

1 Genetic tools weed out misconceptions of strain reliability in *Cannabis sativa*: Implications for a
2 budding industry.

3

4

5 Anna L. Schwabe^{1*¶} and Mitchell E. McGlaughlin^{1*¶}

6

7 ¹School of Biological Sciences, University of Northern Colorado, Greeley, Colorado, United

8 States of America

9 *Corresponding Authors

10

11 Email

12 Anna Schwabe: schw0701@bears.unco.edu (970) 217-3300

13 Mitchell McGlaughlin: Mitchell.McGlaughlin@unco.edu (970) 351- 2139

14 ¶These authors contributed equally to this work

15

16 Date of Submission: May 27, 2018

17 Number of tables: 3

18 Number of Figs: 4 (total), 2 (color in print), 2 (color online only)

19 Supplementary: 3 Figs, 2 tables

20 Word count: 6239

21

22 Highlight: Genetic analyses provide evidence of genetic variation within clonal and stable seed

23 strains of commercially available *Cannabis sativa*, indicating the potential for inconsistent

24 products for medical patients and recreational users.

25 **Abstract**

26 *Cannabis sativa* is listed as a Schedule I substance by the United States Drug Enforcement
27 Agency and has been federally illegal in the United States since 1937. However, the majority of
28 states in the United States, as well as several countries, now have various levels of legal
29 *Cannabis*. Products are labeled with identifying strain names but there is no official mechanism
30 to register *Cannabis* strains, therefore the potential exists for incorrect identification or labeling.
31 This study uses genetic analyses to investigate strain reliability from the consumer point of view.
32 Ten microsatellite regions were used to examine samples from strains obtained from dispensaries
33 in three states. Samples were examined for genetic similarity within strains, and also a possible
34 genetic distinction between Sativa, Indica, or Hybrid types. The analyses revealed genetic
35 inconsistencies within strains. Additionally, although there was strong statistical support dividing
36 the samples into two genetic groups, the groups did not correspond to commonly reported
37 Sativa/Hybrid/Indica types. Genetic differences have the potential to lead to phenotypic
38 differences and unexpected effects, which could be surprising for the recreational user, but have
39 more serious implications for patients relying on strains that alleviate specific medical
40 symptoms.

41

42

43

44 Keywords: *Cannabis indica* – *Cannabis sativa* – consumer – genotype – hemp – marijuana –
45 medical – microsatellite – phenotype – strain

46

47 **List of abbreviations**

48 **US:** United States **HIV:** human immunodeficiency virus **AIDS:** acquired immune deficiency
49 syndrome **PTSD:** post-traumatic stress disorder **THC:** Δ^9 -tetrahydrocannabinol **USDA:** United
50 States Department of Agriculture **PVPA:** The Plant Variety Protection Act **PVPO:** Plant Variety
51 Protection Office **SLO:** San Luis Obispo **DNA:** deoxyribonucleic acid **CTAB:** Acetyl
52 trimethylammonium bromide **PCR:** Polymerase chain reaction **HWE:** Hardy–Weinberg
53 equilibrium **PCoA:** Principle Coordinates Analysis **SD:** standard Deviation **IA:** identical alleles

54

55 **Introduction**

56 *Cannabis sativa* L. is one of the most useful plants (Clarke & Merlin, 2013) with
57 evidence of human cultivation dating back thousands of years (Abel, 2013). *Cannabis*
58 prohibition in the United States began with the Marihuana Tax Act in 1937 (The Marihuana Tax
59 Act of 1937), and the Controlled Substances Act of 1970 classified *Cannabis* as a Schedule I
60 drug with no “accepted medical use in treatment in the United States” (Controlled Substances
61 Act, 1970). *Cannabis* is largely illegal worldwide, but laws allowing *Cannabis* for use as hemp,
62 medicine, and some adult recreational use are emerging (ProCon, 2016a). *Cannabis* is a multi-
63 billion dollar crop, but global restrictions have limited *Cannabis* related research. The origins
64 and genetic identities of many *Cannabis* strains are largely unknown, as there are relatively few
65 genetic studies focused on strains (Lynch *et al.*, 2016).

66 The World Drug Report estimates ~4.5% of the global population, consumes *Cannabis*
67 regularly (United Nations Office on Drugs, Crime, 2010), and there are an estimated ~3.5 million
68 medical marijuana patients in the US (Marijuana Policy Project, 2017). Recent legalization has
69 led to a surge of new strains as breeders are producing new plant varieties with novel chemical
70 profiles with various psychotropic effects, and relief for an array of symptoms associated with

71 medical conditions including (but not limited to): chronic pain, depression, anxiety, PTSD,
72 autism, fibromyalgia, epilepsy, Chron's Disease, and glaucoma (Ogborne *et al.*, 2000; Tomida *et*
73 *al.*, 2004; Borgelt *et al.*, 2013; Naftali *et al.*, 2013; ProCon, 2016b).

74 Research using a variety of techniques consistently finds drug-types and hemp are
75 genetically distinct (de Meijer *et al.*, 1996; Small, 1997; Sawler *et al.*, 2015; Lynch *et al.*, 2016;
76 Dufresnes *et al.*, 2017). Variation within the drug-types is higher than within hemp (Small, 1997;
77 Sawler *et al.*, 2015; Lynch *et al.*, 2016; Vergara *et al.*, 2016). There is limited genetic research on
78 variation within strains, but in studies with multiple accessions of a particular strain, variation is
79 observed (Sawler *et al.*, 2015; Lynch *et al.*, 2016; Soler *et al.*, 2017).

80 There are generally two *Cannabis* usage groups (hemp and drug-types) although the
81 scientific and common nomenclature is conflicted. The current Flora of North America
82 recognizes all forms of *Cannabis* as *Cannabis sativa* L. (Small, 1997), but many breeders and
83 botanists support the polytypic taxonomy of *Cannabis* based on morphological (de Lamarck &
84 Poiret, 1789; Schultes, 1970; Emboden, 1974; Anderson, 1980), chemical (de Meijer *et al.*, 2003;
85 Hillig & Mahlberg, 2004; Hillig, 2005; Hazekamp & Fishedick, 2012) and psychotropic (de
86 Meijer *et al.*, 2003; Hillig & Mahlberg, 2004; Hazekamp & Fishedick, 2012; Clarke & Merlin,
87 2013) differences. However, the suggested putative species are presumed to readily interbreed
88 and therefore violate species concepts that are applicable to plants (De Queiroz, 2007). The
89 common terminology for *Cannabis* products are, that (1) hemp types have $< 0.3\% \Delta^9$ -
90 tetrahydrocannabinol (THC), (2) plants of broad and narrow leaf drug-types as well as hybrid
91 variants with moderate to high THC concentrations are referred to as marijuana, (3) drug-type
92 strains of *Cannabis* are commonly divided into three categories: Sativa, Indica and Hybrid type
93 strains, (4) drug-type strains with low THC and high cannabidiol (CBD) are sought after for

94 medicinal use, and (5) there are thousands of variants of *Cannabis* referred to as strains. Genetic
95 analyses have not provide a clear consensus for higher taxonomic distinction among these
96 commonly described *Cannabis* types (Sawler *et al.*, 2015; Lynch *et al.*, 2016), but both the
97 recreational and medical *Cannabis* communities claim there are distinct differences in effects
98 between Sativa and Indica type strains (Smith, 2012; Leaf Science, 2014). Sativa type strains are
99 associated with tall, loosely branched plants with long, narrow leaflets, and are reported to have
100 energizing or uplifting psychotropic effects (Russo, 2007; Fishedick *et al.*, 2010; Hillig, 2004).
101 Indica type strains are associated with shorter plants with dense branching and broad leaflets, and
102 reportedly exhibit sedating effects and pain relieving properties (Russo, 2007; Fishedick *et al.*,
103 2010; Hillig, 2004). Hybrid types are a mix of varying degrees of the reported effects of Sativa
104 and Indica types.

