Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis

Daniela Vergara [Corresp., 1], **Cellene M Feathers** [1], **Ezra L Huscher** [1], **Ben Holmes** [2], **Jacob A Haas** [3], **Nolan C Kane** [1]

[1] Ebio, University of Colorado at Boulder, Boulder, Colorado, United States

[2] Centennial Seeds, Lafayette, Colorado, United States

[3] DabLogic, 5000 east 41st ave, Denver, Colorado, USA

Corresponding Author: Daniela Vergara
Email address: daniela.vergara@colorado.edu

The flowering plant *Cannabis sativa*, cultivated for centuries for multiple purposes, displays extensive variation in phenotypic traits in addition to its wide array of secondary metabolite production. Notably, *Cannabis* produces two well-known secondary-metabolite cannabinoids: cannabidiolic acid (CBDA) and delta-9-tetrahydrocannabinolic acid (THCA), which are the main products sought by consumers in the medical and recreational market. *Cannabis* has several suggested subspecies which have been shown to differ in chemistry, branching patterns, leaf morphology and other traits. In this study we obtained measurements related to phytochemistry, reproductive traits, growth architecture, and leaf morphology from 297 hybrid individuals from a cross between two diverse lineages. We explored correlations among these characteristics to inform our understanding of which traits may be causally associated. Many of the traits widely assumed to be strongly correlated did not show any relationship in this hybrid population. The current taxonomy and legal regulation within *Cannabis* is based on phenotypic and chemical characteristics. However, we find these traits are not associated when lineages are inter-crossed, which is a common breeding practice and forms the basis of most modern marijuana and hemp germplasms. Our results suggest naming conventions based on leaf morphology do not correspond to the chemical properties in plants with hybrid ancestry. Therefore, a new system for identifying variation within *Cannabis* is warranted that will provide reliable identifiers of the properties important for recreational and, especially, medical use.

Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis

Daniela Vergara[1*], Cellene M. Feathers[1], Ezra L. Huscher[1], Ben Holmes[2], Jacob A. Haas[1,3], and Nolan C. Kane[1*]

1. University of Colorado, Boulder, Department of Ecology and Evolutionary Biology
2. Centennial Seeds, 1216 Commerce Ct. No.7, Lafayette, CO 80026.
3. DabLogic, 5000 east 41st ave Denver CO 80216

*Correspondence: daniela.vergara@colorado.edu or nolan.kane@colorado.edu
University of Colorado Boulder
1900 Pleasant Street
Boulder, CO 80309

Keywords: cannabinoid, chemotype, geometric morphometrics, hemp, marijuana, shape

Abstract

The flowering plant *Cannabis sativa*, cultivated for centuries for multiple purposes, displays extensive variation in phenotypic traits in addition to its wide array of secondary metabolite production. Notably, *Cannabis* produces two well-known secondary-metabolite cannabinoids: cannabidiolic acid (CBDA) and delta-9-tetrahydrocannabinolic acid (THCA), which are the main products sought by consumers in the medical and recreational market. *Cannabis* has several suggested subspecies which have been shown to differ in chemistry, branching patterns, leaf morphology and other traits. In this study we obtained measurements related to phytochemistry, reproductive traits, growth architecture, and leaf morphology from 297 hybrid individuals from a cross between two diverse lineages. We explored correlations among these characteristics to inform our understanding of which traits may be causally associated. Many of the traits widely assumed to be strongly correlated did not show any relationship in this hybrid population. The current taxonomy and legal regulation within *Cannabis* is based on phenotypic and chemical characteristics. However, we find these traits are not associated when lineages are inter-crossed, which is a common breeding practice and forms the basis of most modern marijuana and hemp germplasms. Our results suggest naming conventions based on leaf morphology do not correspond to the chemical properties in plants with hybrid ancestry. Therefore, a new system for identifying variation within *Cannabis* is warranted that will provide reliable identifiers of the properties important for recreational and, especially, medical use.

Words: 229

Introduction

Phenotypic variation within and between populations is an important characteristic to consider for classification purposes. In particular, if phenotypic variation exists between different lineages within a species, then characterizing trait correlations can shed light on how they are inherited, whether they are controlled by the same genes, and if they can be used for taxonomic purposes. Wild populations may carry shared ancestral traits despite being independently inherited, and therefore it may appear as if these traits were associated. Additionally, selection may favor certain trait combinations.

The angiosperm species *Cannabis sativa* has been cultivated for millennia for a range of purposes (Li 1973; Li 1974; Russo 2007) and is currently, by some estimates, one

of the world's most valuable crops (Hutchison et al. 2019). Unfortunately, widespread legal issues have hindered *Cannabis* research.

One of the most notable characteristics of the *Cannabis* plant is its chemistry: the production of a family of molecules known as cannabinoids which are mainly produced and stored in the trichomes of female flowers (Gagne et al. 2012; Sirikantaramas et al. 2005). The most studied of these cannabinoids are cannabidiolic acid (CBDA), and delta-9-tetrahydrocannabinolic acid (THCA), which are produced by the enzymes CBDA and THCA synthases, respectively. These two synthases are found at the final stage of the biochemical pathway, along with cannabichromenic acid synthase, a third less well-studied synthase, that produces Cannabichomenic acid (CBCA; Page & Stout 2017). These three synthases use the same precursor molecule, cannabigerolic acid (CBGA; Laverty et al. 2019; Page & Boubakir 2014; Vergara et al. 2019). The genetic sequences for the three synthases are very similar and at least the genes encoding CBDA and THCA synthase are close in proximity (Weilblen et al. 2015), suggesting they may have originated from the same ancestor gene (Onofri et al. 2015; Padgitt-Cobb et al. 2019). Additionally, in vitro, each of these synthases can produce at least eight different compounds including THCA and CBDA in different ratios (Kovalchuk et al. 2020; Zirpel et al. 2018). These enzymes may be classified as "promiscuous enzymes" due to their considerable similarities, the fact they act on the same precursor molecule, and they can produce each other's compounds (Auldridge et al. 2006; Chakraborty et al. 2013; Franco 2011).

When heated, THCA and CBDA are converted into the neutral forms THC and CBD (Russo 2011), which interact with the human endocannabinoid system (Pertwee 1988; Pertwee 1997; Pertwee 2004). Both THC and CBD have medicinal (Russo 2011; Swift et al.

2013; Volkow et al. 2014) and economic value (Evans 2013; Kirsch 2018), but THC has been intensely selected by breeders and growers (Volkow et al. 2014) due to its psychoactive effects (ElSohly & Slade 2005). Studies have also found CBDA may have medicinal benefits in its acidic form (Takeda et al. 2008; Takeda et al. 2012). Differences in leaf size, leaf shape, plant size, and inflorescence size are used in the *Cannabis* industry to categorize plants and these morphological differences are thought to be useful predictors of cannabinoid content.

Currently recognized lineages within the genus *Cannabis* include the narrowleaf drug types, *C. sativa* ssp. *sativa*, the broadleaf drug type *C. sativa* ssp. *indica*, the northern Eurasian wild *C. sativa* ssp. *ruderalis,* and at least one lineage of hemp (Clarke & Merlin 2013). Among these subspecies, there is substantial phenotypic variation in the production of multiple cannabinoids (McPartland & Russo 2001; Russo et al. 2008; Russo & McPartland 2003) and terpenoids (de la Fuente et al. 2020; Orser et al. 2017; Reimann-Philipp et al. 2019), substantial genotypic variation (Kovalchuk et al. 2020; Lynch et al. 2016; Sawler et al. 2015; Vergara et al. 2016), and observed morphological variation in traits such as branching, internode length, and flowering time (Clarke & Merlin 2013).

The main classifications used in the modern *Cannabis* industry parlance are “indica”, “sativa”, and “hybrids”. Sativa plants are described as tall with narrow leaves and lighter density buds, allegedly producing high levels of THCA, and therefore have uplifting and stimulating psychedelic effects after consumption. Indica plants are described as short with broad leaves and dense buds, and produce high levels of both THCA and CBDA believed to produce a relaxing effect (Clarke & Merlin 2013; McPartland 2017; Vergara et al. 2016). Yet, the associations between these multiple traits have not previously been researched.

Other popular ideas suggest the important distinctions between *Cannabis* lineages related to the effects after consumption are due to differences in terpene profiles rather than to cannabinoid profiles. It may be that terpene profiles are more relevant to differences in *Cannabis* lineages than cannabinoids (de la Fuente et al. 2020; Orser et al. 2017; Reimann-Philipp et al. 2019). Crosses between "sativa" and "indica" plants are referred to as "hybrids" and these have variable phenotypes usually intermediate to the parents (Vergara et al. 2016). Finally, the hemp group has been traditionally used for industrial purposes such as fiber or oil production, however the legal definition of hemp includes any *Cannabis* plant with less than 0.3% THC by weight.

The colloquial naming convention of "indica" and "sativa" do not correspond to the scientific subspecies with similar names. Furthermore, these common distinctions do not reflect evolutionary relationships (Sawler et al. 2015; Schwabe & McGlaughlin 2019; Vergara et al. 2016). This misidentification can be particularly problematic for medical patients who are depending on reliable and consistent products.

