBD), is partially controlled by two additive loci. Here, we present the first study to search for evidence of alternate genetic models describing the inheritance and expression of cannabinoids.

Materials and Methods: Using an information-theoretic approach, we estimated composite genetic effects (CGEs) of four cultivars with pure CBD or pure THC chemotypes, their F₁ and F₂ hybrid progeny, to identify genetic models that explain cannabinoid inheritance patterns. We also estimated the effective number of genetic factors that control differences in cannabinoid concentration (THC, CBD, and cannabichromene [CBC]).

Results: Unlike previous research, we note nonadditive components of cannabinoid inheritance. Concentration of THC is a polygenic trait (three to four genetic factors). Both additive and dominance CGEs best explained THC expression patterns. In contrast, cytoplasmic genomes and additive genes may influence CBD concentration. Maternal additive effects and additive genetic effects apparently influence CBC expression.

Conclusions: Cannabinoid inheritance is more complex than previously appreciated; among other genetic effects, cytogenetic and maternal contributions may be undervalued influences on cannabinoid ratios and concentrations. Further research on the environmental sensitivity of cannabinoid production is advised.

Keywords: composite genetic effects; genetic architecture; hemp; joint-scaling test; line cross analysis; marijuana

Introduction

Selection on chemical traits plays a strong role in crop domestication, which is often characterized by recessive alleles at few loci¹ (e.g., *Cannabis sativa*,^{2–5} *Capsicum annuum*,⁶ *Coffea*,⁷ *Nicotiana tabacum*,⁸ and *Papaver somniferum*^{9,10}). Yet, understanding the genetic basis of chemical differences between crop and wild relatives is in its infancy. *Cannabis sativa* L. is a fascinating example of adaptive chemical divergence, producing more than 100 terpenophenolic secondary metabolites called cannabinoids.^{11–14} Although populations vary in cannabinoid composition, the most common cannabinoids include cannabidiol (CBD^{15,16}), Δ -9-tetrahydrocannabinol (THC¹⁷), cannabigerol (CBG¹⁸), and cannabichromene (CBC¹⁹). Breeding ef-

forts have largely focused on modifying the production of THC and CBD.^{20–23}

Molecular data suggest that the stunning chemical diversification in cannabis cultivars is the result of recent breeding efforts.^{5,24,25} Cultivars of *C. sativa* are typically characterized based on both the ratio of THC:CBD and the overall abundance of these cannabinoids. Generally, psychoactive cultivars have high THC:CBD ratios and produce high quantities of cannabinoids. Fiber or oilseed cultivars (hemp) are characterized as low THC:CBD, and generally produce lower quantities of cannabinoids. Here, we ask: what genetic models describe cannabinoid expression within plants, and how many genes underlie cannabinoid differences between *C. sativa* cultivars?

Department of Chemistry and Biology, Ryerson University, Toronto, Canada.

*Address correspondence to: Lesley G. Campbell, PhD, Department of Chemistry and Biology, Ryerson University, 350 Victoria Street, Toronto, ON, M5B 2K3, Canada, E-mail: lesley.g.campbell@ryerson.ca

In 2003, de Meijer et al.² proposed a genetic model to explain THC and CBD production in *C. sativa* populations. Cannabinoid yield per crop area was described as a complex trait that was the product of total above-ground biomass, the fraction of biomass composed of inflorescence, total cannabinoid content, and cannabinoid purity (equation 1 in Ref.²). To start acquiring data to test this model, de Meijer et al.² elucidated the inheritance of THC:CBD ratios. Since THC and CBD production is influenced by the presence of enzymes that catalyze deacidification, the enzymes themselves seem to be controlled by simple Mendelian additivity at a synthase locus.² Individuals possessing two B_T alleles result in THC synthase (and THC) production. Individuals possessing two B_D alleles result in CBD synthase (and CBD) production. Individuals with one B_T and B_D allele produce both THC and CBD. This hypothesis was supported by sequencing of THC synthase and CBD synthase genes, which exhibit 89% genetic similarity.^{26,27} Furthermore, de Meijer et al.²⁸ proposed that a homozygous genotype possessing two loss-of-function alleles, B_0 , at the same synthase locus controlled the accumulation of the precursor CBG in adult plants. Finally, the expression of CBC was found to be environmentally sensitive to light and its expression did not fit an additive or dominance genetic model.⁴ Thus, de Meijer et al.^{4,28} described the inheritance of THC:CBD ratios and CBC production, and proposed a mathematical model describing cannabinoid abundance as a potentially more complex quantitative trait. Although de Meijer et al.^{4,28} examined the inheritance patterns of cannabinoids, the only genetic model tested was a simple additive model, whereas more complex genetic effects remained unexamined. Moreover, the de Meijer et al. model remained largely unexamined until the recent research by Weiblen et al.⁵ which identified two genes controlling the ratio of THC:CBD and a third gene controlling the abundance of cannabinoids, still relying on a relatively simple model of cannabinoid inheritance.

Although cannabinoid inheritance has only been tested against relatively simple additive models,^{2,4,5,28,29} both published data and anecdotal evidence suggest that more complex genetic models are necessary. Growers repeatedly describe differences among clones (suggesting environmental effects) and mothers (suggesting maternal effects). Furthermore, genetically uniform F_1 offspring of parents homozygous for THC and CBD synthase alleles exhibited significant amounts of phenotypic variation (e.g., THC:CBD range: 0.91–

1.48, figure 2 in Ref.²), suggesting a strong influence of environmental variation on cannabinoid expression profiles. Furthermore, THC and CBD expression patterns are reliant on multiple linked loci on chromosome 6 and gene duplication to produce tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) synthase sequences.⁵ In addition, selection on hemp cultivars for low THC concentrations, to meet government regulations, has failed to completely remove the expression of THC in most hemp cultivars.³⁰ All this evidence suggests more complex inheritance patterns of several cannabinoids than could be described by simple additive or dominance genetic models.