105 Morphological variation is typically used to categorize species, sub-species, and varieties.
106 However, morphological identification can be difficult with closely related taxa and hybrid
107 organisms (Rieseberg, 1995; Rieseberg, 1997; Cattell & Karl, 2004; Mallet, 2005; Zha *et al.*,
108 2008, Schwabe *et al.* 2015). Sexual reproduction generally results in offspring with a blend of
109 traits from both parents. On the other hand, clonal offspring or progeny produced from self-
110 fertilization should be virtually identical to the parent. Unique physical differences (phenotypes)
111 and varying chemical profiles (chemotypes) may result when plants with the same genetic profile
112 (genotype) are impacted by environmental factors (phenotypic plasticity) (Schlichting, 1986;
113 Elzinga *et al.* 2015). Phenotypic plasticity is commonly observed in *Cannabis*, and therefore, the
114 use of chemical profile or other physical characteristics are not ideal to precisely identify
115 *Cannabis* variants (Schultes, 1970; Clarke & Merlin, 2013; Small, 2017)

116 Female flowers of predominantly dioecious *Cannabis* plants produce the majority of
117 cannabinoids and terpenes in glandular trichomes. Female plants are selected based on desirable
118 characters (mother plants) and are reproduced through cloning and, in some cases, self-
119 fertilization to produce seeds (Green, 2005). The offspring will be identical (from clone), or
120 nearly identical (from seed), to the mother plant. Cross-pollination allows for genetic variability
121 and novel strain creation, but generally *Cannabis* growers use cloning to produce consistent
122 products of established and popular strains. Whether propagated through cloning or from
123 germination of self-fertilized seed, genetic variation within strains should be minimal no matter
124 the source of origin.

125 There are an overwhelming number of *Cannabis* strains that vary widely in appearance,
126 taste, smell and psychotropic effects (de Lamarck & Poiret, 1789; Schultes, 1970; Emboden,
127 1974; Anderson, 1980; de Meijer *et al.*, 2003; Hillig & Mahlberg, 2004; Hillig, 2005; Hazekamp
128 & Fishedick, 2012; Clarke & Merlin, 2013). Strains are generally categorized as Indica, Sativa
129 or Hybrid types. Online databases such as Leafly (Leafly, 2018) and Wikileaf (Wikileaf, 2018)
130 provide consumers with information about strains but lack scientific merit for the *Cannabis*
131 industry to regulate the consistency of strains. To our knowledge, there have not been any
132 published scientific studies specifically investigating the genetic consistency of strains at
133 multiple points of sale for *Cannabis* consumers.

134 Of particular interest is how the genetic integrity of named *Cannabis* strains over time in
135 the absence of regulation been maintained (Green, 2014; Stockton, 2015). Other crop varieties
136 are protected by certification through the United States Department of Agriculture (USDA) and
137 The Plant Variety Protection Act of 1970 (PVPA), or similar mechanisms in other countries.
138 This system protects against commercial exploitation, allows for trademarking, and recognizes

139 intellectual property for developers of new plant cultivars (United States Department of
140 Agriculture, 1989). Traditionally, morphological characters were used to define new varieties in
141 crops such as grapes (*Vitis vinifera* L.), olives (*Olea europea* L.) and apples (*Malus domestica*
142 Borkh.). With the rapid development of new varieties in these types of crops, morphological
143 characters have become increasingly difficult to distinguish. Currently, quantitative and/or
144 molecular characters are often used to demonstrate uniqueness among varieties to obtain a plant
145 variety protection certificate from the Plant Variety Protection Office (PVPO) of the Agricultural
146 Marketing Service, USDA (United States Department of Agriculture, 2015). Microsatellite
147 genotyping enables growers and breeders of new cultivars to demonstrate uniqueness through
148 variable genetic profiles (Rongwen *et al.*, 1995). Microsatellite genotyping has been used to
149 distinguish cultivars and hybrid varieties of multiple crop varieties within species (Guilford *et*
150 *al.*, 1997; Hokanson *et al.*, 1998; Cipriani *et al.*, 2002; Belaj *et al.*, 2004; Sarri *et al.*, 2006;
151 Baldoni *et al.*, 2009; Štajner *et al.*, 2011; Costantini *et al.*, 2015; Pellerone *et al.*, 2015).
152 Multiple crop studies have found that 3-12 microsatellite loci are sufficient to accurately identify
153 varieties and detect misidentified individuals (Cipriani *et al.*, 2002; Belaj *et al.*, 2004; Sarri *et al.*,
154 2006; Poljuha *et al.*, 2008; Baldoni *et al.*, 2009; Muzzalupo *et al.*, 2009;). *Cannabis* varieties
155 however, are not afforded any legal protections, as the USDA considers it an “ineligible
156 commodity” (United States Department of Agriculture, 2016), but this system provides a model
157 by which *Cannabis* strains could also be developed, identified, registered, and protected.

158 Currently, the *Cannabis* industry has no way to verify strains. Consequently, suppliers
159 are unable to provide confirmation of strains. Reports of inconsistencies, along with the history
160 of underground trading and growing in the absence of a verification system, reinforce the
161 likelihood that strain names may be unreliable identifiers for *Cannabis* products at the present

162 time. Without verification systems in place, there is the potential for misidentification and
163 mislabeling of plants, creating names for plants of unknown origin, and even re-naming or re-
164 labeling plants with prominent names for better sale. *Cannabis* taxonomy is complex, but given
165 the success of microsatellites to determine varieties in other crops, we suggest the using genetic
166 based approaches to provide identification information for strains in the medical and recreational
167 marketplace.

168 Variable microsatellite markers were developed using the *Cannabis sativa* ‘Purple Kush’
169 draft genome (National Center for Biotechnology Information, accession AGQN000000000.1).
170 These regions were compared within commercially available *C. sativa* strains to determine if
171 products with the same name purchased from different sources have the genetic congruence we
172 expect from propagation of clones or self-fertilized seeds. The unique approach for this study
173 was that of the common retail consumer. Flower samples were purchased legally from
174 dispensaries based on what was available at the time of purchase. All products were purchased
175 as-is, with no additional information provided by the facility, other than the identifying label
176 (strain name). This study aimed to determine if: (1) any genetic distinction separates the common
177 perception of Sativa, Indica and Hybrid types; (2) purported proportions for Sativa, Indica and
178 Hybrid type strains are reflected in the genotypes of multiple strains; (3) consistent genetic
179 identity is found within a variety of different strain accessions obtained from different facilities;
180 (4) there is evidence of misidentification or mislabeling.

181

182 **Materials and Methods**

183 **Genetic Material**

184 *Cannabis* samples for 30 strains were acquired from 20 dispensaries or donors in three
185 states: Colorado - Denver (4), Boulder (3), Fort Collins (3), Garden City (4), Greeley (1),
186 Longmont (1); California - San Luis Obispo (4); and Washington - Union Gap (1) (Table 1). All
187 samples used in this study were obtained legally from either retail (Colorado and Washington),
188 medical (California) dispensaries, or as a donation from legally obtained samples (Greeley 1).
189 DNA was extracted using a modified CTAB extraction protocol (Doyle 1987) with 0.035-0.100
190 grams of dried flower tissue per extraction Proportions of Sativa and Indica phenotypes for each
191 strain were retrieved from Wikileaf (Wikileaf, 2018). Analyses were performed on the full 122-
192 sample dataset (Table 1). A subset of twelve strains in high demand was used throughout the
193 study to emphasize various genetic anomalies and patterns (Table 2). The twelve strains were
194 chosen based on popularity (Leafly, 2018; Wikileaf, 2018) and availability.

195

196 **Microsatellite Development**

197 The *Cannabis* draft genome from ‘Purple Kush’ (GenBank accession AGQN00000000.1)
198 was scanned for microsatellite repeat regions using MSATCOMMANDER-1.0.8-beta (Faircloth,
199 2008). Primers were developed *de-novo* flanking thirty microsatellites with 3-6 nucleotide repeat
200 units (Table S1). One primer in each pair was tagged with a 5’ universal sequence (M13, CAGT
201 or T7) so that a matching sequence with a fluorochrome tag could be incorporated via PCR
202 (Schwabe *et al.*, 2013). Ten of the thirty primer pairs produced consistent peaks within the
203 predicted size range and were used for the genetic analyses herein.

204

205 **PCR and Data Scoring**

206 Microsatellite loci were amplified in 12 μL reactions using 1.0 μL DNA (10-20 ng/ μL),
207 0.6 μL fluorescent tag (5 μM ; FAM, VIC, or PET), 0.6 μL non-tagged primer (5 μM), 0.6 μL
208 tagged primer (0.5 μM), 0.7 μL dNTP mix (2.5mM), 2.4 μL GoTaq Flexi Buffer (Promega,
209 Madison, WI, USA), 0.06 μL GoFlexi taq polymerase (Promega), 0.06 μL BSA (Bovine Serum
210 Albumin 100X), 0.5-6.0 μL MgCl or MgSO₄, and 0.48-4.98 μL dH₂O. Amplified products were
211 combined into multiplexes and diluted with water. Hi-Di formamide and LIZ 500 size standard
212 (Applied Biosystems, Foster City, CA, USA) were added before electrophoresis on a 3730
213 Genetic Analyzer (Applied Biosystems) at Arizona State University. Fragments were sized using
214 GENEIOUS 8.1.8 (Biomatters Ltd).