Cannabis is dioecious (Divashuk et al. 2014; van Bakel et al. 2011), although monoecious plants exist, particularly in the hemp lineage (Hillig 2005; Peil et al. 2003). Dioecious varieties are common for medicinal and recreational purposes, and selection in domestication has been focused on females due to the production of cannabinoids (Gagne et al. 2012; Sirikantaramas et al. 2005), with strong selection against males and hermaphrodites. Many commercially important traits are expressed at maturity, and if breeders could predict their late-stage expression through correlations among these traits earlier during development, selection could be made sooner, accelerating breeding cycles. Also, if traits early in the development allowed for distinguishing between sexes, males

could be culled before pollen production and potential female pollination. This is important because females would undesirably divert energy to seeds instead of cannabinoids after being pollinated (Clarke & Merlin 2013).

In this study, we quantified 18 phenotypic traits of 297 individuals from a first-generation backcross (BC1) between a female 'Carmagnola' hemp and a male marijuana-type plant 'Afghan Kush'. Many of the morphological traits we measured are considered important by the *Cannabis* industry to characterize different plants. We predicted these traits would vary in the BC1, and therefore allow us to investigate the pattern of possible genetic correlations. Furthermore, we determined whether the association between morphological traits and cannabinoid chemistry could be used to characterize *Cannabis* lineages. Given that the current nomenclature is not supported scientific research (Lynch et al. 2016; Sawler et al. 2015; Vergara et al. 2016), it is possible that erroneous associations between morphological traits has contributed to the misnaming issues in the *Cannabis* industry.

Methods

BC cross and measurements

A cross was performed between a female, narrowleaf 'Carmagnola' hemp plant and a male broadleaf 'Afghan Kush' plant at Centennial Seeds in Lafayette Colorado (Figure 1A). One of the female first filial (F1) offspring of this cross was backcrossed to a male sibling of the parental male (Figure 1A). Two hundred ninety-seven individuals from this backcross population (BC1; Figure 1A) were started indoors on April 23, 2015 and on June 6, 2015,

159 when the plants were six weeks old, they were transplanted outdoors in a field in Boulder
160 County, Colorado.

161 Morphological measurements including height, stalk diameter, inner-node length,
162 petiole length, leaf length and width, among other measurements, were obtained at two
163 different time points during the growing cycle (Table S1). We chose these two time points,
164 one at the beginning and one at the end of the growing season, to provide information on
165 possible trait associations during the plant's development. The initial timepoint (IT) was
166 taken at six weeks old (June 6, 2015), and the final timepoint (FT) at 19 weeks old
167 (September 2, 2015) which corresponds to the beginning of the fall season. Additional
168 traits were measured at the FT including bud count, size of biggest bud, length of longest
169 branch, and number of buds on the longest branch (Table S1). At the FT, we also scanned a
170 representative leaf from each plant. Therefore, we had fully extended longest leaf (FELL)
171 measurements from both the IT and FT. At the FT, we determined the sex of all 297
172 individuals, and also measured the concentrations of three cannabinoids -THC, CBD, and
173 CBG- from 100 plants. Here we removed plants identified as male to avoid pollination, after
174 the representative leaf was scanned. Therefore, some of the measurements were taken on
175 fewer individuals than at the IT.

176

177 *Phenotypic trait statistical analyses*

178 To understand the change in individual phenotypes through the growing season, we
179 calculated the difference between the initial measurements and the final measurements
180 (delta Δ) for some traits. Specifically, we calculated Δ for the four traits that were measured
181 at both timepoints (Table S1). We then use these data to estimate the within-time point

182 correlations for both the IT and the FT, and between-time correlations. All correlations
183 were corrected with Bonferroni for multiple comparisons (Weisstein 2004). Finally, we
184 established whether any of the measured traits differed between males and females using
185 t-tests with sex as the explanatory variable.
186
187 *Leaf shape analysis*
188 We carried out a geometric morphometric analysis to develop a quantifiable
189 measure of leaf shape. Specifically, we placed landmark coordinates on each leaf picture
190 with the program TPS Dig2 (Rohlf 2006). We used ten landmarks from the first, central,
191 and last leaflet (Figure 1B) which covers the whole leaf structure. Additionally, we
192 measured the length and width of each of the leaves, counted the serration number in the
193 center leaflet, and counted the number of leaflets.
194 We used the R package *Geomorph* (Adams & Otárola-Castillo 2013) for all geometric
195 morphometric analysis, following the methodology of Vergara et al. (2017b). A Procrustes
196 analyses was used to remove variability caused by position, orientation, and size and to
197 quantify shape variation by superimposing the objects in a joint coordinate system. Then, a
198 Principal Component Analysis (PCA) was used to identify the orthogonal structure in the
199 data and to visually explore morphological variation among individuals.
200 We performed multiple statistical tests to understand whether leaf shape was
201 related to any of the other measured traits at both timepoints and between timepoints.
202 First, we implemented several multivariate analyses of variance (MANOVAs) with shape as
203 the response variable for each of the measured traits in both timepoints and Δ. We then
204 performed MANOVA models within each timepoint and Δ to understand whether the main

205 effects of each trait affected leaf shape. We corroborated the results using multivariate
206 multiple regressions.
207
208 *Cannabinoid concentration measurements*
209 The concentrations for the three cannabinoids -- CBG, THC, and CBD -- were
210 measured using gas chromatography on an SRI 86106 equipment with an MXT-35 column
211 using 197-209 mg of dried flower as described in (Brenneisen & ElSohly 1988). When
212 heated, the acidic compounds CBGA, THCA, and CBDA are turned into the neutral forms
213 CBG, THC, and CBD, which is the reason why gas chromatography only quantifies the
214 neutral forms of the compounds.
215 Given that the production of these three cannabinoids may be correlated because
216 they are part of the same biochemical pathway (Page & Boubakir 2014; Page & Stout 2017;
217 Vergara et al. 2019) and both CBDA and THCA synthases compete for the same precursor
218 molecule -CBGA-, we analyzed them using a Principal Component Analysis (PCA) to account
219 for multicollinearity and to avoid redundancies. We used a K-means cluster analysis on PC1
220 vs PC2 to visualize the different cannabinoid groups. We also added the total cannabinoid
221 concentration and measured the ratio of each cannabinoid over this total concentration
222 (Table S1).
223
224 *Statistical analyses*
225 We examined the associations between the production of each cannabinoid and
226 each of the measured traits at both timepoints and the Δ. We also used cannabinoids as the
227 explanatory variables for several MANOVA models to determine whether cannabinoid

production explained differences among the measured traits. We also corroborated the MANOVA results with multivariate multiple regressions, and correlated leaf shape to cannabinoid content to understand whether any association exists between those traits. Finally, we generated a variance-covariance matrix to establish the association within and between all phenotypic traits.

These data were added to the dryad repository (https://doi.org/10.5061/dryad.6t1g1jwxh). Statistical analyses, including leaf geometric morphometrics, were done using R (Team 2013) and the associated code is available on github (https://bit.ly/38DpE8D). All figures were generated in the R Studio platform version 1.1.383 (Team 2013) and enhanced with Adobe Illustrator 2019 (v23.0.6).

Results

Phenotypic trait statistical analyses (including males and females)

Our results show that some phenotypic traits from the IT (Table S2) are correlated with each other after correcting with Bonferroni for multiple comparisons. For example, height is significantly correlated to the number of branches and the number of nodes even though these two traits are not significantly correlated to each other (Table S2). The positive correlation between traits related to height such as number of nodes and number of branches is expected. In other words, it is expected that tall plants will have multiple branches and nodes. It is also expected that traits that are not related to height, such as leaf-related characteristics, lack a significant correlation.

Similarly, the FT also shows that some traits are correlated at this stage (Table S3). Some of the height-related traits show a significant correlation. For example, tall plants

have long side branches as well as thicker stalks. However, as expected, some traits lack association, such as stalk diameter and inflorescence number or size.

However, many of the significant associations within either the IT or FT are lost when both timepoints are correlated between them (Table S4). These various phenotypic traits are not predictive between time periods (Table S4); whether a young plant is tall or short is not indicative of the adult plant's height, and thus plants that are tall at the IT are not always the same ones that are tall at the FT. Therefore, we cannot establish whether, for example, tall plants also have thick stalks and numerous nodes since traits which are correlated while young, are not significantly correlated while adults. In other words, the plants exhibited different patterns of growth, irrespective of their initial size at the beginning of the growing season. The lack of correlations between the timepoints suggests that some of the statistically significant correlations may not be due to true biological variation but instead due to chance and to the multiple comparisons, despite correcting with Bonferroni.

The lack of significance between the Δ correlations when compared to either the IT or FT (Tables S2 and S3) suggest that some of these correlations may be spurious. The non-significant correlations between the traits and their Δ indicates that the measured characteristics do not follow a trend as they change over time. Therefore, the changes during the plant's lifetime impede future phenotypic predictions and the initial plant measurements cannot be used as an indication of future success as an adult, or how the phenotype will change during the plant's lifetime.

Similarly, these phenotypic traits are not different between males and females (Table S5). In other words, males cannot be distinguished from females with any of the physical characteristics that we measured in this study. However, some trait correlations

do differ between the sexes (Table S6), but again are not consistent between the timepoints. The only significantly different trait between both groups is the number of buds in the main branch, which was taken at the FT, where males have a larger average number of buds (35.75) compared to females (24.80). However, this comparison is between only four males and 19 females, as most most males were removed from the field before these measurements were taken, and therefore this result may again not hold any true biological meaning.