To predict the pace and direction of cannabinoid evolution in response to artificial selection, the genetic architecture that underlies variation in cannabinoid production must be understood. Two original studies^{2,29} compared chemotype data to either a single- or two-locus model with multiple alleles using chi-square tests to assess model fit, but did not consider alternative models. We use an information-theoretic (I-T) approach developed by Blackmon and Demuth³¹ to estimate the composite genetic effects (CGEs) contributing to variation in cannabinoid production among divergent psychoactive and fiber/oilseed cultivars, and their F_1 and F_2 offspring. In addition, to estimate the number of genes that drive the expression of THC:CBD chemotype and variation in the absolute production of CBD, THC, and CBC, we used the Castle–Wright estimator. Given that the studied lineages were a product of inbreeding and selection,³² we expected to confirm the previous findings that the difference in cannabinoid production among the cultivars examined was governed by few loci with mostly additive effects. However, the use of an I-T approach³¹ revealed several cannabinoids to be the product of multigene complexes and illuminated additional genetic effects contributing to cultivar differences that were undetected using previous approaches.

Materials and Methods

Parental and offspring lineages

We used two independent data sets with known pedigrees to more thoroughly assess the genetic architecture of cannabinoid inheritance (raw data sets were provided by the authors of Refs.^{2,29}). Detailed methods for establishing the parental and offspring lineages are described elsewhere^{2,29} and summarized here (summarized data in Appendix Table A1). We encourage readers interested in reanalysis to contact the original authors^{2,29} who very generously shared the raw data.

In 2003,² four isolated crosses between inbred female and sex-reversed male plants with contrasting cannabinoid profiles were established from genotypes used at HortaPharm in 1998–2000. Resulting F₁ offspring were grown to maturity. A subset of F₁ females were sex reversed to act as pollen parents and produce the F₂ generation. Resulting F₂ plants were grown to maturity. Each generation was raised under similar greenhouse conditions and strict isolation.

In 2014,²⁹ one chemically produced male (from a THC-predominant clone) pollinated two inbred female plants (from two CBD-predominant clones). To produce F₂ populations, F₁ female plants were crossed with sex-reversed F₁ females. Then, 230 F₂ seedlings were grown in a greenhouse and flowering was induced through photoperiod adjustments.

Chemical assessment

Sample extraction from mature inflorescences of every plant for gas chromatography (GC) analysis was conducted.^{2,29} Both studies used the same equipment and the results are directly comparable; however, there were differences in how cannabinoids were identified and quantified (described in those publications). Briefly, the two studies used compared GC sample peaks with pure standard peaks to identify compounds. The GC peak areas were converted into dry weight con-

centrations using a linear calibration curve from CBD standards. The cannabinoid composition of THC and CBD fractions was reported as the weight proportion (%) of the individual cannabinoids in the total cannabinoid fraction.² The cannabinoid composition was expressed as the proportion of the weight of individual cannabinoids in the total cannabinoid fraction (%). We also analyzed the inheritance patterns of the ratio of THC:CBD. CBC concentrations were also measured in a similar manner.²

Analysis

With these data, we estimated the expected mean phenotype of each parental lineage (P_{THC}, P_{CBD}), along with F₁ and F₂ populations and used this information to distinguish among seven CGEs (Appendix Table A2).³³ One strategy, although not used in any previous research on cannabinoid inheritance,^{2,4,5,28,29} to distinguish among genetic models is line cross analysis (LCA) where genetic models are compared using estimates of mean and standard deviation from a few line crosses (e.g., parental, F₁, F₂ populations). However, joint scaling tests (one method of performing LCA) favor simpler models of genetic architecture and do not allow for uncertainty in genetic model selection.³¹ Simple genetic models may suffer from bias against more complex genetic architecture and may overlook alternative explanations for patterns

Table 1. Parameter Estimates (±Unconditional Standard Error) for Models of the Genetic Architecture of Cannabinoid Inheritance: Δ -9-Tetrahydrocannabinol

Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)						
(1) de Meijer et al. ²		Aa	Ad	Ca	Med	AaAa	AdAd	Mea
Aa	0.285	2.43 (1.27)						
Aa, Ad	0.255	3.28 (0.17)	−1.25 (0.11)					
Ca	0.146			1.12 (0.78)				
Med	0.095				0.68 (0.61)			
AaAa	0.076					−1.71 (1.77)		
AdAd	0.045						−0.47 (0.93)	
Ad	0.038		−0.36 (1.33)					
Mea	0.036							−0.11 (0.68)
Variable importance ->		0.551	0.293	0.146	0.095	0.076	0.065	0.052

Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)					
(2) Staginuss et al. ²⁹		Aa		Ca		AaAa	
Aa	0.442			44.57 (0.53)			
Ca, AaAa	0.176				44.50 (0.08)		53.37 (0.92)
Ca, Aa	0.176			53.37 (0.92)	−8.86 (0.92)		
Aa, AaAa	0.176			44.50 (0.08)	8.86 (0.92)		
Variable importance ->				0.822	0.352		0.351

Cannabinoid inheritance is described for THC based on data from either (1) de Meijer et al.² or (2) Staginuss et al.²⁹ and an assessment of variable importance within the model and the Akaike weight associated with each genetic effect. AIC corrected for finite sample size (AIC_c) weights were used to select the models for presentation when the minimum number of models whose weights sum to >95%.

AIC, Akaike information criterion; CBC, cannabichromene; CBD, cannabidiol; SE, standard error; THC, Δ -9-tetrahydrocannabinol.

of phenotypic expression.³⁴ When assessing the importance of epistatic or environmental effects, the Akaike information criterion (AIC) approach to conducting LCA appears to reduce bias.

Thus, we estimated the relative contribution of genetic effects (Appendix Table A2) on the inheritance of cannabinoid production. Using an I-T approach, we compared expected and observed mean cannabinoid concentration and standard error (SE) of the parental populations (P_1 , P_2) and two subsequent generations (F_1 , F_2). All analyses were computed using the software package SAGA v.2.0 (R 3.2.3; R Foundation for Statistical Computing 2015; <https://github.com/coleoguy/SAGA>; Appendix 3).