215

216 **Genetic Statistical Analyses**

217 GENALEX ver. 6.4.1 (Peakall & Smouse, 2006; Peakall & Smouse, 2012) was used to
218 calculate deviation from Hardy–Weinberg equilibrium (HWE). Linkage disequilibrium was
219 tested using GENEPOP ver. 4.0.10 (Raymond & Rousset, 1995; Rousset, 2008). The possibility
220 of null alleles was assessed using MICRO-CHECKER (Van Oosterhout *et al.*, 2004). Genotypes
221 were analyzed using the Bayesian cluster analysis program STRUCTURE ver. 2.4.2 (Pritchard *et al.*,
222 2000). Burn-in and run-lengths of 50,000 generations were used with ten independent
223 replicates for each STRUCTURE analysis. STRUCTURE HARVESTER (Earl, 2012), which
224 implements the Evanno method (Evanno *et al.*, 2005), was used to determine the K value that
225 best describes the number of genetic groups for the data set. GENALEX was used to conduct a
226 Principal Coordinate Analysis (PCoA) to examine variation in the dataset. Lynch & Ritland
227 (Lynch & Ritland, 1999) pairwise genetic relatedness (r) values were reported for each sample
228 within a strain using GENALEX. Mean pairwise relatedness (r) statistics were calculated

229 between all 122 samples resulting in 7381 pairwise r -values showing degrees of relatedness. A
230 genetic pairwise relatedness heat map of the data set was generated in Microsoft EXCEL. For all
231 strains the r -mean and standard deviation (SD) was calculated averaging among all samples.
232 Obvious outliers were determined by calculating the lowest r -mean and iteratively removing
233 those samples to determine the relatedness among the remaining samples in the subset. A graph
234 was generated for the twelve popular strains to show how the r -mean value change within a
235 strain when outliers were removed.

236

237 **Results**

238 The microsatellite analyses show genetic inconsistencies in *Cannabis* strains acquired
239 from different facilities. The samples used in this study are drug-type strains and are categorized
240 as Sativa, Indica and Hybrid type according to Wikileaf (Wikileaf, 2018). While some popular
241 strains were widely available, some strains were found only at two dispensaries (Table 1 & 2).
242 Since the aim of the research was not to identify specific locations where strain inconsistencies
243 were found, the names for each dispensary are coded to protect the identity of businesses.

244 There was no evidence of linkage-disequilibrium when all the samples were treated as a
245 single population. All loci deviate significantly from HWE when all samples were treated as a
246 single population, and all but one locus was monomorphic in at least two strains. All but one
247 locus had excess homozygosity and therefore possibly null alleles. Given the inbred nature and
248 extensive hybridization of *Cannabis*, deviations from neutral expectations are not surprising, and
249 the lack of linkage-disequilibrium indicates that the markers are spanning multiple regions of the
250 genome. There was no evidence of null alleles due to scoring errors.

251 STRUCTURE HARVESTER calculated high support ($\Delta K=146.56$) for two genetic
252 groups, $K=2$ (Fig. 1). STRUCTURE assignment for all samples is shown in Fig. 2 with the
253 strains ordered by the purported proportions of Sativa phenotype (Wikileaf, 2018) and then
254 alphabetically within each strain by city. A clear genetic distinction between Sativa and Indica
255 types would assign 100% Sativa strains ('Durban Poison') to one genotype, and assign 100%
256 Indica strains ('Purple Kush') to the other genotype (Table 2, Fig. 2). Division of the genotypes
257 into two genetic groups does not support the commonly described Sativa and Indica phenotypes.
258 For the assigned 100% Sativa type strain 'Durban Poison', seven of nine samples show greater
259 than 96% assignment to genotype 1 (blue; Fig. 2). For the assigned 100% Indica type 'Purple
260 Kush' three of four samples of show greater than 89% assignment to genotype 2 (yellow; Fig. 2).
261 However, samples of 'Hawaiian' (90% Sativa) and 'Grape Ape' (100% Indica) do not show
262 consistent patterns of predominant assignment to genotype 1 or 2. Interestingly, 'Durban Poison'
263 (100% Sativa, $n = 9$) and 'Sour Diesel' (90% Sativa, $n = 7$) have 86% and 14% average
264 assignment to genotype 1, respectively. Hybrid strains should result in some proportion of shared
265 ancestry, with assignment to both genotype 1 and 2. The strains 'Blue Dream' and 'Tahoe OG'
266 are reported as 50-50% Sativa-Indica Hybrid strains, but eight of nine samples of 'Blue Dream'
267 show $> 80\%$ assignment to genotype 1, and three of four samples of 'Tahoe OG' show $< 7\%$
268 assignment to genotype 1.

269 Principal Coordinate Analyses (PCoA) were conducted using GENALEX for (1) all
270 samples (Fig. 2) and (2) twelve popular strains (Fig. S2). The samples in the PCoA of all 30
271 strains are organized from 100% Sativa types (red), through all levels of Hybrid types, to 100%
272 Indica types (purple; Fig. 4). Strain types with the same reported proportions are the same color
273 but have different symbols. The PCoA of all strains represents 14.90% of the variation in the

274 data on coordinate axis 1, 9.56% on axis 2, and 7.07% on axis 3 (not shown). The second PCoA
275 of twelve popular strains specifically examines the genetic relationship within strains that are in
276 high demand (Fig. S2). The results from this analysis found that 15.30% of the variation in the
277 data is explained by coordinate axis 1, 12.98% on axis 2, and 7.96% on axis 3 (not shown).

278 Lynch & Ritland (Lynch & Ritland, 1999) pairwise genetic relatedness (r) between all
279 122 samples was calculated in GENALEX. The resulting 7380 pairwise r -values were converted
280 to a heat map using purple to indicate the lowest pairwise relatedness value (-1.09) and green to
281 indicate the highest pairwise relatedness value (1.00; Fig. S3. Comparisons are detailed for six
282 popular strains (Fig. 3) to illustrate the relationship of samples from different sources and the
283 impact of outliers. Values of close to 1.00 indicate a high degree of relatedness (Lynch &
284 Ritland, 1999), which could be indicative of clones or seeds from the same mother (Green, 2005;
285 SeedFinder, 2017). First order relatives (full siblings or mother-daughter) share 50% genetic
286 identity (r -value = 0.50), second order relatives (half siblings or cousins) share 25% genetic
287 identity (r -value = 0.25), and unrelated individuals are expected to have an r -value of 0.00 or
288 lower. Negative values arise when individuals are less related than expected under normal
289 panmictic conditions (Moura *et al.*, 2013; Norman *et al.*, 2017). Values ranged from -1.09
290 (between ‘Purple Haze’ Greeley 1 and ‘Girl Scout Cookies’ Union Gap 1) indicating low levels
291 of relatedness, to 1.00 (e.g., between ‘Durban Poison’ samples from Boulder 3 and Fort Collins
292 3).

293 Individual pairwise r -values were averaged within strains to calculate the overall r -mean
294 as a measure of genetic similarity within strains. The overall r -means within strains ranged from
295 -0.22 (‘Tangerine’) to 0.68 (‘Island Sweet Skunk’) (Table 3). Standard deviations ranged from
296 0.04 (‘Jack Herer’) to 0.51 (‘Bruce Banner’). The strains with higher standard deviation values

297 indicate a wide range of genetic relatedness within a strain, while low values indicate that
298 samples within a strain share similar levels of genetic relatedness. In order to determine how
299 outliers impact the overall relatedness in a strain, the farthest outlier (lowest pairwise r -mean
300 value) was removed and the overall r -means and SD values within strains were recalculated
301 (Table 3). In all strains, the overall r -means increased when outliers were removed. In strains
302 with more than three samples, a second outlier was removed and the overall r -means and SD
303 values were recalculated. Overall r -means were used to determine degree of relatedness as clonal
304 (or from stable seed; overall r -means > 0.9), first or higher order relatives (overall r -means 0.46
305 – 0.89), second order relatives (overall r -means 0.26 - 0.45), low levels of relatedness (overall r -
306 means 0.00 - 0.25), and not related (overall r -means < 0.00). Initial overall r -means indicate only
307 three strains are first or higher order relatives (Table 3). Removing outliers revealed samples
308 within ten of the remaining 22 strains are first or higher order relatives. After outliers were
309 removed, 15 of the 30 strains are comprised of first or higher order relatives, indicating outliers
310 are often responsible for variability within strains. Removing outliers revealed samples within
311 seven of the twelve popular strains are of first or higher order relatives (Table 3, Fig. 4). Three
312 strains are comprised of second order relatives with overall r -means ranging from 0.22 - 0.25.
313 Two strains show low levels of relatedness with overall r -means ranging from 0.13 - 0.16 even
314 after outliers are removed (Table 3). The impact of outliers can be clearly seen in the heat map
315 for ‘Durban Poison’ which shows the relatedness for 36 comparisons (Fig. 3A), six of which are
316 nearly identical (r -value 0.90 - 1.0), six of which are first order siblings (r -value 0.46 - 0.89), six
317 of which are second order relatives (r -value 0.26 - 0.45), five of which have low levels of
318 relatedness (r -value 0.00 - 0.25), and 13 which are not related (r -value < 0.00). However, removal

319 of two outliers, Denver 1 and Garden City 2, reduces the number of comparisons ranked as not
320 related from 13 to zero, and low level of relatedness from five to one.