Leaf shape analysis

Our geometric morphometric analysis on leaf shape revealed that 82.3% of all variation in leaf shape is explained by the two first principal components (Figure 2). The deformation grids in the top left and bottom right corners show the extreme trends in leaf morphologies. Even though these morphologies are not seen in any individual plant, these are the tendencies of the leaves in these furthest points of the morphospace. The individuals on the top-left side of the morphospace tend to have shorter and broader leaves, and as seen in the deformation grid, the first and last leaflets are pointing outwards. On the other hand, individuals in the bottom right side of the morphospace tend to have a longer middle leaflet, and the first and last leaflets are clumped together pointing downward. Our morphometric analysis shows no significant relationship between leaf shape and the plant's sex. Therefore, both male and female plants can have similar leaf shapes.

Our results suggest there are some trait correlations that describe leaf shape, but these are not correlated to growth rates, plant size, branching architecture,

298 phytochemistry, or plant sex (Figure S1). It appears that there could be a within-leaf effect
299 because the FELL measurements correlate within them in the IT, and serration, leaf length,
300 and number of leaflets correlate with leaf shape in the FT. However, the leaf measurements
301 show no association between timepoints (Table S7).

302 The overall trend shows leaf shape is not explained by any of the plant traits
303 measured on either timepoint (Table S7). The lack of association between a particular leaf
304 shape and plant height, or any of the other plant traits, suggests that a tall plant can have
305 broad or narrow leaves or high or low cannabinoid levels.

306 Furthermore, the MANOVA models with leaf shape as the response variable (PC1
307 and PC2, Table S7) show that the only traits that may explain leaf shape correspond to
308 other leaf measurements which are leaf length, serration, and the number of leaflets from
309 that same leaf at the FT (Figure S1). These MANOVA results confirm the within-leaf effect
310 where these measured leaf traits are associated within the same leaf. The additional
311 models that include the main effects of multiple traits support the individual MANOVA
312 results given that no significant trend in any other phenotype explains leaf shape in the IT.
313 The FT results of the within-leaf effect are confirmed with a further model including all of
314 the leaf traits (leaf length: Wilk's λ= 0.975, F= 3.560, P=0.029740; serration: Wilk's λ=
315 0.967, F= 4.625, P<0.012; no. leaflets Wilk's λ = 0.8105, F= 32.483, P= 2.1e-13). However,
316 the only significant interaction effect was between leaf length and leaf width (Wilk's λ=
317 0.967, F= 4.739, P< 0.00095). The multivariate multiple regressions confirmed the
318 MANOVA results.
319
320 *Cannabinoid concentration measurements*

321 Two clear chemotype clusters were identified in the PCA of cannabinoid chemistry
322 determined by the antagonistic relationships between THC versus CBD and CBG (Figure 3).
323 PC1 and PC2 explain 94.2% of the variation and the two distinct groups identified in figure
324 3 (represented in triangles and diamonds) covary in a pronounced way, which show
325 covariation of the original variables with two clear chemotype clusters (Figure S2).
326 The positive loadings of both CBD and CBG on PC1 (0.71 and 0.70, respectively)
327 compared to the negative loading of THC (-0.02) indicates that when CBD and CBG tend to
328 increase, THC decreases and covaries in a different direction. Therefore, the value of PC1
329 increases when CBG or CBD increase. However, on PC2, both CBG and THC have a negative
330 loading (-0.31 and -0.91 respectively), indicating a high association, while CBD has a
331 positive loading (0.28). PC2 is primarily determined by THC given its high loading value. In
332 both PC1 and PC2, CBD and THC go in different directions.
333 Cannabinoid content showed no correlation with any of the other measured
334 phenotypic traits at either time point (IT or FT) nor with the Δs (Table S8). These results
335 were confirmed with the MANOVAs and multivariate multiple regressions.
336
337 *Leaf shape vs cannabinoid content*
338 We found no relationship between leaf shape and cannabinoid content using PC1 for
339 leaf shape and PC1 for cannabinoid variation (Figure 4). Therefore, leaf shape is not
340 predictive of cannabinoid content, and individuals that are high in a particular cannabinoid
341 can have elongated or short leaves.

Discussion

In this study we examined correlations among various phenotypic traits from a morphologically diverse first-generation backcross (BC1) population to understand whether these multiple traits covaried with each other. We interpret these patterns of correlations as being genetically-based due to our development of a diverse array of progeny of known parentage all grown in a common environment. Our results suggest these traits are not constrained by strong genetic correlations and the initial associations between the various morphological traits in the parent generation can be broken by recombination. The lack of apparent genetic correlation between these traits suggest they can be selected for independently. Therefore, these traits are inherited independently, and would evolve separately unless selection acts to increase or maintain correlations among them. The lack of genetic correlations between the morphological traits, was also reflected in a variable pattern of growth across the growing season. The dramatic trait changes and the distribution variation over the course of plant growth and development (Coleman et al. 1994) may explain the lack of correlations between the two timepoints. This lack of correlation could also signify phenotypic plasticity which is common in plants and may be a form of adaptation (Schlichting 1986; Sultan 1995). Finally, in other species such as sugarcane, yield has not been associated to characteristics in the stalk (ie. length, weight, diameter, number) nor to other traits such as plant height (Aitken et al. 2008; Kang et al. 1983), so these patterns are likely not unique to *Cannabis*.

The lack of sexual dimorphism in the measured traits for this study may be specific to this population, and particularly the measurements in the FT may be problematic due to the lack of males. Theoretical models suggest differences between males and females

365 particularly in wind-pollinated plants (Friedman & Barrett 2009). Additionally, sex
366 differences in *Cannabis* have been found in traits not measured here, such as
367 photosynthetic rates and senescence (Dzhaparidze 1969; Geber et al. 2012). Future studies
368 may include these traits to examine differences between the sexes.

369 The PCA analysis facilitates the examination of shape variation for each structure
370 independently (Adams et al. 2004), allowing us to distinguish differences in leaf shape
371 (Figures 1B and 2). As size is removed during the Procrustes superimposition, it does not
372 determine the variation of the first principal component (PC1) as it does in traditional
373 morphometrics, assuring that the main source of variation explored is shape. With this
374 geometric morphometric analysis, we found that leaf shape is not related to sex (Figure 2),
375 cannabinoid production (Figure 4), or to multiple other phenotypic traits (Table S8),
376 suggesting all of these traits segregate independently. However, we did find within-leaf
377 associations between shape, leaf length, serration, and number of leaflets (Figure S1)
378 suggesting that within a single leaf some characteristics may be related to each other.

379 Because THC and CBD have attracted the most research and popular attention of all
380 the cannabinoids and their synthases both compete for the same precursor CBGA, the
381 relationship between these compounds revealed significant patterns. PC1 and PC2 for
382 cannabinoid variation (Figure 3) explain 94.2% of the variation due to the fact that there
383 are only three variables that compose the original matrix. We used a principle component
384 analysis because of the high association between these enzymes which compete for the
385 same precursor molecule (Page & Boubakir 2014; Page & Stout 2017), have similar
386 chemical structures (Brenneisen 2007; Flores-Sanchez & Verpoorte 2008) and genetic
387 sequences (Onofri et al. 2015; Vergara et al. 2019), and may exemplify “sloppy enzymes”

(Auldridge et al. 2006; Chakraborty et al. 2013; Franco 2011). Our results show that despite the evident competition for the same precursor, as seen with the negative correlation between THC and CBD (Figure S2C), all of these compounds can be present together. Additionally, CBG is always seen in lower levels when compared to THC and CBD (Figure S2), implying that THCA and CBDA synthases are efficiently converting CBGA into THCA and CBDA respectively in this population.

Studies suggest that THC has been selected for by breeders and growers and that varieties have been bread for higher THC potency (ElSohly et al. 2016; Volkow et al. 2014). Our results confirm these studies given that THC is always produced in higher quantities than CBD (Figure S2), implying that THCA synthase may be a better competitor than CBDA synthase in this population.

Variation in THC production is probably a result of gene sequence variation (Onofri et al. 2015), expression levels, and gene copy number variation (Vergara et al. 2019), and there are multiple genes throughout the genome associated with its production (Grassa et al. 2018; Laverty et al. 2019). However, expression of these genes could be due to environmental effects such as cultivation conditions (Elzinga et al. 2015), which have not yet been quantified. Even though the parent plants were grown under different conditions than the BC1 offspring, all of the offspring were grown under the same conditions minimizing the environmental effects on the expression of these genes.

Although some correlations among traits are significant and make biological sense, the traits that are associated with purported groups (i.e. *indica* and *sativa*) within *Cannabis* are not correlated because of shared genetic basis. Therefore, trait correlations observed are due to either shared ancestry, in the case of comparisons among subspecies or other

411 major lineages, or correlated selection, in the case of modern hybrids. In other words,
412 correlations between leaf shape and phytochemistry may not be due to causal
413 relationships, but rather because breeders have intentionally (or unintentionally) selected
414 for these trait combinations. If these traits were associated due to shared ancestry or
415 correlated selection, their association can be broken by recombination.