We first developed C-matrices for each trait to describe the potential contribution of CGEs to F_0 , F_1 , and F_2 cohort means. We used a C-matrix that was scaled to the midparent mean and included 7 of the 23 CGEs available in SAGA (Appendix Table A2), thus presenting the first study of cannabinoid inheritance with lowered risk that results are biased by experimenter selection of the CGEs tested. For each CGE, we have calculated coefficients for four potential crosses, considering only mixed sex cohorts.

SAGA reduced the C-matrix to include only the CGEs that can be partitioned with the four available cohorts (P_1 , P_2 , F_1 , F_2). Then, SAGA generated all possible models (maximum number of models = number

FIG. 1. Results from replicate analyses of the genetic architecture of THC concentration in *Cannabis sativa*. **(A, C)** Traditional line cross analysis plot of THC concentration with respect to proportion of the genome inherited from CBD-rich parent (P_1). If THC concentration was inherited in a Mendelian manner, the points should fall on the dotted line and the F_1 and F_2 data points should fall directly on top of each other. **(B, D)** Model-weighted parameter estimates for the genetic architecture underlying THC concentration in *C. sativa*. Points are sized based on v_i scores and indicate the magnitude of the genetic effects. Whiskers indicate the unconditional SEs. Only parameter estimate >0 , where $v_i > 0.01$ and $w_i > 0.001$, was included. The genetic effects axis describes the weighted parameter estimates calculated by SAGA in R.³¹ Data presented in **(A)** and **(B)** are from de Meijer et al.²; data presented in **(C)** and **(D)** are from Staginnus et al.²⁹ Model acronyms include autosomal additive (Aa), autosomal dominant (Ad), complex additive (Ca), maternal effect additive (Mea), maternal effect dominance (Med), autosomal additive by additive epistasis (AaAa), and autosomal dominance by dominance epistasis (AdAd); same acronyms are explained in Appendix Table A2. CBD, cannabidiol; SE, standard error; THC, Δ -9-tetrahydrocannabinol.

of cohorts minus 2df) and performed a weighted least-squares regression.³¹ SAGA produced parameter ($\pm$ SE) estimates along with AIC_C scores to adjust for model complexity, goodness of fit, and sample size.³⁵ This approach (formerly termed the joint-scaling test) allowed us to distinguish among the seven CGEs. Once all models were evaluated, SAGA calculated Δ AIC_C, comparing each model AIC_C score to the minimum AIC_C score calculated across all possible models. Models were ranked according to Δ AIC_C, and an Akaike weight (w_i) was generated.

With quantitative tools, namely the Castle–Wright estimator,^{36,37} one can estimate the minimum number of genes involved in a quantitative trait before investing significant resources in mapping studies of those quantitative trait loci (QTL). When the likely CGEs included an additive model, we applied the Castle–Wright estimator³³ to estimate the effective number of loci involved in THC, CBD, and CBC trait expression ($\hat{n}_e$), underlying differences in cannabinoid profiles between cultivars, based on the parental means ($\bar{X}_{P_1}$, $\bar{X}_{P_2}$) and variances ($\sigma_{P_1}^2$, $\sigma_{P_2}^2$) as well as the variances of the F₁ and F₂ hybrid offspring ($\sigma_{F_2}^2$, $\sigma_{F_1}^2$) using the following equation:

$$\hat{n}_e = \frac{(\bar{X}_{P_1} - \bar{X}_{P_2})^2 - \sigma_{P_1}^2 - \sigma_{P_2}^2}{8(\sigma_{F_2}^2 - \sigma_{F_1}^2)}$$

This method assumes: (1) genes have an equal and additive effect on phenotype; and (2) these alleles are at unlinked loci and are stable and dependent on DNA sequence (i.e., not epigenetic). When assumptions are violated, the Castle–Wright method generally underestimates the effective number of loci involved.^{38,39} The estimated number of genetic factors controlling CBD and THC was greater than 20 in the 2014 data set. Since the Castle–Wright estimator performs poorly when the number of genetic factors is large,⁴⁰ we did not report these values.

Results

There was a moderately high amount of model selection uncertainty for the inheritance of THC concentration in both data sets. The best model in both data sets, with high Akaike weight, considered autosomal additive effects (Aa, $w_i=0.29$, $w_i=0.442$, respectively; Table 1; Fig. 1). In 2003, the second-best model, with approximately half the Akaike weight of the Aa model ($w_i=0.26$) considered both additive and autosomal dominant ge-

netic effects (Aa+Ad), and the third-best model included cytotype additive genetic effects (Ca; $w_i=0.146$). Moreover, in 2014, there were three second-best models (Ca+AaAa, Ca+Aa, Aa+AaAa) that included Aa, Ca, and autosomal additive by additive epistatic (AaAa) effects. The w_i scores and v_i scores suggest that Aa and Ca are both important contributors to mean THC concentration in 2014 lineages. AaAa CGEs may also be an important but small contributor to line means, since AaAa appeared in the top three models identified by w_i scores in both data sets and had a relatively high v_i score in 2014.

There was a moderate amount of model selection uncertainty in both data sets in the inheritance of CBD concentration, after considering three models to describe the 2003 data set and four models to describe the 2014 data set (Table 2). The best and second-best models flip-flopped in rank order among the two data sets. Whereas the best model in 2014 was additive ($w_i=0.82$), the best model in the 2003² data set was a model that considered cytotype additive effects (Ca, $w_i=0.56$; Table 2; Fig. 2). The second-best model in 2003 was Aa ($w_i=0.37$), whereas the second-best model, with a relatively low Akaike weight, in 2014 was the Ca model ($w_i=0.075$).