321

322 **Discussion**

323 The legal status and social attitudes toward *Cannabis* are changing worldwide, with more
324 than half the states in the U.S. having sanctioned medical *Cannabis* use (ProCon, 2016a).
325 *Cannabis* types and strains are becoming an ever-increasing topic of discussion, so it is
326 important that scientists and the public can discuss *Cannabis* in a similar manner. Currently, not
327 only are Sativa and Indica types disputed, but also experts are at odds about nomenclature for
328 *Cannabis* (Clarke & Merlin, 2015; Small, 2015b). We investigated the possibility of a genetic
329 distinction in commonly described Sativa and Indica strains. Previous genetic research found
330 genetic variability among seeds from the same strain supplied from a single source, indicating
331 genotypes within strains are variable (Sohler *et al.*, 2017). However, it was unclear if the seeds in
332 the study were produced from multiple parent plants, which could have introduced a source for
333 genetic variation. The focus of this study is that genetic profiles from strains with the same
334 identifying name should have identical, or at least, highly similar genotypes no matter the source
335 of origin. It is important that strain names reflect consistent genetic identity, especially for those
336 who rely on *Cannabis* to alleviate specific medical symptoms. An important element for this
337 study is that samples were acquired from multiple locations to maximize the potential for
338 variation among samples. The multiple genetic analyses used here address important questions
339 and bring scientific evidence to support claims that inconsistent products are being distributed.
340 Genotype analysis can be used to ensure higher levels of consistency within strains. Maintenance
341 of the genetic integrity of strains is possible only following evaluation of genetic consistency,

342 and continuing to overlooking this aspect will to promote variability and phenotypic variation.
343 Addressing strain variability at the molecular level is of the utmost importance while the industry
344 is still relatively new.

345 Genetic analyses have consistently found genetic distinction between hemp and
346 marijuana, but no clear distinction has been shown between the common description of Sativa
347 and Indica types (de Meijer *et al.*, 1996; Small, 1997; Lynch *et al.*, 2016; Sawler *et al.*, 2015;
348 Vergara *et al.*, 2016; Dufresnes *et al.*, 2017; Soler *et al.*, 2017). We found high support for two
349 genetic groups in the data (Fig. 1) but no discernable distinction or pattern between the described
350 Sativa and Indica strains. The color-coding of strains in the PCoA for all 122 samples allows for
351 visualization of clustering among similar phenotypes by color Sativa (red/orange), Indica
352 (blue/purple) and Hybrid (green) type strains (Fig. 2). However, there is no evidence of
353 clustering in the three commonly described types. If genetic differentiation of the commonly
354 perceived Sativa and Indica types previously existed, it is no longer detectable in the neutral
355 genetic markers used here. Extensive hybridization and selection has presumably created a
356 homogenizing effect and erased evidence of potentially divergent historical genotypes.

357 Wikileaf maintains that the proportions of Sativa and Indica reported for strains are
358 largely based on genetics and lineage (Dan Nelson, Wikileaf, personal communication). This has
359 seemingly become convoluted over time (Russo, 2007; Small, 2015a; Clarke & Merlin, 2013;
360 Small, 2017). Our results show that commonly reported levels of Sativa, Indica and Hybrid type
361 strains are often not reflected in the average genotype. For example, two sought-after Sativa
362 strains, ‘Durban Poison’ and ‘Sour Diesel’, were found to have contradicting genetic
363 assignments (Fig. 1, Table 2). ‘Durban Poison’, described as 100% Sativa, has an 86% average
364 assignment to genotype 1, while ‘Sour Diesel’, described as 90% Sativa, has a 14% average

365 assignment to genotype 1. This analysis indicates strains with similar reported proportions of
366 Sativa or Indica may have differing genetic assignments. Further illustrating this point is that
367 ‘Bruce Banner’, ‘Flo’, ‘Jillybean’, ‘Pineapple Express’, ‘Purple Haze’, and ‘Tangerine’ are all
368 reported to be 60/40 Hybrid type strains, but clearly have differing levels of admixture both
369 within and among these reportedly similar strains (Table 2, Fig. 1). From these results, we can
370 conclude that reported ratios or differences between Sativa and Indica phenotypes are not
371 discernable using these genetic markers. Given the lack of genetic distinction between Indica and
372 Sativa types, it is not surprising that reported ancestry proportions are also not supported.

373 To accurately address reported variation within strains, samples were purchased from
374 various locations, as a customer, with no information of strains other than publically available
375 online information. Evidence for genetic inconsistencies is apparent within many strains and
376 supported by multiple genetic analyses. In our analyses of 30 strains, only 4 strains had
377 consistent STRUCTURE genotype assignment and admixture among all samples: ‘Chemdawg’
378 (n=7), ‘Island Sweet Skunk’ (n=3), ‘Larry OG’ (n=3) and ‘Jack Flash’ (n = 2; Fig. 2). However,
379 it is clear that many strains contained one or more obvious genetic outliers (e.g. Durban Poison –
380 Denver 1; Fig 1, 3A). With the removal of one obvious outlier, the remaining samples of eleven
381 strains were classified as first order relatives based on pairwise genetic relatedness *r*-values
382 (overall *r*-mean >0.45; Table 3, Fig. 4). The removal of a second outlier resulted in 15 of the 30
383 strains having an overall *r*-mean >0.45 (Table 3, Fig. 4). Together, these results indicate that half
384 of the strains used in this analysis showed relatively stable genetic identity among most samples
385 within a strain. Six of the strains with inconsistent patterns had only two samples, both of which
386 were different (e.g., ‘Trainwreck’ and ‘Headband’). The remaining nine strains in the analysis
387 had more than one obvious outlier (e.g., ‘Sour Diesel’) or had no consistent genetic pattern

388 among the samples within the strain (e.g., ‘Girl Scout Cookies’; Table 3, Fig. 1, Fig. 2, Fig. S2).
389 It is noteworthy that many of the strains used here fell into a range of genetic relatedness
390 indicative of first order siblings (r -value 0.46 - 0.89) when samples with high genetic divergence
391 were isolated and removed from the data set (Table 4; Figs. 3, 4).

392 Relationships within the twelve popular strains were analyzed separately to determine if
393 (1) strains with more samples show a higher degree of clustering, and (2) strains in higher
394 demand have a higher degree of genetic relatedness. The analysis of genetic variation for the
395 subset of twelve popular strains shows some clustering within strains (Fig. S2), but clustering is
396 not seen for all strains, and outliers are apparent. This analysis represents more of the variation in
397 the data compared to the PCoA for all 30 strains and shows clustering of some strains, such as
398 ‘Durban Poison’, ‘Golden Goat’ and ‘Blue Dream’. However, all clusters have at least one
399 sample that is removed from the other samples in the group. From this we argue that samples
400 representing the popular strains may be slightly more likely to have a higher degree of genetic
401 relatedness, but more sampling would be required to determine this with confidence.

402 A pairwise genetic heat map based on Lynch & Ritland (Lynch & Ritland, 1999)
403 pairwise genetic relatedness (r -values) was generated to visualize genetic relatedness throughout
404 the data set (Fig. S3). Values of 1.00 (or close to) are assumed to be clones or plants from self-
405 fertilized seed. Six examples of within-strain pairwise comparison heat maps were examined to
406 illustrate common patterns (Fig.7). The heat map shows that many strains contain samples that
407 are first order relatives or higher (r -value > 0.49). For example ‘Sour Diesel’ (Fig. 3??) has 12
408 comparisons of first order or above, and six have low/no relationship. There are also values that
409 could be indicative of clones or plants from a stable seed source such as ‘Blue Dream’ (Fig.
410 3???), which has 10 nearly identical comparisons (r -value 0.90-1.00), and no comparisons in

411 ‘Blue Dream’ have negative values. While ‘Blue Dream’ has an initial overall r -mean indicating
412 first order relatedness within the samples (Table 3, Fig. 4), it still contains more variation than
413 would be expected from a clone only strain (SeedFinder, 2017). Other clone-only strains
414 (SeedFinder, 2017), e.g. ‘Girl Scout Cookies’ (Table 3, Fig. 3??) and ‘Golden Goat’ (Table 3,
415 Fig. 3??), have a high degree of genetic variation resulting in low overall relatedness values.
416 Outliers were calculated and removed iteratively to demonstrate how they affected the overall r -
417 mean within the twelve popular strains (Table 3, Fig. 4). In all cases, removing outliers increased
418 the mean r -value, as illustrated by ‘Bruce Banner’, which increased substantially, from 0.3 to 0.9
419 when samples with two outlying genotypes removed. The outliers are evidence of
420 inconsistencies within strains and when removed, genetic relatedness greatly improves. There are
421 unexpected areas in the heat map that indicate high degrees of relatedness between different
422 strains (Fig. S3). For example, comparisons between ‘Golden Goat’ and ‘Island Sweet Skunk’
423 (overall r - mean 0.37) are higher than within samples of ‘Sour Diesel’. Interestingly, ‘Golden
424 Goat’ is reported to be a hybrid descendant of ‘Island Sweet Skunk’ (Leafly, 2018), which
425 explains the high genetic relatedness between these strains. However, most of the between strain
426 overall r - mean are negative (e.g., ‘Golden Goat’ to ‘Durban Poison’ -0.03 and ‘Chemdawg’ to
427 ‘Durban Poison’ -0.22; Fig. S3), indicative of limited recent genetic relationship.