416 This is particularly noticeable in most of the modern cultivars which are hybrids
417 from the supposed two main groups. Therefore, our study also suggests that common
418 assumptions about associations between leaf shape and chemistry may exacerbate the
419 misnaming problems of *Cannabis* varieties by the industry (Sawler et al. 2015; Vergara et
420 al. 2016). Given the lack of association between cannabinoids and other morphological
421 traits, the accepted standards for categorizing *Cannabis* types by the industry are deeply
422 flawed because their naming convention is based on sets of traits that could be
423 disassociated to each other. Additionally, other studies have shown that name is not
424 indicative of cannabinoid potency or overall chemical composition (Elzinga et al. 2015),
425 and that varieties are grouped based on reported flavors and aromas, regardless of genetic
426 relationships misclassifying closely related individuals (de la Fuente et al. 2020). This
427 misnaming problem in the *Cannabis* industry for both varieties and groupings ('sativa' and
428 'indica') is greatly magnified by the fact that scientists can only study the *Cannabis*
429 produced by the federal government despite its inferiority in potency and diversity, and the
430 fact that it does not reflect the products distributed in consumer markets (Schwabe et al.
431 2019; Vergara et al. 2017a). It is crucial for *Cannabis* researchers to disseminate accurate
432 information to the public. This is not being done adequately because the scientific literature
433 is not effectively informing public policy, medical decisions, or providing correct

information on harm reduction (Hutchison et al. 2019). This lack of information has major ramifications for growers, breeders, regulators, and consumers, particularly for medical patients who must understand what they are consuming to achieve the greatest benefit for their individual needs.

In order to improve the quality and efficacy of the *Cannabis* consumed by medical patients, it is important that unbiased, accurate, and precise chemotype testing should be made mandatory. However, testing facilities do not have universally established standards, as cannabinoid measurements vary widely across laboratories (Jikomes & Zoorob 2018), and there are no supervising institutions that oversee testing entities or their methodologies, making differences in cannabinoid reporting inevitable .

Conclusions

The fact that most of the phenotypic traits are not genetically correlated has significant implications for both *Cannabis* breeders and commercial growers. If these traits are not linked, as previously thought, then it is possible to select for new combination of traits when breeding for novel varieties. This expands the possibility of generating varieties with a unique combination of traits providing unforeseen medicinal and industrial value. Future breeding can be done to maximize combinations of these traits.

Acknowledgments

We would like to thank J. A. Fuentes for help with the morphometric analysis, figure construction and for comments on the manuscript, to C. Pogoda, A. Schwabe, and R. Miller

for comments on the manuscript, and to A. Holloway and two anonymous reviewers for their informative comments which greatly improved this paper.

Author Contributions

D.V. and C.M.F. analyzed the data and wrote the first draft of the manuscript; E.L.H. calculated the Gmatrix, and measured BC1 plants; B.H. donated, grew, and cared for all plants-parentals, F1, and BC1; J.H. grew, cared, and measured the BC1 plants; D.V and N.C.K. conceived and lead the project and measured BC1 plants. All authors contributed to manuscript preparation.

Legends Supplementary Tables and Figures

Table S1. All data. Data used for all analyses with each plant (column 1) its sex (column 2), unique id (column 3), the IT date (column 4), measurements for the IT (columns 5-12), the FT (column 13), measurements for the FT (columns 14-21), deltas for four measurements (columns 22-25), flower weight for cannabinoid measurement (column 26), the addition of all cannabinoids (column 27), the percentage of each cannabinoid over the total (columns 28-30), the ratio of each cannabinoid to each other (columns 31-33), the PC values for the cannabinoids (columns 34-36), leaf measurements for the FT (columns 37-39), and the leaf shape PC values (columns 41-56).

Table S2 Within IT and delta (Δ) correlations. Correlations within the initial time point and Δ for the eight traits measured in time point one and the four Δ changes

Table S3 Within FT and delta (Δ) correlations. Correlations within the final time point and Δ for the eight traits measured in the final timepoint, the four Δ changes and the four leaf-morphologic traits from the leaves used for geometric morphometrics (leaflet number, serration in the middle leaflet, leaf width and leaf length).

Table S4 Between time-point correlations. Correlations between the eight traits from time-point one against the eight traits from time point two and against the four leaf-morphologic traits from the leaves used for geometric morphometrics (leaflet number, serration in the middle leaflet, leaf width and leaf length).

Table S5. t-tests between sex and phenotypic traits. t-tests between sex and the phenotypic traits measured on both timepoints including leaf shape measurements in the IT, FT, and Δ.

Table S6. Difference in trait correlations between sex. Difference between the correlations of the various phenotypic traits between males and females in the IT, FT, and Δ.

Table S7. MANOVA for leaf shape against all phenotypic traits. MANOVA associations between PC1 and PC2 for leaf shape against all phenotypic traits from the IT, FT, and Δ.

Table S8. Correlations between three cannabinoids CBG, CBD, and THC, and all phenotypic traits. Correlations between the three cannabinoids all phenotypic traits from the IT, FT, and Δ.

Table S8. GMatrix for variance and covariance between all phenotypic traits measured. Variances and covariances between all phenotypic traits measured at both timepoints.

Figure S1. PC1 for shape versus three traits. Leaf length is not correlated to PC1 **(A)**, while serration in the middle leaflet **(B)** and number of leaflets **(C)** are marginally and significantly correlated, respectively. Each panel at the top has the deformation grids at the minimum (left) and maximum (right) values for length **(A)**, serration in the middle leaflet **(B)**, and number of leaflets **(C).** Males are shown in squares, females in open circles. The insets show the p and r values and the variance and covariance between both variables, also given in the variance-covariance matrix (Table S8).

Figure S2. Relationships between percent THC, CBD, and CBG. A and **B** show a positive correlation between CBG vs CBD, and THC vs CBG, respectively. **C** displays the negative linear relationship between THC and CBD. Notice the differences in the axis representing the disparity in production of the different cannabinoids. Males are shown in squares, females in open circles. The insets show the p and r values and the variance and covariance between both variables, also given in the variance-covariance matrix (Table S8).

References

Adams DC, and Otárola-Castillo E. 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4:393-399.

Adams DC, Rohlf FJ, and Slice DE. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* 71:5-16.

Aitken K, Hermann S, Karno K, Bonnett G, McIntyre L, and Jackson P. 2008. Genetic control of yield related stalk traits in sugarcane. *Theoretical and Applied Genetics* 117:1191-1203.

Auldridge ME, McCarty DR, and Klee HJ. 2006. Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Current opinion in plant biology* 9:315-321.

Brenneisen R. 2007. Chemistry and analysis of phytocannabinoids and other Cannabis constituents. *Marijuana and the Cannabinoids*: Springer, 17-49.

Brenneisen R, and ElSohly MA. 1988. Chromatographic and spectroscopic profiles of Cannabis of different origins: Part I. *Journal of Forensic Science* 33:1385-1404.

Chakraborty S, Minda R, Salaye L, Dandekar AM, Bhattacharjee SK, and Rao BJ. 2013. Promiscuity-based enzyme selection for rational directed evolution experiments. *Enzyme Engineering: Methods and Protocols*:205-216.

Clarke R, and Merlin M. 2013. *Cannabis: evolution and ethnobotany*: Univ of California Press.

Coleman JS, McConnaughay KD, and Ackerly DD. 1994. Interpreting phenotypic variation in plants. *Trends in Ecology & Evolution* 9:187-191.

de la Fuente A, Zamberlan F, Ferrán AS, Carrillo F, Tagliazucchi E, and Pallavicini C. 2020. Relationship among subjective responses, flavor, and chemical composition across more than 800 commercial cannabis varieties. *Journal of Cannabis Research* 2:1-18.

Divashuk MG, Alexandrov OS, Razumova OV, Kirov IV, and Karlov GI. 2014. Molecular cytogenetic characterization of the dioecious Cannabis sativa with an XY chromosome sex determination system. *Plos One* 9:e85118.

Dzhaparidze L. 1969. Sex in plants, part II. Biochemical and physiological sex differences in dioecious plants: problem of influencing sex formation. Jerusalem: IPST Press.

ElSohly MA, Mehmedic Z, Foster S, Gon C, Chandra S, and Church JC. 2016. Changes in cannabis potency over the last 2 decades (1995–2014): analysis of current data in the United States. *Biological psychiatry* 79:613-619.

ElSohly MA, and Slade D. 2005. Chemical constituents of marijuana: the complex mixture of natural cannabinoids. *Life sciences* 78:539-548.

Elzinga S, Fischedick J, Podkolinski R, and Raber J. 2015. Cannabinoids and terpenes as chemotaxonomic markers in cannabis. *Nat Prod Chem Res* 3:2.

Evans D. 2013. The economic impacts of marijuana legalization.

Flores-Sanchez IJ, and Verpoorte R. 2008. Secondary metabolism in cannabis. *Phytochemistry reviews* 7:615-639.

Franco OL. 2011. Peptide promiscuity: an evolutionary concept for plant defense. *FEBS letters* 585:995-1000.

Friedman J, and Barrett SC. 2009. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Annals of Botany* 103:1515-1527.

Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, and 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* 109:12811-12816.

Geber MA, Dawson TE, and Delph LF. 2012. *Gender and sexual dimorphism in flowering plants*: Springer Science & Business Media.

Grassa CJ, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, Schwartz CJ, and Weiblen GD. 2018. A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv*. 10.1101/458083

Hillig KW. 2005. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genetic Resources and Crop Evolution* 52:161-180. 10.1007/s10722-003-4452-y

Hutchison KE, Bidwell LC, Ellingson JM, and Bryan AD. 2019. Cannabis and Health Research: Rapid Progress Requires Innovative Research Designs. *Value in Health*.