Table 2. Parameter Estimates ($\pm$ Unconditional Standard Error) for Models of the Genetic Architecture of Cannabinoid Inheritance: Cannabidiol

Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)			
(1) de Meijer et al. ²		Ca	Aa	AaAa	
Ca	0.558	1.40 (0.07)			
Aa	0.374		2.77 (0.16)		
AaAa	0.040			−2.79 (0.29)	
Variable importance ->		0.560	0.391	0.040	
Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)			
(2) Staginnus et al. ²⁹		Aa	Ca	Ad	Mea
Aa	0.817	48.79 (3.67)			
Ca	0.075		28.63 (3.96)		
Aa, Ad	0.043	42.95 (1.98)		11.37 (1.98)	
Mea	0.026				31.63 (5.77)
Variable importance ->		0.884	0.098	0.044	0.027

Cannabinoid inheritance is described for CBD based on data from either (1) de Meijer et al.² or (2) Staginnus et al.²⁹ and an assessment of variable importance within the model and the Akaike weight associated with each genetic effect. AIC corrected for finite sample size (AIC_C) weights were used to select the models for presentation when the minimum number of models whose weights sum to >95%.

FIG. 2. Results from replicate analyses of the genetic architecture of CBD concentration in *Cannabis sativa*. **(A, C)** Traditional line cross analysis plot of CBD concentration with respect to proportion of the genome inherited from THC-rich parent (P₁). If CBD concentration was inherited in a Mendelian manner, the points should fall on the dotted line and the F₁ and F₂ data points should fall directly on top of each other. **(B, D)** Model-weighted parameter estimates for the genetic architecture underlying CBD concentration in *C. sativa*. Points are sized based on v_i scores and indicate the magnitude of the genetic effects. Whiskers indicate the unconditional SEs. Only parameter estimate >0 , where $v_i > 0.01$ and $w_i > 0.001$, was included. The genetic effects axis describes the weighted parameter estimates calculated by SAGA in R.³¹ Data presented in **(A)** and **(B)** are from de Meijer et al.,² whereas data presented in **(C)** and **(D)** are from Staginnus et al.²⁹

There was a small amount of model selection uncertainty in both data sets in the inheritance of THC:CBD ratio, after considering seven models to describe each of the 2003 and 2004 data sets. The best model in the 2003 data set was a model that considered additive and additive maternal effects ($w_i = 0.98$) and the distant second-best model considered dominant epistatic effects (AdAd) and additive maternal effects ($w_i = 0.10$). In the 2004 data set, the best model considered only additive effects ($w_i = 0.63$) and the second-best model considered additive and dominant effects (Aa, Ad; $w_i = 0.094$). These results pair closely with the conclusions of de Meijer et al.²

There was a moderate amount of model selection uncertainty in 2003 and no model selection uncertainty in 2014 for the inheritance of CBC concentration (Table 3). In 2003, the best model was additive ($w_i = 0.62$), and the single variable in the model was

an important variable ($v_i = 0.63$) and had a relatively large parameter estimate (Table 3; Fig. 3). The second-best model included both autosomal dominant and additive maternal effects (Ad, Mea; $w_i = 0.25$), which were both moderately important variables (for both, $v_i = 0.25$) but with low parameter estimates. The third-best model involved epistasis among autosomal additive genetic factors (AaAa, $w_i = 0.12$), and the single variable in the model had relatively low importance ($v_i = 0.12$) but a relatively large parameter estimate. In 2014, the only model that explained chemical variation among parental, F₁ and F₂ populations was one that included both autosomal additive and dominance effects (Aa, Ad; $w_i = 0.99$), which were both important variables in the only model (for both, $v_i = 0.99$) and large parameter estimates.

Differences in THC, CBD, and CBC concentrations were predicted to be controlled by 3.56, 1.22, and 1.68

Table 3. Parameter Estimates ($\pm$ Unconditional Standard Error) for Models of the Genetic Architecture of Cannabinoid Inheritance: Cannabichromene

Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)			
(1) de Meijer et al. ²		Aa	Ad	Mea	AaAa
Aa	0.625	−0.26 (0.02)			
Ad, Mea	0.247		−0.004 (0.05)	−0.15 (0.004)	
AaAa	0.118				−0.26 (0.03)
Variable importance ->		0.628	0.250	0.248	0.121
Model	AIC _C weight (<i>w_i</i>)	Estimate (SE)			
(2) Staginnus et al. ²⁹		Aa		Ad	
Aa, Ad	0.999	−1.15 (0.002)		0.79 (0.004)	
Variable importance ->		0.999		0.999	

Cannabinoid inheritance is described for CBC based on data from either (1) de Meijer et al.² or (2) Staginuss et al.²⁹ and an assessment of variable importance within the model and the Akaike weight associated with each genetic effect. AIC corrected for finite sample size (AIC_c) weights were used to select the models for presentation when the minimum number of models whose weights sum to >95%.

factors, respectively, in 2003. Thus, one to two genes controlled the expression of CBD and CBC concentration and three to four genes controlled the expression of THC concentration. Difference in CBC expression was predicted to be controlled by 0.78 factors for the 2014 population, and therefore, this trait is likely controlled by a single gene.