428 While collecting samples from various dispensaries, it was noted that strains of
429 ‘Chemdawg’ had various different spellings of the strain name, as well as numbers and/or letters
430 attached to the name. Without knowledge of the history of ‘Chemdawg’, the assumption was that
431 these were local variations. These were acquired to include in the study to determine if and how
432 these variants were related. Upon investigation of possible origins of ‘Chemdawg’, an interesting
433 history was uncovered, especially in light of the results (Backes & Weil, 2014). Legend has it

434 that someone named “Chemdog” (a person) grew the variations (‘Chemdawg 91’, ‘Chemdawg
435 D’, ‘Chemdawg 4’, ‘Chemdog 1’) from seeds he found in an ounce he purchased at a Grateful
436 Dead concert. This illustrates how *Cannabis* strains may have come to market in a non-
437 traditional manner. The history of ‘Chemdawg’ is currently unverifiable, but the analysis
438 supports that these variations could be from seeds of the same plant. Genetic analyses can add
439 scientific support to the stories behind vintage strains and possibly help clarify the history of
440 specific strains.

441 Possible facilitation of inconsistencies may come from both suppliers and growers of
442 *Cannabis* clones and stable seed, because currently they can only assume the strains they possess
443 are true to name. There is a chain of events from seed to sale that relies heavily on the supplier,
444 grower, and dispensary to provide the correct product, but there is currently no reliable way to
445 verify *Cannabis* strains. The possibility exists for errors in plant labeling, misplacement,
446 misspelling, and/or relabeling along the entire chain of production. Although the expectation is
447 that plants are labeled carefully and not re-labeled with a more desirable name for a quick sale,
448 these misgivings must be considered. Identification by genetic markers has largely eliminated
449 these types of mistakes in other widely cultivated crops such as grapes, olives and apples.
450 Modern genetic applications can accurately identify varieties and can clarify ambiguity in closely
451 related and hybrid species, [e.g., Rongwen *et al.*, 1995; Guilford *et al.*, 1997; Belaj *et al.* 2004;
452 Muzzalupo *et al.*, 2009; S’tajner *et al.*, 2011).

453 Matching genotypes within the same strains were expected, but highly similar genotypes
454 between samples of different strains could be the result of mislabeling or misidentification,
455 especially when acquired from the same source. The pairwise genetic relatedness r -values were
456 examined for incidence of possible mislabeling or re-labeling. There were instances in which

457 different strains had r -values = 1.0 (Fig. S3), indicating clonal genetic relationships. Two
458 samples with matching genotypes were obtained from the same location ('Larry OG' and 'Tahoe
459 OG' from San Luis Obispo 3). This could be evidence for mislabeling or misidentification
460 because these two samples have similar names. It is unlikely that these samples from reportedly
461 different strains have identical genotypes, and more likely that these samples were mislabeled at
462 some point. Misspelling may also be a source of error, especially when facilities are handwriting
463 labels. An example of possible misspelling may have occurred in the sample labeled 'Chemdog
464 1' from Garden City 1. 'Chemdawg 1', a described strain, could have easily been misspelled, but
465 it is unclear whether this instance is evidence for mislabeling or renaming a local variant.
466 Inadvertent mistakes may carry through to scientific investigation where strains are spelled or
467 labeled incorrectly. For example, Vergara et al. (2016) reports genome assemblies for
468 'Chemdog' and 'Chemdog 91' as they are reported in GenBank (GCA_001509995.1), but
469 neither of these labels are recognized strain names. It is likely that these are 'Chemdawg' and
470 'Chemdawg 91' (Leafly, 2018; Wikileaf, 2018) although it is possible these strains are
471 unreported variants. Another example that may lead to confusion is how information is reported
472 in public databases. For example, data is available for the reported monoisolate of 'Pineapple
473 Banana Bubba Kush' in GenBank (SAMN06546749), and while 'Pineapple Kush', 'Banana
474 Kush' and 'Bubba Kush' are known strains (Leafly, 2018; Wikileaf, 2018), the only record of
475 'Pineapple Banana Bubba Kush' is in Genbank. This study has highlighted several possible
476 sources of error and how genotyping can serve to uncover sources of variation. Although this
477 study was unable to confirm sources of error, it is important that producers, growers and
478 consumers are aware that there are errors and they should be documented and corrected
479 whenever possible.

480

481 **Conclusion**

482 Over the last decade, the legal status of *Cannabis* has shifted and is now legal for medical
483 use, and some recreational adult use, in the majority of the United States as well as several other
484 countries that have legalized or decriminalized *Cannabis*. The recent legal changes have led to
485 an unprecedented increase in the number of strains available to consumers. There are currently
486 no baseline genotypes for any strains, but steps should be taken to ensure products marketed as a
487 particular strain are genetically congruent. Although the sampling in this study was not
488 exhaustive, the results are clear: strain inconsistency is evident and is not limited to a single
489 source, but rather exists among dispensaries across cities in multiple states. Various suggestions
490 for naming the genetic variants do not seem to align with the current widespread definitions of
491 Sativa, Indica, Hybrid, and Hemp (Hillig, 2005; Clarke & Merlin, 2013). As our *Cannabis*
492 knowledge base grows, so does the communication gap between scientific researchers and the
493 public. Currently, there is no way for *Cannabis* suppliers, growers or consumers to definitively
494 verify strains. Exclusion from protection, due to the Federal status of *Cannabis* as a Schedule I
495 drug, has created avenues for error and inconsistencies. Presumably, the genetic inconsistencies
496 will often manifest as differences in overall effects (Backes, 2014). Differences in characteristics
497 within a named strain may be surprising for a recreational user, but differences may be more
498 serious for a medical patient who relies on a particular strain for alleviation of specific
499 symptoms.

500 This study shows that in neutral genetic markers, there is no consistent genetic
501 differentiation between the widely held perceptions of Sativa and Indica *Cannabis* types.

502 Moreover, the genetic analyses do not support the reported proportions of Sativa and Indica
503 within each strain, which is expected given the lack of genetic distinction between Sativa and
504 Indica. Instances were found where samples within strains are not genetically similar, which is
505 unexpected given the manner in which *Cannabis* plants are propagated. Although it is impossible
506 to determine the source of these inconsistencies as they can arise at multiple points throughout
507 the chain of events from seed to sale, we theorize misidentification, mislabeling, misplacement,
508 misspelling, and/or relabeling are all possible. Especially where names are similar, there is the
509 possibility for mislabeling, as was shown here. In many cases genetic inconsistencies within
510 strains were limited to one or two samples. We feel that there is a reasonable amount of genetic
511 similarity within many strains, but currently there is no way to verify the “true” genotype of any
512 strain. Although the sampling here includes merely a fragment of the available *Cannabis* strains,
513 our results give scientific merit to claims that strains can be unpredictable.

514

515 **Supplementary Data**

516 Table S1: Primer information used in this research.

517

518 Fig. S1: STRUCTURE HARVESTER graph indicating K=2 is highly supported.

519

520 Fig. S2: Principal Coordinates Analysis (PCoA) for twelve popular strains.

521

522 Fig. S3: Pairwise genetic relatedness (r) heat table with values for 122 samples.

523

524

525 *Acknowledgements*

526 We thank Gerald Bresowar and Nolan Kane for comments on an earlier draft of this manuscript.

527 The University of Northern Colorado School of Biological Sciences supported this research, and

528 we are grateful to the Graduate Student Association and the Gerald Schmidt Memorial Biology
529 Scholarship for providing partial funding to carry out this research.