Jikomes N, and Zoorob M. 2018. The cannabinoid content of legal cannabis in Washington state varies systematically across testing facilities and popular consumer products. *Scientific Reports* 8:4519.

Kang MS, Miller JD, and Tai P. 1983. Genetic and Phenotypic Path Analyses and Heritability in Sugarcane 1. *Crop Science* 23:643-647.

Kirsch J. 2018. Taxation and Welfare in the Cannabis Industry: Evidence from Colorado Edibles 2014-2016.

Kovalchuk I, Pellino M, Rigault P, van Velzen R, Ebersbach J, R. Ashnest J, Mau M, Schranz M, Alcorn J, and Laprairie R. 2020. The Genomics of Cannabis and Its Close Relatives. *Annual review of plant biology* 71.

Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, Deikus G, Sebra R, Hughes TR, and Page JE. 2019. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Research* 29:146-156.

Li HL. 1973. An archaeological and historical account of cannabis in China. *Economic Botany* 28:437-448.

Li HL. 1974. Origin and use of Cannabis in Eastern Asia; Linguistic-cultural implications. *Economic Botany* 28:293-301. 10.1007/bf02861426

Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, deCesare K, Land DP, and Kane NC. 2016. Genomic and Chemical Diversity in Cannabis. *Critical Reviews in Plant Sciences* 35:349-363. 10.1080/07352689.2016.1265363

McPartland JM. 2017. Cannabis sativa and Cannabis indica versus “Sativa” and “Indica”. *Cannabis sativa L-Botany and Biotechnology*: Springer, 101-121.

McPartland JM, and Russo EB. 2001. Cannabis and cannabis extracts: greater than the sum of their parts? *Journal of Cannabis Therapeutics* 1:103-132.

Onofri C, de Meijer EPM, and Mandolino G. 2015. Sequence heterogeneity of cannabidiolic-and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. *Phytochemistry*.

Orser C, Johnson S, Speck M, Hilyard A, and Afia I. 2017. Terpenoid Chemoprofiles Distinguish Drug-type Cannabis sativa L. Cultivars in Nevada. *Natural Products Chemistry and Research* 6. 10.4172/2475-7675.1000304s

Padgitt-Cobb LK, Kingan SB, Wells J, Elser J, Kronmiller B, Moore D, Concepcion G, Peluso P, Rank D, Jaiswal P, Henning J, and Hendrix DA. 2019. A phased, diploid assembly of the Cascade hop (Humulus lupulus) genome reveals patterns of selection and haplotype variation. *bioRxiv*:786145. 10.1101/786145

Page JE, and Boubakir Z. 2014. Aromatic prenyltransferase from Cannabis. Google Patents.

Page JE, and Stout JM. 2017. Cannabichromenic acid synthase from Cannabis sativa. Google Patents.

Peil A, Flachowsky H, Schumann E, and Weber WE. 2003. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (Cannabis sativa L.). *Theoretical and Applied Genetics* 107:102-109.

Pertwee RG. 1988. The central neuropharmcology of psychotropic cannabinoids. *Pharmacology & therapeutics* 36:189-261.

Pertwee RG. 1997. Pharmacology of cannabinoid CB1 and CB2 receptors. *Pharmacology & therapeutics* 74:129-180.

Pertwee RG. 2004. Pharmacological and therapeutic targets for Δ 9 tetrahydrocannabinol and cannabidiol. *Euphytica* 140:73-82.

Reimann-Philipp U, Speck M, Orser C, Johnson S, Hilyard A, Turner H, Stokes AJ, and Small-Howard AL. 2019. Cannabis Chemovar Nomenclature Misrepresents Chemical and Genetic Diversity; Survey of Variations in Chemical Profiles and Genetic Markers in Nevada Medical Cannabis Samples. *Cannabis and Cannabinoid Research*.

Russo EB. 2007. History of cannabis and its preparations in saga, science, and sobriquet. *Chemistry & Biodiversity* 4:1614-1648. 10.1002/cbdv.200790144

628 Russo EB. 2011. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage
629 effects. *British Journal of Pharmacology* 163:1344-1364. 10.1111/j.1476-5381.2011.01238.x
630 Russo EB, Jiang H-E, Li X, Sutton A, Carboni A, del Bianco F, Mandolino G, Potter DJ, Zhao Y-X, Bera S,
631 Zhang Y-B, Lue E-G, Ferguson DK, Hueber F, Zhao L-C, Liu C-J, Wang Y-F, and Li C-S. 2008.
632 Phytochemical and genetic analyses of ancient cannabis from Central Asia. *Journal of*
633 *Experimental Botany* 59:4171-4182. 10.1093/jxb/ern260
634 Russo EB, and McPartland JM. 2003. Cannabis is more than simply Δ9-tetrahydrocannabinol.
635 *Psychopharmacology* 165:431-432.
636 Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, and Myles S. 2015. The Genetic
637 Structure of Marijuana and Hemp. *Plos One* 10:e0133292.
638 Schlichting CD. 1986. The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and*
639 *Systematics* 17:667-693.
640 Schwabe AL, Hansen CJ, Hyslop RM, and McGlaughlin ME. 2019. Research grade marijuana supplied by
641 the National Institute on Drug Abuse is genetically divergent from commercially available
642 Cannabis. *bioRxiv*:592725. 10.1101/592725
643 Schwabe AL, and McGlaughlin ME. 2019. Genetic tools weed out misconceptions of strain reliability in
644 Cannabis sativa: implications for a budding industry. *Journal of Cannabis Research* 1:3.
645 10.1186/s42238-019-0001-1
646 Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, and Shoyama Y. 2005.
647 Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is
648 secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology* 46:1578-
649 1582.
650 Sultan SE. 1995. Phenotypic plasticity and plant adaptation. *Acta botanica neerlandica* 44:363-383.
651 Swift W, Wong A, Li KM, Arnold JC, and McGregor IS. 2013. Analysis of cannabis seizures in NSW,
652 Australia: cannabis potency and cannabinoid profile. *Plos One* 8:e70052.
653 Takeda S, Misawa K, Yamamoto I, and Watanabe K. 2008. Cannabidiolic acid as a selective
654 cyclooxygenase-2 inhibitory component in cannabis. *Drug metabolism and disposition* 36:1917-
655 1921.
656 Takeda S, Okajima S, Miyoshi H, Yoshida K, Okamoto Y, Okada T, Amamoto T, Watanabe K, Omiecinski
657 CJ, and Aramaki H. 2012. Cannabidiolic acid, a major cannabinoid in fiber-type cannabis, is an
658 inhibitor of MDA-MB-231 breast cancer cell migration. *Toxicology letters* 214:314-319.
659 Team RC. 2013. R: A language and environment for statistical computing.
660 van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, and Page JE. 2011. The draft genome
661 and transcriptome of Cannabis sativa. *Genome Biology* 12. 10.1186/gb-2011-12-10-r102
662 Vergara D, Baker H, Clancy K, Keepers KG, Mendieta JP, Pauli CS, Tittes SB, White KH, and Kane NC.
663 2016. Genetic and Genomic Tools for Cannabis sativa. *Critical Reviews in Plant Sciences* 35:364-
664 377. 10.1080/07352689.2016.1267496
665 Vergara D, Bidwell LC, Gaudino R, Torres A, Du G, Ruthenburg TC, deCesare K, Land DP, Hutchison KE,
666 and Kane NC. 2017a. Compromised External Validity: Federally Produced Cannabis Does Not
667 Reflect Legal Markets. *Scientific Reports* 7:46528.
668 Vergara D, Fuentes JA, Stoy KS, and Lively CM. 2017b. Evaluating shell variation across different
669 populations of a freshwater snail. *Molluscan Research* 37:120-132.
670 Vergara D, Huscher EL, Keepers KG, Givens RM, Cizek CG, Torres A, Gaudino R, and Kane NC. 2019. Gene
671 copy number is associated with phytochemistry in Cannabis sativa. *bioRxiv*:736181.
672 10.1101/736181
673 Volkow ND, Baler RD, Compton WM, and Weiss SRB. 2014. Adverse Health Effects of Marijuana Use.
674 *New England Journal of Medicine* 370:2219-2227. 10.1056/NEJMra1402309
675 Weisstein EW. 2004. Bonferroni correction. *https://mathworld wolfram com/*.

Zirpel B, Kayser O, and Stehle F. 2018. Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L. *Journal of biotechnology* 284:17-26.