Discussion

Here, we show that inheritance of cannabinoids is far more complex than previously appreciated.^{2,4,5,28,29} In fact, not only are cannabinoid concentrations polygenic but we have also documented that they are influenced by maternal and cytoplasmic genetic effects. These complex genetic models alter the variance components of the one-locus, two-allele model in fundamental and enlightening ways, as detailed below. By using an I-T approach to identify the best genetic models, rather than a typical LCA approach to elucidate cannabinoid inheritance, we have had more power to predict variation in cannabinoids. However, it is difficult to disentangle the differences in these results because the genetic algorithms of SAGA are unclear in the original article and R documentation. The role of genetic variation on the inheritance of THC:CBD ratio is well described (and our modeling efforts confirmed the general consensus developed in the literature that the ratio is largely an additive inheritance

process in *C. sativa*).^{2,5} Importantly, it appears as though there is a trade-off between the production of THCA and CBDA, which influences the relative cannabinoid abundance. Nevertheless, how absolute cannabinoid abundance is controlled by genetic factors still requires significant research. Our model predicts that the absolute abundance of cannabinoids, especially THC and CBD, is controlled by different genetic architectures. This discrepancy between the models of relative and absolute abundance hinders careful and predictable breeding efforts of cannabis cultivars, and experimentation is required to reconcile these two conclusions. Because we did not have access to data from unselected parental populations, reciprocal crosses for F₁ generations, or backcrossed offspring,^{2,33} the available data were less likely to yield complex genetic effects. Yet, despite the reduced probability of detecting complex genetic effects and nonadditivity, we found evidence for complexity in the inheritance of all three cannabinoids. Therefore, the elucidated patterns are likely important and understudied aspects of cannabinoid inheritance. For example, altered CBD and THC concentration appears to result from cytoplasmic effects; maternal effects may explain CBC concentrations; and allelic dominance likely influences THC and CBC concentrations. Earlier studies, using traditional analytical approaches, did not explore phenotypic variance explained by the polygenic nature of these traits and non-additive genetic effects and so their conclusions were limited to additive genetic effects.^{2–4,28} By identifying genetic effects on cannabinoid concentrations, we expect our results will make it easier to predict and control cannabinoid production.

Population-level studies involve limited parental germplasm and thus provide information of CGEs acting on that population alone (i.e., results are not globally generalizable). Our use of two complementary data sets with unique parental genotypes, and the comparison of our results with a third study using unique parental genotypes,⁵ highlights those CGEs common to multiple breeding populations. This confirms that cannabinoid abundance is, in part, controlled by a non-functional CBDA synthase, in contrast to a single-locus additive gene, as originally proposed.^{2,29} Furthermore, we detected the presence of multiple genes acting on the concentration of THC in both 2003² and 2014,²⁹ consistent with de Meijer's original quantitative model² and the mapped "cannabinoid quantity" QTL.⁵ However, we present only summarized data because the raw data set was a gift from de Meijer et al.^{2,29} We

FIG. 3. Results from replicate analyses of the genetic architecture of CBC concentration in *Cannabis sativa*. **(A, C)** Traditional line cross analysis plot of CBC concentration with respect to proportion of the genome inherited from CBC-depauperate parent (P_1). If CBC concentration was inherited in a Mendelian manner, the points should fall on the dotted line and the F_1 and F_2 data points should fall directly on top of each other. **(B, D)** Model-weighted parameter estimates for the genetic architecture underlying CBC concentration in *C. sativa*. Points are sized based on v_i scores and indicate the magnitude of the genetic effects (listed on the x-axis). Whiskers indicate the unconditional SE (note that in **D**, SE is obscured by the point size). Only parameter estimate >0 , where $v_i > 0.01$ and $w_i > 0.001$, was included. The genetic effects axis describes the weighted parameter estimates calculated by SAGA in R.³¹ Data presented in **(A)** and **(B)** are from de Meijer et al.,² whereas data presented in **(C)** and **(D)** are from Staginnus et al.²⁹ CBC, cannabichromene.

encourage readers who want to explore our approach to contact the authors of these studies directly.

CBD is likely controlled by at least one major gene and a series of smaller genetic effects. In contrast to previous research, cytogenetic effects appear to contribute to the inheritance of CBD in both populations, suggesting a role of mitochondrial or plastid genomes in its expression (Fig. 3; Appendix Table A2). Comparative sequencing of chloroplast and mitochondrial DNA noted the utility of extranuclear genomes in differentiating among fiber and drug cultivars,⁴¹ which are differentiated by THC concentrations but often also differ in CBD concentration.²³ However, this remains understudied because other studies have removed “sequences obviously originating from organelles or ribosomal

RNA.”⁴² Finally, CBC concentrations are likely the product of two major genes, one additive and one dominant, given the consistent role of these two CGEs across both data sets.

Our analyses also detected differences in the role of CGEs among interbreeding populations, suggesting that convergent evolution is responsible for some of the chemical diversity within *C. sativa* populations. For instance, where cytotype and epistatic additive effects appear to play a role in the expression of THC concentration in the 2014 populations,²⁹ they play more minor roles in the 2003 population.² Maternal effects influenced CBC and chemotype expression in the 2003 data set,² whereas they were not included in the model in the 2014 data set.²⁹ These differences in

genetic effects may help explain the broad difficulties for hemp breeders when attempting to limit the production of THC, and for cannabis growers in the production of consistent concentrations of cannabinoids between and within crop varieties. Furthermore, more general epistatic effects have not been considered and may contribute to fluctuating chemotype expression.

The Castle–Wright estimator is commonly used to explore the genetic basis of trait evolution and has been commonly used to evaluate crop traits that have likely involved adaptation through altered expression of second metabolites (e.g., resistance to fungal infections, insect resistance, and heat tolerance^{43–45}). Previous QTL and modeling approaches suggest that the Castle–Wright estimator may be robust for traits with few loci, even if assumptions are not met.^{40,46,47} Our Castle–Wright estimator results for THC, CBD, and CBC in 2003² are consistent with the independent CGE analyses run in SAGA, in terms of our expectations of the number of genes involved in a trait. However, the results differ dramatically between the two data sets, suggesting that populations may differ dramatically in the mechanisms of inheritance and the number of genes involved.

Our results suggest that variation in cannabinoid expression was not previously fully described. We hypothesize that nonadditive genetic effects play prominent roles in cannabinoid genetics. A formal test of integration, whether phenotypic,⁴⁸ genetic,^{49,50} or developmental,⁵¹ would make significant advances in our understanding of cannabinoid inheritance. Thus, new crosses that map loci relative to psychoactive chemicals and pathways in the F₁ and F₂ generations of diverse chemotypes⁵² will complement ongoing efforts to characterize the genome of this plant.^{53,54}

Acknowledgments

The authors gratefully acknowledge the Natural Sciences and Engineering Research Council (NSERC) Discovery Grants program (Nos. 402305-2011 to L.C., and 2015-04405 to S.A.S.) and the Ryerson University for funding; the previous authors for their original work and shared data.