Figure Legends

Fig. 1

Bar plot graphs generated from STRUCTURE analysis for 122 individuals from 30 strains dividing genotypes into two genetic groups, $K=2$. Samples were arranged by purported proportions from 100% Sativa to 100% Indica (Wikileaf, 2018) and then alphabetically within each strain by city. Each strain includes reported proportion of Sativa in parentheses (Wikileaf, 2018) and each sample includes the coded location and city from where it was acquired. Each bar indicates proportion of assignment to genotype 1 and genotype 2.

Fig. 2

Principal Coordinates Analysis (PCoA) generated in GENALEX. Samples are a color-coded continuum by proportion of Sativa (Table 2) with the strain name given for each sample: Sativa type (red: 100% Sativa proportion, Hybrid type (dark green: 50% Sativa proportion), and Indica type (purple: 0% Sativa proportion). Different symbols are used to indicate different strains within reported phenotype. Coordinate axis 1 explains 14.29% of the variation, coordinate axis 2 explains 9.56% of the variation, and Coordinate axis 3 (not shown) explains 7.07%.

Fig. 3

Heat maps of six prominent strains using Lynch & Ritland (1999) pairwise genetic relatedness (r) values: purple indicates no genetic relatedness (minimum value -1.09) and green indicates a high degree of relatedness (maximum value 1.0). Sample strain names and location of origin are indicated along the top and down the left side of the chart. Pairwise genetic relatedness (r) values are given in each cell and cell color reflects the degree to which two individuals are related.

Fig. 4

This graph indicates the mean pairwise genetic relatedness (r) initially (light gray) and after the removal of one (medium gray) or two (dark gray) outlying samples in 12 prominent strains.

References

- Abel EL.** 2013. *Marihuana: the first twelve thousand years*. Springer Science & Business Media.
- Anderson LC.** 1980. Leaf variation among *Cannabis* species from a controlled garden. Botanical Museum Leaflets, Harvard University **28**, 61-9.
- Backes M, Weil A.** 2014. *Cannabis pharmacy: the practical guide to medical marijuana*. Black Dog & Leventhal.
- Baldoni L, et al.** 2009. A consensus list of microsatellite markers for olive genotyping. Molecular Breeding. **24**, 213-31.
- Belaj A, Cipriani G, Testolin R, Rallo L, Trujillo I.** 2004. Characterization and identification of the main Spanish and Italian olive cultivars by simple-sequence-repeat markers. HortScience. 2004. **39**, 1557-61.
- Borgelt LM, Franson KL, Nussbaum AM, Wang GS.** 2013. The pharmacologic and clinical effects of medical cannabis. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. **33**, 195-209.
- Cattell MV, Karl SA.** 2004. Genetics and morphology in a *Borrchia frutescens* and *B. arborescens* (Asteraceae) hybrid zone. *American Journal of Botany*. **91**, 1757-66.
- Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R.** 2002. Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theoretical and Applied Genetics*. **104**, 223-8.
- Clarke R, Merlin M.** 2013. *Cannabis: Evolution and Ethnobotany*. University of California Press.
- Clarke R, Merlin M.** 2015. Letter to the Editor: Small, Ernest. 2015. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. *The Botanical Review*. **81**, 295-305.
- Controlled Substances Act.** 1970. [Pub. L. 91-513, title II](#), § 101, Oct 27, 1970, [84 Stat. 1242](#).
- Costantini LA, Monaco A, Vouillamoz JF, Forlani M, Grando MS.** 2015. Genetic relationships among local *Vitis vinifera* cultivars from Campania (Italy). *VITIS-Journal of Grapevine Research*. **44**, 25.
- De Queiroz K.** 2007. Species concepts and species delimitation. *Systematic biology*. **56**, 879-86.
- Doyle JJ.** 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue.

- Phytochemical Bulletin. **19**, 11-5.
- Dufresnes C, Jan C, Bienert F, Goudet J, Fumagalli L.** 2017. Broad-Scale Genetic Diversity of *Cannabis* for Forensic Applications. PloS one. 2017. **12**, e0170522.
- Earl DA.** 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources. **4**, 359-61.
- Elzinga S, Fishedick J, Podkolinski R, Raber JC.** 2015. Cannabinoids and terpenes as chemotaxonomic markers in cannabis. Natural Products Chemistry & Research. **3**.
- Emboden WA.** 1974. *Cannabis*—a polytypic genus. Economic Botany. **28**, 304-10.
- Evanno G, Regnaut S, Goudet J.** 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology. **14**, 2611-20.
- Faircloth BC.** 2008. Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources. **8**, 92-4.
- Fishedick JT, Hazekamp A, Erkelens T, Choi YH, Verpoorte R.** 2010. Metabolic fingerprinting of *Cannabis sativa* L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. Phytochemistry. **71**, 2058-73.
- Green G.** 2005. The Cannabis Breeder's Bible. San Francisco: Green Candy Press.
- Green J.** 2014. How Many Marijuana Strain Names Are There? Marijuana Business News. <http://www.theweedblog.com/how-many-marijuana-strains-are-there/>. Accessed July 14 2016.
- Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R.** 1997. Microsatellites in *Malus x domestica* (apple): abundance, polymorphism and cultivar identification. Theoretical and Applied Genetics. **94**, 249-54.
- Hazekamp A, Fishedick JT.** 2012. *Cannabis*-from cultivar to chemovar. Drug Testing and Analysis. **4**, 660-7.
- Hillig KW.** 2004. A chemotaxonomic analysis of terpenoid variation in *Cannabis*. Biochemical Systematics and Ecology. **32**, 875-91.
- Hillig KW, Mahlberg PG.** 2004. A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). American Journal of Botany. **91**, 966-75.
- Hillig KW.** 2005. Genetic evidence for speciation in *Cannabis* (Cannabaceae). Genetic Resources and Crop Evolution. **52**, 161-80.

- Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR.** 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus × domestica* Borkh. core subset collection. *Theoretical and Applied Genetics*. **97**, 671-83.
- de Lamarck JB, Poiret JL.** 1789. *Encyclopédie méthodique: botanique*. chez Panckoucke.
- Leaf Science.** 2014. Indica vs. Sativa: Understanding The Differences. <http://www.leafscience.com/2014/06/19/indica-vs-sativa-understanding-differences/>. Accessed June 19 2016.
- Leafly.** 2017. Cannabis Strain and Infused Product Explorer. <https://www.leafly.com> Accessed May 31 2017.
- Lynch M, Ritland K.** 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*. **152**, 1753-66.
- Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, deCesare K, Land DP, Kane NC.** 2016. Genomic and chemical diversity in *Cannabis*. *Critical Reviews in Plant Sciences*. **35**, 349-63.
- Mallet J.** 2005. Hybridization as an invasion of the genome. *Trends in ecology & evolution*. **20**, 229-37.
- Marijuana Policy Project.** Medical Marijuana Patient Numbers. 2017. <https://www.mpp.org/issues/medical-marijuana/state-by-state-medical-marijuana-laws/medical-marijuana-patient-numbers/>. Accessed May 30 2017.
- de Meijer EP, Bagatta M, Carboni A, Crucitti P, Moliterni VC, Ranalli P, Mandolino G.** 2003. The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics*. **163**, 335-46.
- de Meijer ED, Keizer LC.** 1996. Patterns of diversity in *Cannabis*. Genetic resources and crop evolution. **43**, 41-52.
- Moura AE, Natoli A, Rogan E, Hoelzel AR.** 2013. Atypical panmixia in a European dolphin species (*Delphinus delphis*): implications for the evolution of diversity across oceanic boundaries. *Journal of Evolutionary Biology*. **26**, 63-75.
- Muzzalupo I, Stefanizzi F, Perri E.** 2009. Evaluation of olives cultivated in southern Italy by simple sequence repeat markers. *HortScience*. **44**, 582-8.
- Naftali T, Schleider LB, Dotan I, Lansky EP, Benjaminov FS, Konikoff FM.** 2013. *Cannabis* induces a clinical response in patients with Crohn's disease: a prospective placebo-controlled study. *Clinical Gastroenterology and Hepatology*. **11**, 1276-80.
- Norman AJ, Stronen AV, Fuglstad GA, Ruiz-Gonzalez A, Kindberg J, Street NR, Spong G.**