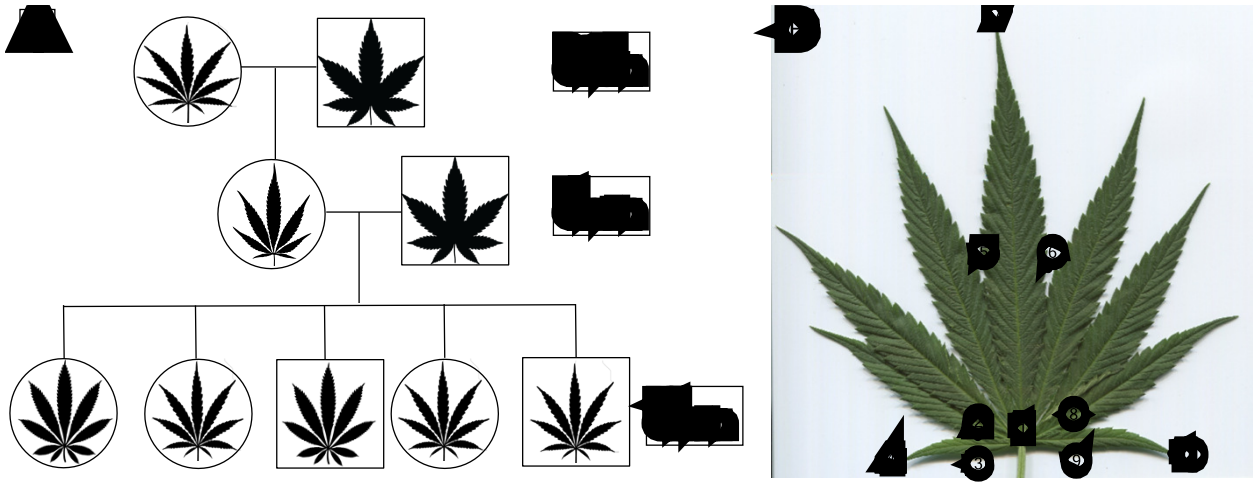
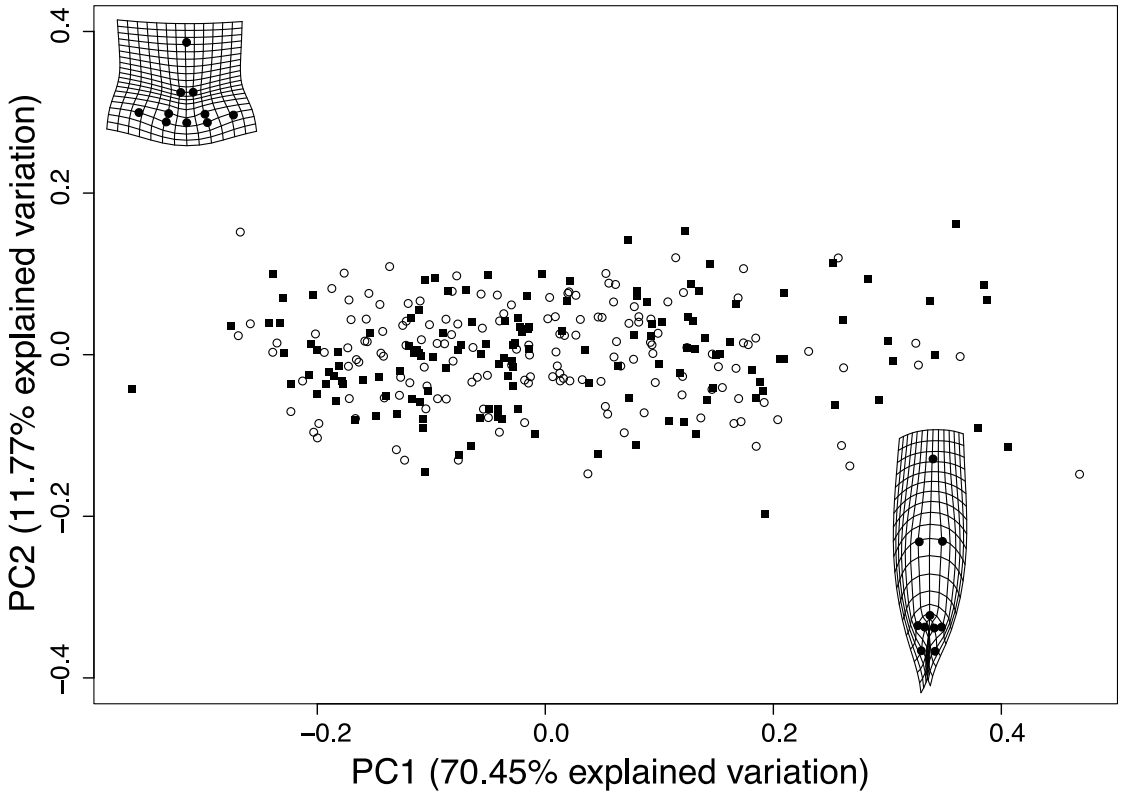


Figure 1. Pedigree and landmarks. Panel A: Pedigree of a first-generation backcross (BC1) between a male marijuana-type Afghan Kush and a female Carmagnola hemp. The F1 generation was backcrossed with a brother from the original male Afghan Kush to produce the BC1 Generation. Panel B: Exemplar leaf depicting the ten points used for leaf shape analysis. The ten points measured the first, central, and last leaflets.

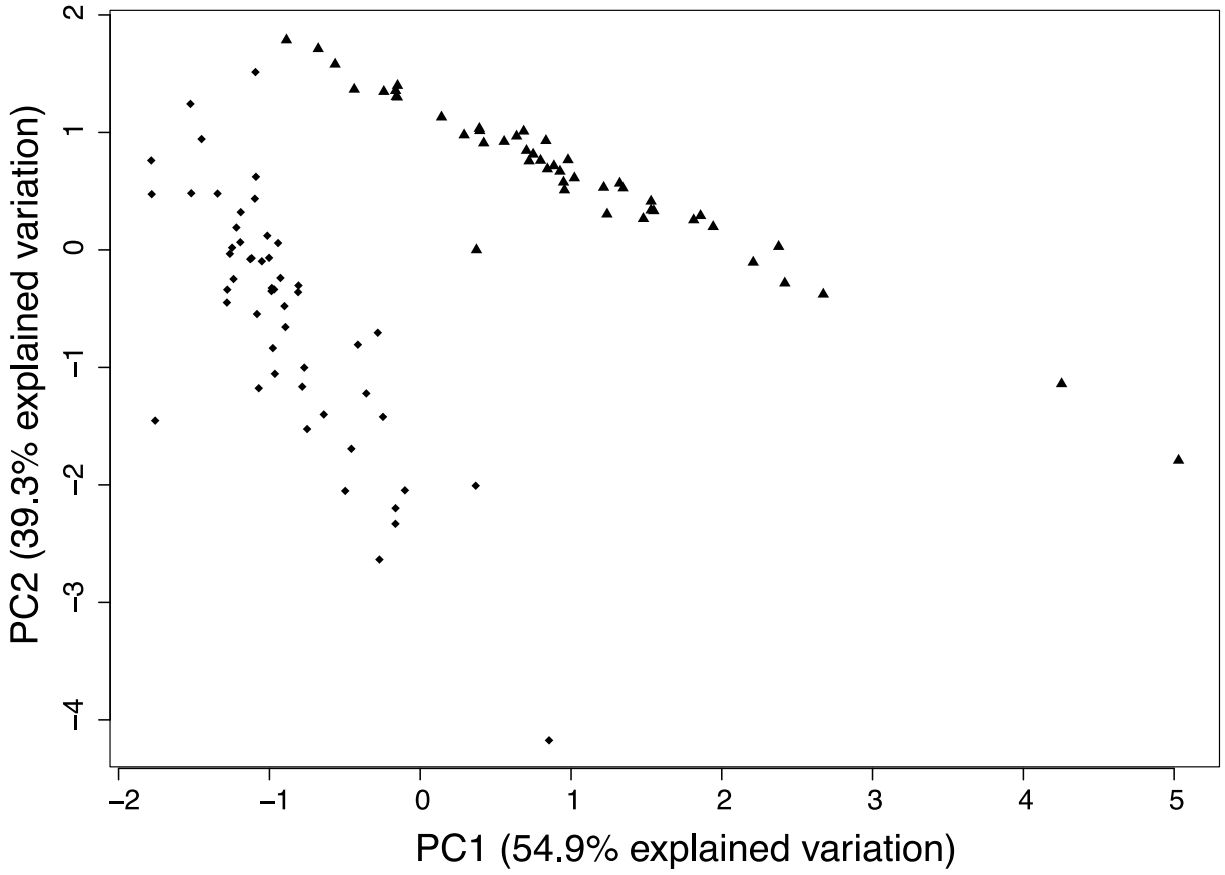


691

692 **Figure 2**: **Geometric morphometric analysis of leaf shape.** The two first PC explain 82.3% of the leaf shape
693 variation, which is not related to sex (males are squares, females are open circles). The deformation grids
694 mostly show the leaf deformations on PC1 which contains most of the variation. The grids show that
695 individuals in the top left have squatter, broader leaves with the first and last leaflets pointing outwards,
696 while those on the bottom right side of the morphospace have thinner leaves with a long mid-leaflet and the
697 two outer leaflets lumped together pointing downwards.
698