Author Disclosure Statement

No competing financial interests exist.

References

- Gepts P. Origins of plant agriculture and major crop plants. In: Tolba MK (ed). Our fragile world: challenges and opportunities for sustainable development, EOLSS Publishers: Oxford, UK, pp. 627–637, 2001.
- de Meijer EPM, Bagatta M, Carboni A, et al. The inheritance of chemical phenotype in *Cannabis sativa* L. Genetics. 2003;163:335–346.
- de Meijer EPM, Hammond KM. The inheritance of chemical phenotype in *Cannabis sativa* L. (II): cannabigerol predominant plants. Euphytica. 2005;145:189–198.
- de Meijer EPM, Hammond KM, Micheler M. The inheritance of chemical phenotype in *Cannabis sativa* L. (III): variation in cannabichromene proportion. Euphytica. 2009;165:293–311.
- Weiblen GD, Wenger JP, Craft KJ, et al. Gene duplication and divergence affecting drug content in *Cannabis sativa*. New Phytol. 2015;208:1241–1250.
- Kang BC, Nahm SH, Huh JH, et al. An interspecific (*Capsicum annuum* x *C. chinense*) F₂ linkage map in pepper using RFLP and AFLP markers. Theor Appl Genet. 2001;102:531–539.
- Mondego JMC, Vidal RO, Carazzolle MF, et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. BMC Plant Biol. 2011;11:30.
- Lewis RS, Bowen SW, Keogh MR, et al. Three nicotine demethylase genes mediate nornicotine biosynthesis in *Nicotiana tabacum* L.: functional characterization of the CYP82E10 gene. Phytochemistry. 2010;71:1988–1998.
- Yadav HK, Shukla S, Singh SP. Genetic divergence in parental genotypes and its relation with heterosis, F₁ performance and general combining ability (GCA) in opium poppy (*Papaver somniferum* L.). Euphytica. 2007;157:123–130.
- Ziegler J, Diaz-Chávez ML, Kramell R, et al. Comparative microarray analysis of morphine containing *Papaver somniferum* and eight morphine free *Papaver* species identifies an O-methyltransferase involved in benzyloisoquinoline biosynthesis. Planta. 2005;222:458–471.
- Mechoulam R. Marijuana chemistry. Science. 1970;168:1159–1166.
- Mechoulam R. Plant cannabinoids: a neglected pharmacological treasure trove. Br J Pharmacol. 2005;146:913–915.
- Pertwee RG. Cannabinoid pharmacology: the first 66 years. Br J Pharmacol. 2006;147:5163–5171.
- Turner CE, Bouwisma OJ, Billets S, et al. Constituents of *Cannabis sativa* L. XVIII—Electron voltage selected ion monitoring study of cannabinoids. Biomed Mass Spectrom. 1980;7:247–256.
- Adams R, Hunt M, Clark JH. Structure of cannabidiol, a product isolated from the marijuana extract of Minnesota wild hemp. J Am Chem Soc. 1940;62:196–200.
- Mechoulam R, Shvo Y. Hashish-I, the structure of cannabidiol. Tetrahedron. 1963;19:2073–2078.
- Gaoni Y, Mechoulam R. Isolation, structure and partial synthesis of an active constituent of hashish. J Am Chem Soc. 1964;86:1646–1647.
- Gaoni Y, Mechoulam R. The structure and synthesis of cannabigerol a new hashish constituent. Proc Chem Soc. 1964;March:82.
- Gaoni Y, Mechoulam R. Cannabichromene, a new active principle in hashish. Chem Comm. 1966;1:20–21.
- de Meijer EPM, Van Der Kamp HJ, Van Eeuwijk FA. Characterisation of Cannabis accessions with regard to cannabinoid content in relation to other plant characters. Euphytica. 1992;62:187–200.
- Fournier G. Les chemotypes du chanvre (*Cannabis sativa* L.). Intérêt pour un programme de selection. Agronomie. 1981;1:679–688.
- Fournier G, Paris MR. Le chanvre papetier (*Cannabis sativa* L.) cultivé en France: le point sur ses constituants (Paper hemp cultivated in France: a point on the constituents). Plantes Médicines Phytothérapie (Phytotherapeutic Medicinal Plants). 1979;13:116–121.
- Small E, Beckstead HD. Common cannabinoid phenotypes in 350 stocks of *Cannabis*. Lloydia. 1973;36:144–165.
- Akhmetova G, Sarsenbayev KN, Boribay E, et al. DNA heterogeneity in the *Cannabis* population of Chui valley. J Biotechnol. 2014;185:S62–S63.
- Gao C, Xin P, Cheng C, et al. Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. PLoS One. 2014;9:e110638.
- Taura F, Morimoto S, Shoyama Y. First direct evidence for the mechanisms of delta-1-tetrahydrocannabinolic acid biosynthesis. J Am Chem Soc. 1995;38:9766–9767.
- Taura F, Morimoto S, Shoyama Y. Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L. J Biol Chem. 1996;271:17411–17416.
- de Meijer EPM, Hammond KM, Sutton A. The inheritance of chemical phenotype in *Cannabis sativa* L. (IV): cannabinoid-free plants. Euphytica. 2009;168:95–112.