2017. Landscape relatedness: detecting contemporary fine-scale spatial structure in wild populations. *Landscape Ecology*. **32**, 181-94.
- Ogborne AC, Smart RG, Weber T, Birchmore-Timney C.** 2000. Who is using cannabis as a medicine and why: an exploratory study. *Journal of Psychoactive Drugs*. **32**, 435-43.
- Peakall RO, Smouse PE.** 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology notes*. **6**, 288-95.
- Peakall RO, Smouse PE.** 2012. GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. **28**, 2537-9.
- Pellerone FI, Edwards KJ, Thomas MR.** 2015. Grapevine microsatellite repeats: isolation, characterisation and use for genotyping of grape germplasm from Southern Italy. *VITIS-Journal of Grapevine Research*. **40**, 179.
- Poljuha D, Sladonja B, Šetić E, Milotić A, Bandelj D, Jakše J, Javornik B.** 2008. DNA fingerprinting of olive varieties in Istria (Croatia) by microsatellite markers. *Scientia Horticulturae*. **115**, 223-30.
- Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. *Genetics*. **155**, 945-59.
- ProCon (a).** 2016. States with Pending Legislation or Ballot Measures to Legalize Medical Marijuana – Medical Marijuana – ProCon.org. <http://medicalmarijuana.procon.org> Accessed 31 May 2017.
- ProCon (b).** 2016. For Which Symptoms or Conditions Might Marijuana Provide Relief? <http://medicalmarijuana.procon.org>. Accessed August 6 2016.
- Raymond M, Rousset F.** 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of heredity*. **86**, 248-9.
- Rieseberg LH.** 1995. The role of hybridization in evolution: old wine in new skins. *American Journal of Botany*. **82**, 944-53.
- Rieseberg LH.** 1997. Hybrid origins of plant species. *Annual review of Ecology and Systematics*. **28**, 359-89.
- Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB.** 1995. The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics*. **90**, 43-8.
- Rousset F.** 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*. **8**, 103-6.

- Russo EB.** 2007. History of cannabis and its preparations in saga, science, and sobriquet. *Chemistry & Biodiversity*. **4**, 1614-48.
- Sarri V, Baldoni L, Porceddu A, Cultrera NG, Contento A, Frediani M, Belaj A, Trujillo I, Cionini PG.** 2006. Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. *Genome*. **49**, 1606-15.
- Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S.** 2015. The genetic structure of marijuana and hemp. *PloS one*. **10**, e0133292.
- Schlichting CD.** 1986. The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics*. **17**, 667-93.
- Schultes RE.** 1970. The botanical and chemical distribution of hallucinogens. *Annual Review of Plant Physiology*. **21**, 571-98.
- Schwabe AL, Hubbard AR, Neale JR, McGlaughlin ME.** 2013. Microsatellite loci development for rare Colorado *Sclerocactus* (Cactaceae). *Conservation Genetics Resources*. **5**, 69-72.
- Schwabe AL, Neale JR, McGlaughlin ME.** 2015. Examining the genetic integrity of a rare endemic Colorado cactus (*Sclerocactus glaucus*) in the face of hybridization threats from a close and widespread congener (*Sclerocactus parviflorus*). *Conservation Genetics*. **16**, 443-57.
- SeedFinder.** 2017, Clone Only Strains. <http://en.seedfinder.eu/database/strains/cloneonly/>. Accessed May 31 2017.
- Small E.** 1997. Cannabaceae. *Flora of North America* Editorial Committee, editors. *Flora of North America North of Mexico*. New York and Oxford. vol. **3**, p. 381-387.
- Small E.** (a) 2015. Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. *The Botanical Review*. **81**, 189-294.
- Small E.** (b) 2015. Response to the erroneous critique of my *Cannabis* monograph by RC Clarke and MD Merlin. *The Botanical Review*. **81**, 306-16.
- Small E.** 2017. *Cannabis: A Complete Guide*. CRC Press: Taylor and Francis.
- Smith MH.** 2012. *Heart of Dankness: Underground Botanists, Outlaw Farmers, and the Race for the Cannabis Cup*. Broadway Books.
- Soler S, Gramazio P, Figàs MR, Vilanova S, Rosa E, Llosa ER, Borràs D, Plazas M,**

- Prohens J.** 2017. Genetic structure of *Cannabis sativa* var. *indica* cultivars based on genomic SSR (gSSR) markers: Implications for breeding and germplasm management. *Industrial Crops and Products*. **104**, 171-8.
- S'tajner N, Rusjan D, Korosec-Koruza Z, Javornik B.** 2011. Genetic characterization of old Slovenian grapevine varieties of *Vitis vinifera* L. by microsatellite genotyping. *American Journal of Enology and Viticulture*. ajev-2011.
- Stockton N.** 2015. Sorry, but the names for weed strains are kinda meaningless. *Wired - Science*. <http://www.wired.com/2015/08/sorry-names-weed-strains-kind-meaningless/>. Accessed August 15 2016.
- The Marihuana Tax Act of 1937.** 1937. Pub. 238, 75th Congress, Aug 2, 1937, 50 Stat. 55.
- Tomida I, Pertwee RG, Azuara-Blanco A.** 2004. Cannabinoids and glaucoma. *British Journal of Ophthalmology*. **88**, 708-13.
- United Nations Office on Drugs,** 2010. Crime. World Drug Report. United Nations Publications. https://www.unodc.org/documents/wdr/WDR_2010/World_Drug_Report_2010_lo-res.pdf. Accessed 31 May 2017.
- United States Department of Agriculture.** 1989. United States Plant Variety Protection Act of 24 December 1970. USDA. <https://www.ams.usda.gov/sites/default/files/media/Plant%20Variety%20Protection%20Act.pdf>. Accessed May 31 2017.
- United States Department of Agriculture.** 2015. Agricultural Marketing Service, Plant Variety Protection Office Application Requirements, Guidelines Exhibit B- Statement of Distinctness. <https://www.ams.usda.gov/sites/default/files/media/Exhibit%20B.pdf>. Accessed August 8 2016.
- United States Department of Agriculture.** 2016. Agricultural Marketing Service, What is a Specialty Crop? <https://www.ams.usda.gov/services/grants/scbgp/specialty-crop> Accessed August 2 2016.
- Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P.** 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. **4**, 535-8.
- Vergara D, Baker H, Clancy K, Keepers KG, Mendieta JP, Pauli CS, Tittes SB, White KH, Kane NC.** 2016. Genetic and genomic tools for *Cannabis sativa*. *Critical Reviews in Plant Sciences*. **35**, 364-77.
- Wikileaf.** 2018. Cannabis Strain Research Center. 2017. <http://www.wikileaf.com>. Accessed April 30 2018.

Zha HG, Milne RI, Sun H. 2008. Morphological and molecular evidence of natural hybridization between two distantly related *Rhododendron* species from the Sino-Himalaya. *Botanical Journal of the Linnean Society*. **156**, 119-29.

Table 1

Cannabis samples (122) from 30 strains with the reported proportion of Sativa from Wikileaf (Wikileaf, 2018) and the city location and state where each sample was acquired. (SLO: San Luis Obispo).