699

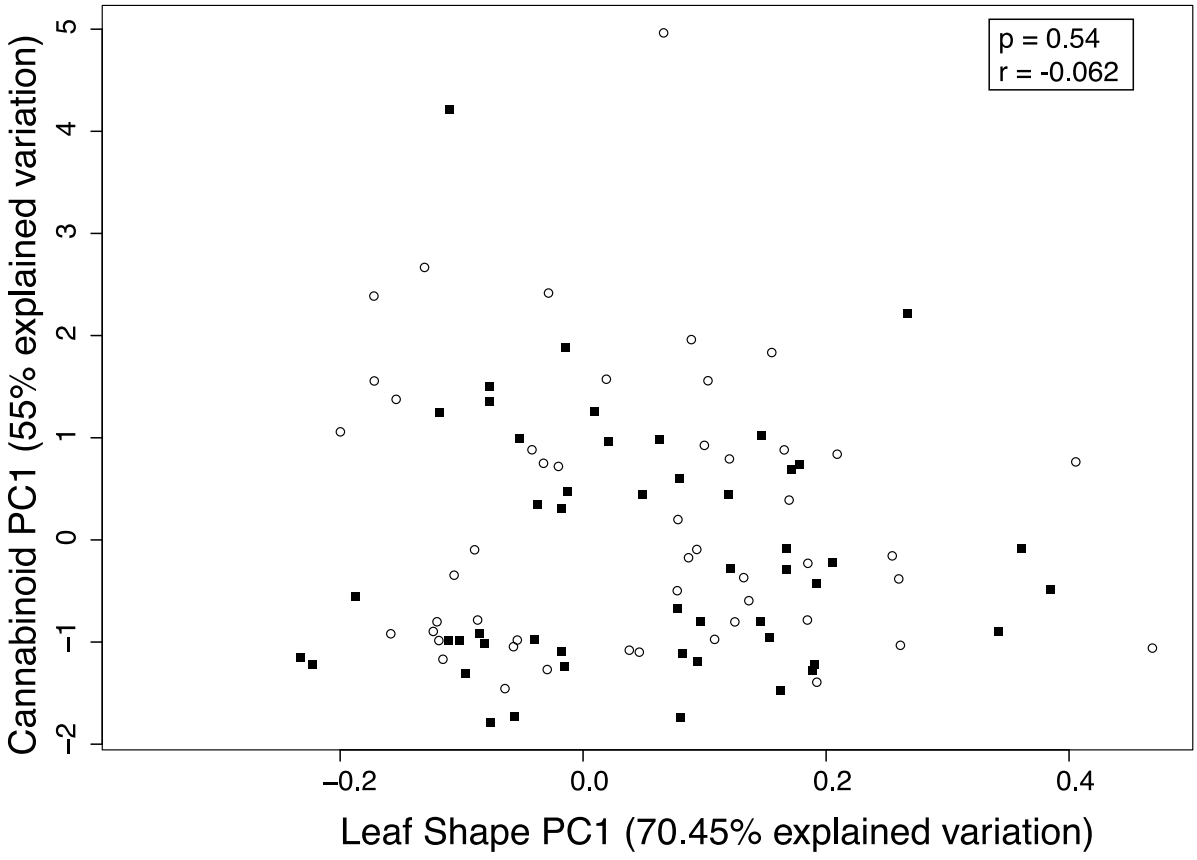
700



701

702 **Figure 3. PC1 and PC2 for cannabinoid variation.** PC1 and PC2 explain 94.2% of the overall cannabinoid
703 variation for the three cannabinoids measured. There are two clear groups in the graph, squares and
704 triangles, which correspond to the two clear cannabinoid clusters. The overall trend shows that the squares
705 have low CBG and CBD and high THC. The triangles show high CBD and low THC.
706

707



708

709 **Figure 4: Correlation between leaf shape (PC1) and cannabinoid variation (PC1)**. Leaf shape is not
710 significantly to cannabinoid variation even at the most extreme points in the morphospace. Males are shown
711 in squares and females in open circles.
712

Figure 1

Figure 1. Pedigree and landmarks.

Panel A: Pedigree of a first-generation backcross (BC1) between a male marijuana-type Afghan Kush and a female Carmagnola hemp. The F1 generation was backcrossed with a brother from the original male Afghan Kush to produce the BC1 Generation. Panel B: Exemplar leaf depicting the ten points used for leaf shape analysis. The ten points measured the first, central, and last leaflets.

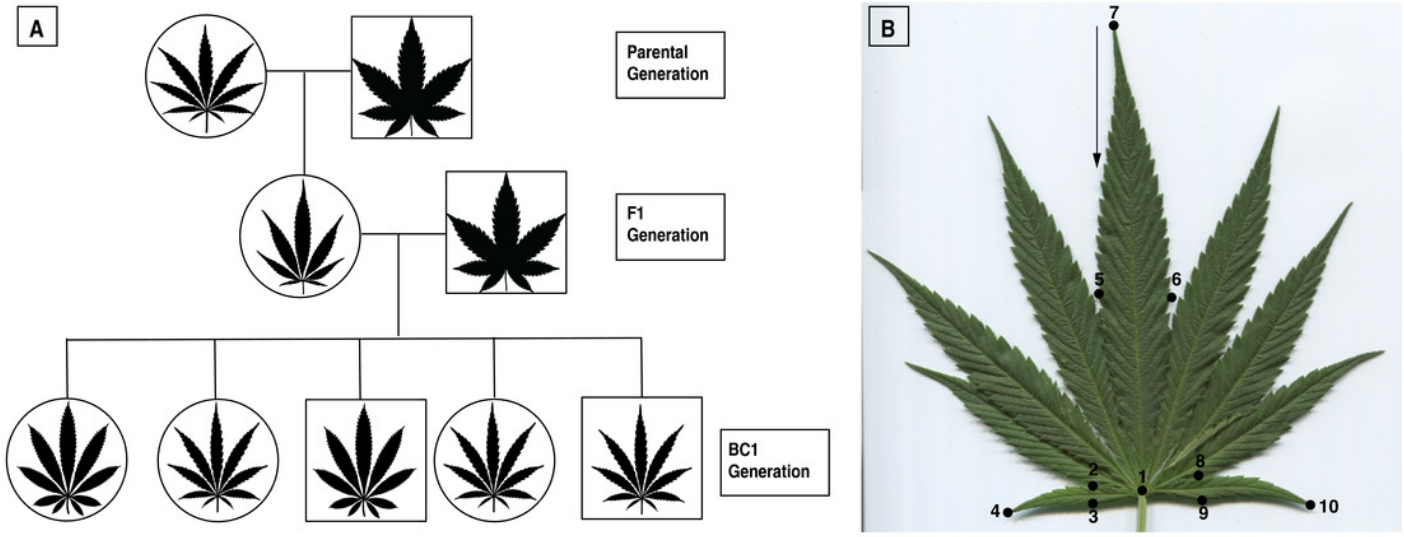


Figure 2

Figure 2: Geometric morphometric analysis of leaf shape.

The two first PC explain 82.3% of the leaf shape variation, which is not related to sex (males are squares, females are open circles). The deformation grids mostly show the leaf deformations on PC1 which contains most of the variation. The grids show that individuals in the top left have squatter, broader leaves with the first and last leaflets pointing outwards, while those on the bottom right side of the morphospace have thinner leaves with a long mid-leaflet and the two outer leaflets lumped together pointing downwards.

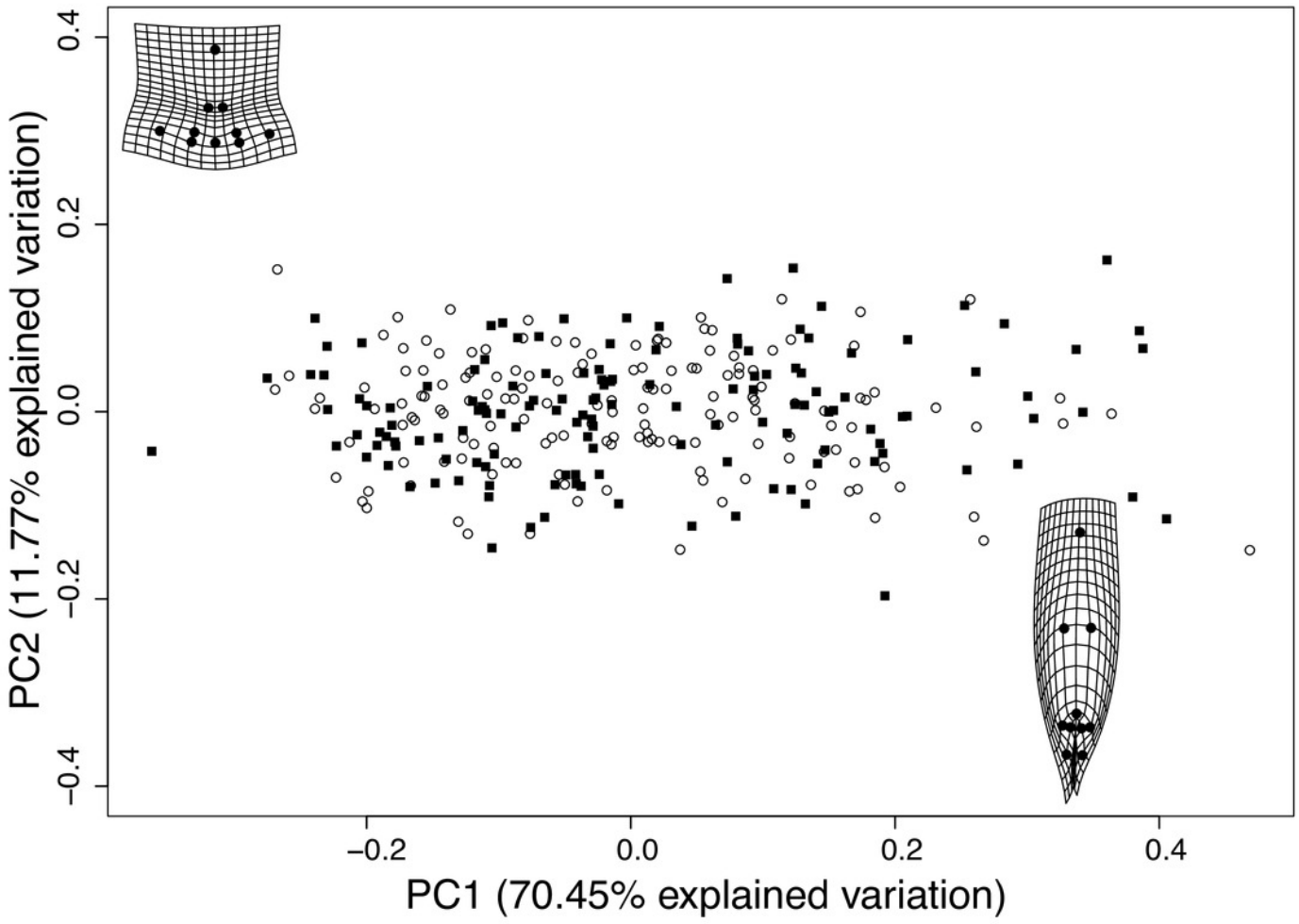


Figure 3

Figure 3. PC1 and PC2 for cannabinoid variation.