29. Staginuss C, Zörntlein S, de Meijer EPM. A PCR marker linked to a THCA synthase polymorphism is a reliable tool to discriminate potentially THC-rich plants of *Cannabis sativa* L. *J Forensic Sci.* 2014;59:919–926.
30. Mandolino G, Carboni A. Potential of marker-assisted selection in hemp genetic improvement. *Euphytica.* 2004;140:107–120.
31. Blackmon H, Demuth JP. An information-theoretic approach to estimating the composite genetic effects contributing to variation among generation means: moving beyond the joint-scaling test for line cross analysis. *Evolution.* 2016;70:420–432.
32. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sinauer Associates: Sunderland, MA, 1998.
33. Wade MJ. Epistasis, complex traits, and mapping genes. *Genetica.* 2001; 112:59–69.
34. Demuth JP, Wade MJ. Experimental methods for measuring gene interactions. *Annu Rev Ecol Syst.* 2006;37:289–316.
35. McQuarrie AD, Tsai CL. Regression and time series model selection. World Scientific Publishing Company: Singapore, 1998.
36. Castle WE. An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Proc Nat Acad Sci USA.* 1921;81:6904–6907.
37. Wright S. Evolution and the genetics of populations. University of Chicago Press: Chicago, IL, 1968.
38. Zeng ZB. Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics.* 1992;131:987–1001.
39. Zeng ZB, Houle D, Cockerham CC. How informative is Wright's estimator of the number of genes affecting a quantitative character? *Genetics.* 1990;126:235–247.
40. Otto SP, Jones CD. Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics.* 2000;156:2093–2107.
41. Gilmore S, Peakall R, Robertson J. Organelle DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*. *For Sci Int.* 2007;172:179–190.
42. Marks MD, Tian L, Wenger JP, et al. Identification of candidate genes affecting delta 9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J Exp Bot.* 2009;60:3715–3726.
43. Chen X, Line RF. Gene number and heritability of wheat cultivars with durable, high-temperature, adult-plant (HTAP) resistance and interaction of HTAP and race-specific seedling resistance to *Puccinia striiformis*. *Phytopathology.* 1995;85:573–578.
44. Adamczyk JJ, Meredith WR. Genetic basis for variability of Cry1Ac expression among commercial transgenic *Bacillus thuringiensis* (Bt) cotton cultivars in the United States. *J Cotton Sci.* 2004;8:17–23.
45. Jones IT. Inheritance of adult plant resistance to mildew in oats. *Ann Appl Biol.* 2008;109:187–192.
46. Peichel CL, Nereng KS, Ohgi KA, et al. The genetic architecture of divergence between three-spine stickleback species. *Nature.* 2001;414:901–905.
47. Westerbergh A, Doebley J. Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci. *Evolution.* 2002;56:273–283.
48. Herrera CM, Cerdá X, García MB, et al. Floral integration, phenotypic covariance structure and pollinator variation in bumblebee-pollinated *Helieborus foetidus*. *J Evol Biol.* 2002;15:108–121.
49. Cheverud JM. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution.* 1982;36:499–516.
50. Delph LF. The genetic integration of sexually dimorphic traits in the dioecious plant, *Silene latifolia*. In: Fairbairn DJ, Blanckenhorn WU, Szekely T (eds). Sex, size and gender roles. Evolutionary studies of sexual size dimorphism. Oxford University Press: Oxford, UK, pp. 115–123, 2007.
51. Schlichting CD, Pigliucci M. Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Inc.: Sunderland, MA, 1998.
52. Beavis WD. QTL analysis: power, precision, and accuracy. In: Paterson AH (ed). Molecular dissection of complex traits. CRC Press: Boca Raton, FL, pp. 145–162, 1997.
53. Lynch RC, Vergara D, Tittes S, et al. Genomic and chemical diversity in Cannabis. *Critical Reviews in Plant Sciences.* 2016;35:349–363.
54. Van Bakel H, Stout JM, Cote AG, et al. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 2011;12:R102.

Cite this article as: Campbell LG, Dufresne J, Sabatinos SA (2020) Cannabinoid inheritance relies on complex genetic architecture, *Cannabis and Cannabinoid Research* 5:1, 105–116, DOI: 10.1089/can.2018.0015.

Abbreviations Used

AIC = Akaike information criterion
 CBC = cannabichromene
 CBD = cannabidiol
 CBDA = cannabidiolic acid
 CBG = cannabigerol
 CGEs = composite genetic effects
 GC = gas chromatography
 I-T = information-theoretic
 LCA = line cross analysis
 SE = standard error
 THC = Δ -9-tetrahydrocannabinol
 THCA = tetrahydrocannabinolic acid

Appendix

APPENDIX TABLE A1. Summary of the Data Used for the Analysis

	De Meijer et al. ^{A1}		Staginnus et al. ^{A2}	
	<i>n</i>	Mean (SE)	<i>n</i>	Mean (SE)
THC				
P ₁ :daughters	9	7.59 (1.27)	1	92.91 (1.00 × 10 ⁻⁷)
P ₂ :sons	11	0.56 (0.36)	2	3.9 (0.2)
F ₁ (P ₂ × P ₁):mixed	546	2.56 (0.08)	12	38.91 (1.43)
F ₂ (F ₁ × F ₁):mixed	101	3.16 (0.10)	230	40.89 (2.09)
CBD				
P ₁ :daughters	11	3.97 (0.77)	2	85, 2.9
P ₂ :sons	12	0 (1.00 × 10 ⁻⁷)	1	0 (1.00 × 10 ⁻⁷)
F ₁ (P ₂ × P ₁):mixed	101	2.97 (0.10)	12	53.90 (1.40)
F ₂ (F ₁ × F ₁):mixed	546	2.61 (0.10)	230	50.47 (2.08)
CBC				
P ₁ :daughters	9	0 (1.00 × 10 ⁻⁷)	1	1.29 (1.00 × 10 ⁻⁷)
P ₂ :sons	10	0.32 (0.07)	2	3.6 (0.1)
F ₁ (P ₂ × P ₁):mixed	101	0.26 (0.01)	12	3.24 (0.07)
F ₂ (F ₁ × F ₁):mixed	118	0.28 (0.01)	230	2.84 (0.07)

Data gathered from de Meijer et al.^{A1} and Staginnus et al.^{A2} Data are presented with two decimal places. *Note:* From de Meijer et al.,^{A1} the P₁ group included the following genotypes (and their selfed offspring): 94.4.12, 94.5.2, 55.28.1, 94.4.12. The P₂ group included the following genotypes (and their selfed offspring): 55.22.7, 55.24.4, 55.28.1. From Staginnus et al.,^{A2} the P₁ group included the following genotypes (and their selfed offspring): 2005.40.69, whereas the P₂ group included 99.2.21.30.16.15.3 and 2001.22.6.20.14.1.