PC1 and PC2 explain 94.2% of the overall cannabinoid variation for the three cannabinoids measured. There are two clear groups in the graph, squares and triangles, which correspond to the two clear cannabinoid clusters. The overall trend shows that the squares have low CBG and CBD and high THC. The triangles show high CBD and low THC.

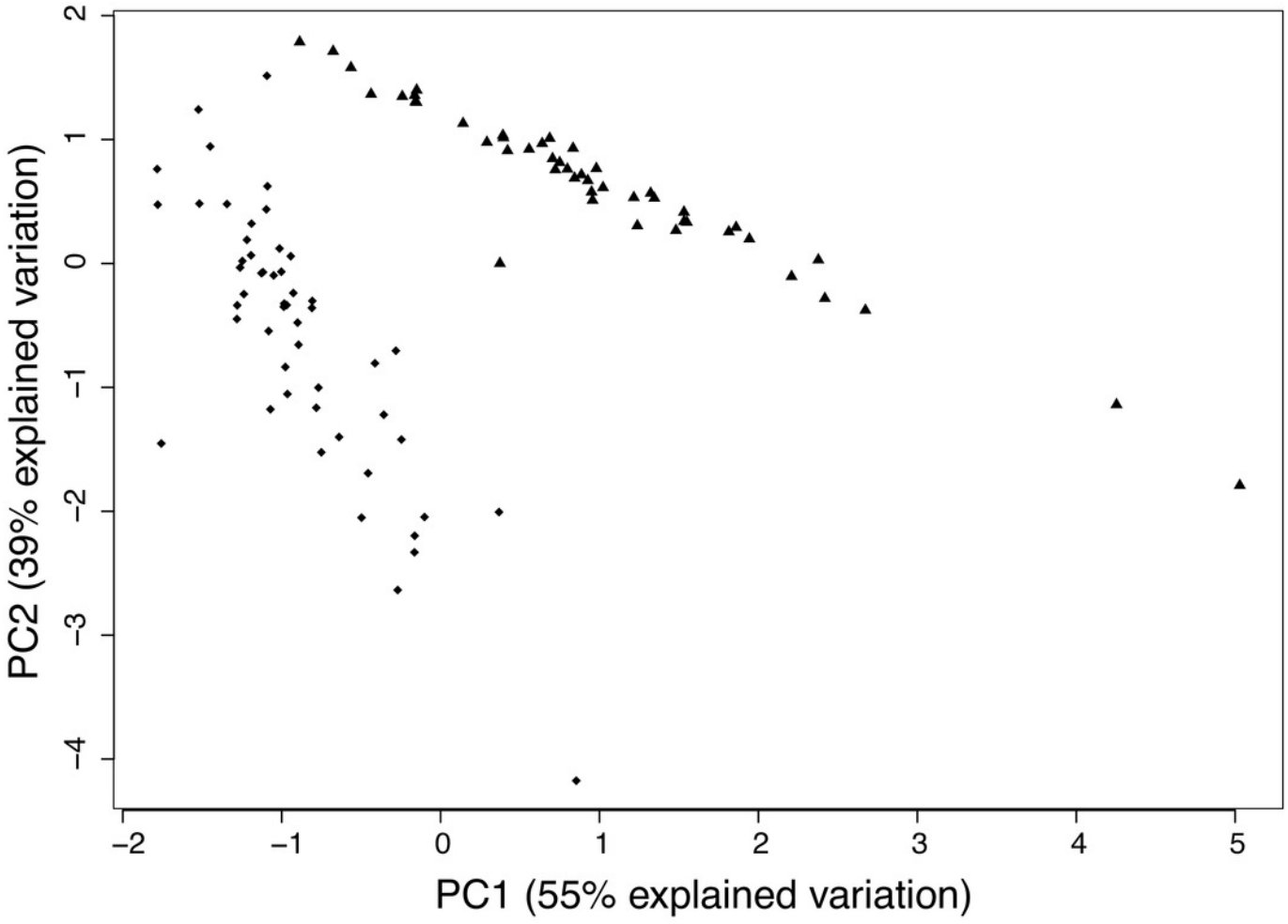
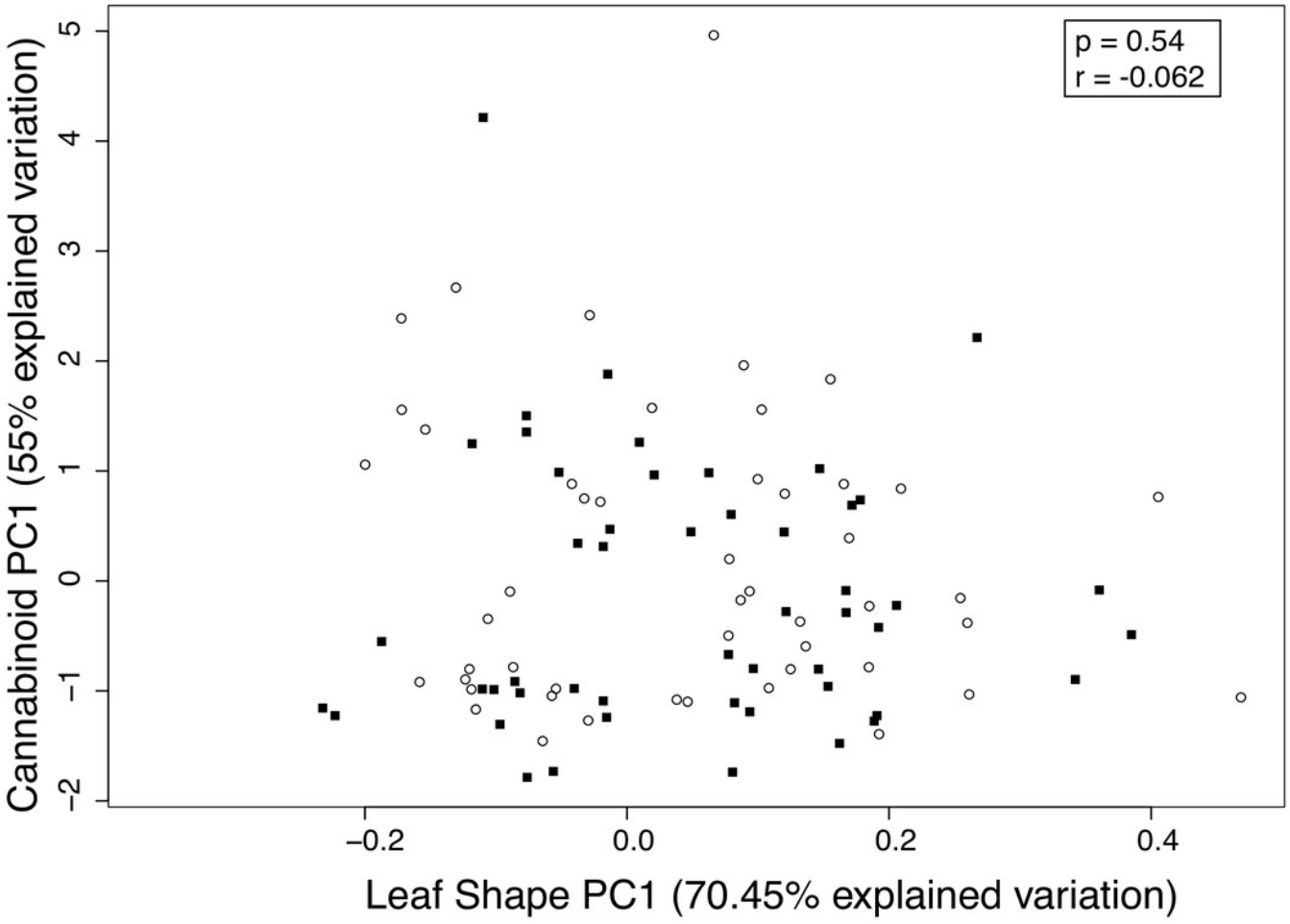


Figure 4

Figure 4: Correlation between leaf shape (PC1) and cannabinoid variation (PC1).

Leaf shape is not significantly to cannabinoid variation even at the most extreme points in the morphospace. Males are shown in squares and females in open circles.