CBC, cannabichromene; CBD, cannabidiol; SE, standard error; THC, Δ-9-tetrahydrocannabinol.

APPENDIX TABLE A2. The Composite Genetic Effects, Their Abbreviation, and a Brief Definition That Were Compared Using an Information Theory Approach with the Analysis of Line Cross Data

Name	Abbreviation	Definition
Autosomal additive	Aa	When the combined effects of alleles at different loci are equal to the sum of their individual effects. Autosomal genes are located on one of the numbered, or non-sex, chromosomes. ^{A3,A4}
Autosomal dominance	Ad	In dominant genetic models, a single allele copy of the mutation is enough to cause expression of the trait. ^{A5}
Cytotype additive	Ca	Genetic variation in cytoplasmic genomes (i.e., the combined mitochondrial and plastid genomes) can influence trait expression. ^{A3,A6}
Maternal effect additive	Mea	Maternal effects arise when the genetic and environmental characteristics of a mother influence the phenotype of her offspring, beyond the direct inheritance of alleles. In this case, transgenerational expression of induced genes with additive genetic effects. ^{A3,A7}
Maternal effect dominance	Med	In this case, transgenerational expression of induced genes with dominance genetic effects. ^{A3}
Autosomal additive by additive epistasis	AaAa	The single locus additive value for a given locus (e.g., A) changes depending on the genotype at a second locus (e.g., B) and <i>vice versa</i> . When an alternate allele at locus A is present, it will result in smaller phenotype values when the B-locus is homozygous for the primary genotype and will result in large phenotype values when the B-locus is homozygous for the alternate genotype. ^{A8}
Autosomal dominance by dominance epistasis	AdAd	The single locus dominance genotype value for a given locus (e.g., A) changes depending on the genotype at a second locus (e.g., B) and <i>vice versa</i> . The single locus dominance genotypic value is underdominant when the alternate locus genotype is homozygous and overdominant when the alternate locus genotype is heterozygous. ^{A8}

For quantitative traits such as cannabinoid concentrations, their expression may be determined by genetic factors, environmental factors, and parental environment. Thus, when assessing the inheritance of chemotypes, a diversity of genetic models must be considered.^{A3}

APPENDIX 3. Sample Code Used to Analyze the Data. This Particular Code was Used to Analyze CBC Data from de Meijer et al.^{A1}

```
#open the SAGA R library
library(SAGA)
#create a data matrix called cbc that is 4 rows by 3
columns
cbc <- matrix(,4,3)

#give the columns names
colnames(cbc) <- c("Cohort.ID," "Mean," "SE")

#populate the cells of the matrix with data
cbc[1,] <- c(1, 0, 0.0000001)
cbc[2,] <- c(5, 0.321014118, 0.065581287)
cbc[3,] <- c(9, 0.257499841, 0.009045537)
cbc[4,] <- c(15, 0.27703782, 0.012610792)

#show the matrix to confirm its organized correctly
cbc

#Perform the analysis of the available crosses and
produce a graph to describe some of the results
results3 <- AnalyzeCrossesMM(cbc, graph=T, cex.
names=.8)

#In this example, there were three models for
which their summed weights accounted for 95% of
the  $\Delta AIC_C$  calculated the program. The value 3 will
change depending on the results above. Next, I rank
their model weights (so my table is arranged nicely and
I know for which models I should produce parameter
estimates).
order(results3[[3]])[1:3]

#Returns parameter estimates conditional on a sin-
gle model (where 1, 15, and 6 represent the three best
```

models identified in the previous step for this sample data set).

```
EvaluateModel(results3, 1, cex.names=.7, cex.-
main=.7)
EvaluateModel(results3, 15, cex.names=.7, cex.-
main=.7)
EvaluateModel(results3, 6, cex.names=.7, cex.-
main=.7)
```

```
#Examine the information returned from Analyze-
CrossesMM - first look at the names of the elements
in this list
names(results3)
```

```
#Access variable importance in element 4
results3[[4]]
```

```
#Convert a vector of  $\Delta AIC$  or  $\Delta AIC_C$  values to
model weights
AICtoMW(results3[[3]])
```

Appendix References

- A1. de Meijer EPM, Bagatta M, Carboni A, et al. The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*. 2003;163:335–346.
- A2. Staginnus C, Zörntlein S, de Meijer EPM. A PCR marker linked to a THCA synthase polymorphism is a reliable tool to discriminate potentially THC-rich plants of *Cannabis sativa* L. *J Forensic Sci*. 2014;59:919–926.
- A3. Wade MJ. Epistasis, complex traits, and mapping genes. *Genetica*. 2001;112:59–69.
- A4. Kempthorne O. An introduction to genetic statistics. John Wiley & Sons: New York, 1957.
- A5. Lynch M, Walsh B. *Genetics and analysis of quantitative traits*. Sinauer Associates: Sunderland, MA, 1998.
- A6. Dobler R, Rogell B, Budar F, Dowling DK. A meta-analysis of the strength and nature of cytoplasmic genetic effects. *J Evol Biol*. 2014;27:2021–2034.
- A7. Mousseau TA, Fox CW, eds. 1998. *Maternal Effects as Adaptations*. New York, USA: Oxford University Press.
- A8. Cheverud JM. Detecting epistasis among quantitative trait loci. In: Wolf JB, Brodie ED, Wade MJ (eds). *Epistasis and the evolutionary process*. Oxford University Press: Oxford, UK, pp. 58–81, 2000.