Name	Sativa	City	State	Name	Sativa	City	State
Durban Poison	100	Boulder 1	CO	OG Kush	55	Denver 3	CO
Durban Poison	100	Boulder 3	CO	OG Kush	55	Fort Collins 3	CO
Durban Poison	100	Denver 1	CO	OG Kush	55	Garden City 2	CO
Durban Poison	100	Denver 2	CO	OG Kush	55	SLO 1	CA
Durban Poison	100	Fort Collins 3	CO	Blue Dream	50	Boulder 1	CO
Durban Poison	100	Fort Collins 4	CO	Blue Dream	50	Boulder 2	CO
Durban Poison	100	Garden City 1	CO	Blue Dream	50	Boulder 3	CO
Durban Poison	100	Garden City 2	CO	Blue Dream	50	Denver 1	CO
Durban Poison	100	Union Gap 1	WA	Blue Dream	50	Garden City 4	CO
Hawaiian	90	Boulder 1	CO	Blue Dream	50	Garden City 4	CO
Hawaiian	90	Fort Collins 2	CO	Blue Dream	50	SLO 2	CA
Sour Diesel	90	Boulder 1	CO	Blue Dream	50	SLO 3	CA
Sour Diesel	90	Boulder 3	CO	Blue Dream	50	SLO 4	CA
Sour Diesel	90	Greeley 1	CO	Tahoe OG	50	Boulder 1	CO
Sour Diesel	90	Denver 4	CO	Tahoe OG	50	Denver 1	CO
Sour Diesel	90	Fort Collins 3	CO	Tahoe OG	50	Fort Collins 4	CO
Sour Diesel	90	Garden City 1	CO	Tahoe OG	50	SLO 3	CA
Sour Diesel	90	Garden City 2	CO	ChemdawgD	40	Boulder 1	CO
Trainwreck	90	Denver 1	CO	ChemDawg	45	Boulder 2	CO
Trainwreck	90	Garden City 1	CO	ChemDawg	45	Boulder 3	CO
Island Sweet Skunk	80	Boulder 1	CO	ChemdawgD	40	Denver 1	CO
Island Sweet Skunk	80	Garden City 1	CO	Chemdawg 91	40	Denver 5	CO
Island Sweet Skunk	80	Garden City 2	CO	Chemdog 1	40	Garden City 1	CO
AK-47	65	Boulder 1	CO	ChemDawg	45	Garden City 2	CO
AK-47	65	Denver 3	CO	Headband	45	Garden City 1	CO
AK-47	65	SLO 2	CA	Headband	45	Greeley 1	CO
Golden Goat	65	Boulder 1	CO	Banana Kush	40	Denver 1	CO
Golden Goat	65	Boulder 2	CO	Banana Kush	40	Garden City 1	CO
Golden Goat	65	Boulder 3	CO	Banana Kush	40	Garden City 2	CO
Golden Goat	65	Denver 1	CO	Banana Kush	40	Greeley 1	CO
Golden Goat	65	Garden City 1	CO	Girl Scout Cookies	40	Boulder 1	CO
Golden Goat	65	Garden City 1	CO	Girl Scout Cookies	40	Denver 1	CO
Golden Goat	65	Garden City 2	CO	Girl Scout Cookies	40	Fort Collins 2	CO
Green Crack	65	Fort Collins 2	CO	Girl Scout Cookies	40	Garden City 2	CO
Green Crack	65	Garden City 1	CO	Girl Scout Cookies	40	Garden City 3	CO
Green Crack	65	SLO 2	CA	Girl Scout Cookies	40	SLO 3	CA
Bruce Banner	60	Boulder 1	CO	Girl Scout Cookies	40	SLO 4	CA
Bruce Banner	60	Denver 1	CO	Girl Scout Cookies	40	Union Gap 1	WA
Bruce Banner	60	Denver 4	CO	Jack Flash	55	Boulder 1	CO
Bruce Banner	60	Fort Collins 3	CO	Jack Flash	55	Denver 3	CO
Bruce Banner	60	Fort Collins 4	CO	Larry OG	40	Boulder 1	CO
Bruce Banner	60	Garden City 1	CO	Larry OG	40	Denver 4	CO
Flo	60	Boulder 1	CO	Larry OG	40	SLO 3	CA
Flo	60	Denver 1	CO	G-13	30	Boulder 3	CO
Flo	60	Fort Collins 2	CO	G-13	30	Fort Collins 3	CO
Flo	60	Garden City 1	CO	G-13	30	Garden City 2	CO
Jillybean	60	Garden City 1	CO	Lemon Diesel	30	Boulder 1	CO
Jillybean	60	Garden City 2	CO	Lemon Diesel	30	Garden City 2	CO
Jillybean	60	Greeley 1	CO	Hash Plant	20	Fort Collins 3	CO
Pineapple Express	60	Boulder 1	CO	Hash Plant (Australian)	20	Garden City 1	CO
Pineapple Express	60	Denver 1	CO	Hash Plant	20	Garden City 1	CO
Pineapple Express	60	Garden City 2	CO	Hash Plant	20	Garden City 2	CO
Pineapple Express	60	Longmont 1	CO	Bubba Kush 98	20	Denver 1	CO
Pineapple Express	60	Union Gap	WA	Pre-98 Bubba Kush	15	Fort Collins 3	CO
Purple Haze	60	Denver 4	CO	Grape Ape	0	Boulder 1	CO
Purple Haze	60	Greeley 1	CO	Grape Ape	0	Union Gap 1	WA
Purple Haze	60	Fort Collins 1	CO	Purple Kush	0	Denver 1	CO
Tangerine	60	Denver 1	CO	Purple Kush	0	Garden City 3	CO
Tangerine	60	Garden City 1	CO	Purple Kush	0	Garden City 4	CO

Jack Herer	55	Garden City 3	CO
Jack Herer	55	SLO 1	CA
Jack Herer	55	Union Gap 1	WA

Table 2

Cannabis samples (122) from 30 strains with the reported proportion of Sativa retrieved from Wikileaf (Wikileaf, 2018). Strains arranged by proportion of Sativa, from reported pure Sativa to pure Indica (which has no reported proportion of Sativa) and the proportions of membership for genotype 1 and genotype 2 from the STRUCTURE (Fig. 2) are reported as a percentage according to the proportion of inferred ancestry.

Asterisk indicates the twelve popular strains used in further analyses

Diamond indicates clone only strains (SeedFinder, 2018)

Strain	# Samples	Sativa Percentage	Genotype 1 (% average)	Genotype 2 (% average)	Standard Deviation
Durban Poison*	9	100	86	14	9.9
Hawaiian	2	90	61	39	27.58
Sour Diesel*	7	90	14	86	53.74
Trainwreck	2	90	59	41	21.92
Island Sweet Skunk	3	80	93	7	9.19
AK-47	3	65	55	45	7.07
Golden Goat**	7	65	68	32	2.12
Green Crack*	3	65	60	40	3.54
Bruce Banner*	6	60	19	81	28.99
Flo*	4	60	38	62	15.56
Jillybean	3	60	73	27	9.19
Pineapple Express*	5	60	62	38	1.41
Purple Haze	3	60	77	23	12.02
Tangerine	2	60	53	47	4.95
Jack Herer	3	55	66	34	7.78
OG Kush**	4	55	28	72	19.09
Blue Dream**	9	50	80	20	21.21
Tahoe OG	4	50	26	74	16.97
Chemdawg*	7	45	9	91	25.46
Headband	2	45	57	43	8.49
Banana Kush*	4	40	52	48	8.49
Girl Scout Cookies**	8	40	25	75	10.61
Jack Flash	2	40	96	4	39.6
Larry OG	3	40	7	93	23.33
G-13	3	30	50	50	14.14
Lemon Diesel*	2	30	85	15	38.89
Hash Plant	4	20	37	63	12.02
Pre98-Bubba Kush	2	15	7	93	5.66
Grape Ape	2	0	55	45	38.89
Purple Kush**	4	0	29	71	20.51

Table 3

Lynch & Ritland (1999) pairwise relatedness comparisons of overall *r*-means (Mean) and standard deviations (SD) for samples of 30 strains including *r*-mean and SD after the first and second (where possible) outliers were removed. Outliers were samples with the lowest *r*-mean. The twelve popular strains are indicated with an asterisk. Diamonds indicate clone-only strains (SeedFinder, 2018)

Strain	# Samples	Measure	All samples	Outlier 1 removed	Outlier 2 removed
Durban Poison*	9	Mean	0.31	0.43	0.58
		SD	0.40	0.37	0.30
Hawaiian	2	Mean	-0.115	-	-
		SD			
Sour Diesel*	7	Mean	0.44	0.57	0.60
		SD	0.29	0.22	0.18
Trainwreck	2	Mean	-0.001	-	-
		SD			
Island Sweet Skunk	3	Mean	0.682	1.000	-
		SD			
AK-47	3	Mean	0.158	0.446	-
		SD			
Golden Goat**	7	Mean	0.25	0.31	0.46
		SD	0.32	0.36	0.36
Green Crack*	3	Mean	0.375	0.885	-
		SD			
Bruce Banner*	6	Mean	0.30	0.51	0.90
		SD	0.51	0.50	0.05
Flo*	4	Mean	0.29	0.55	-
		SD	0.38	0.39	-
Jillybean	3	Mean	-0.033	0.039	-
		SD			
Pineapple Express*	5	Mean	0.02	0.04	0.13
		SD	0.16	0.17	0.19
Purple Haze	3	Mean	0.041	0.263	-
		SD			
Tangerine	2	Mean	-0.219	-	-
		SD			
Jack Herer	3	Mean	0.102	0.127	-
		SD			
OG Kush**	4	Mean	0.13	0.25	-
		SD	0.19	0.22	-
Blue Dream**	9	Mean	0.50	0.63	0.76
		SD	0.39	0.34	0.24
Tahoe OG	4	Mean	0.210	0.406	0.539
		SD			
Chemdawg*	7	Mean	0.42	0.51	0.64
		SD	0.31	0.31	0.28

Headband	2	Mean	0.107	-	-
		SD			
Banana Kush*	4	Mean	0.13	0.24	-
		SD	0.20	0.13	-
Girl Scout Cookies**	8	Mean	0.08	0.13	0.22
		SD	0.27	0.30	0.32
Jack Flash	2	Mean	0.621	-	-
		SD			
Larry OG	3	Mean	0.316	0.671	-
		SD			
G-13	3	Mean	0.286	0.562	-
		SD			
Lemon Diesel*	2	Mean	0.102	-	-
		SD			
Hash Plant	4	Mean	0.250	0.250	0.427
		SD			
Pre98-Bubba Kush	2	Mean	-0.024	-	-
		SD			
Grape Ape	2	Mean	-0.050	-	-
		SD			
Purple Kush**	4	Mean	0.03	0.16	-
		SD	0.21	0.22	-

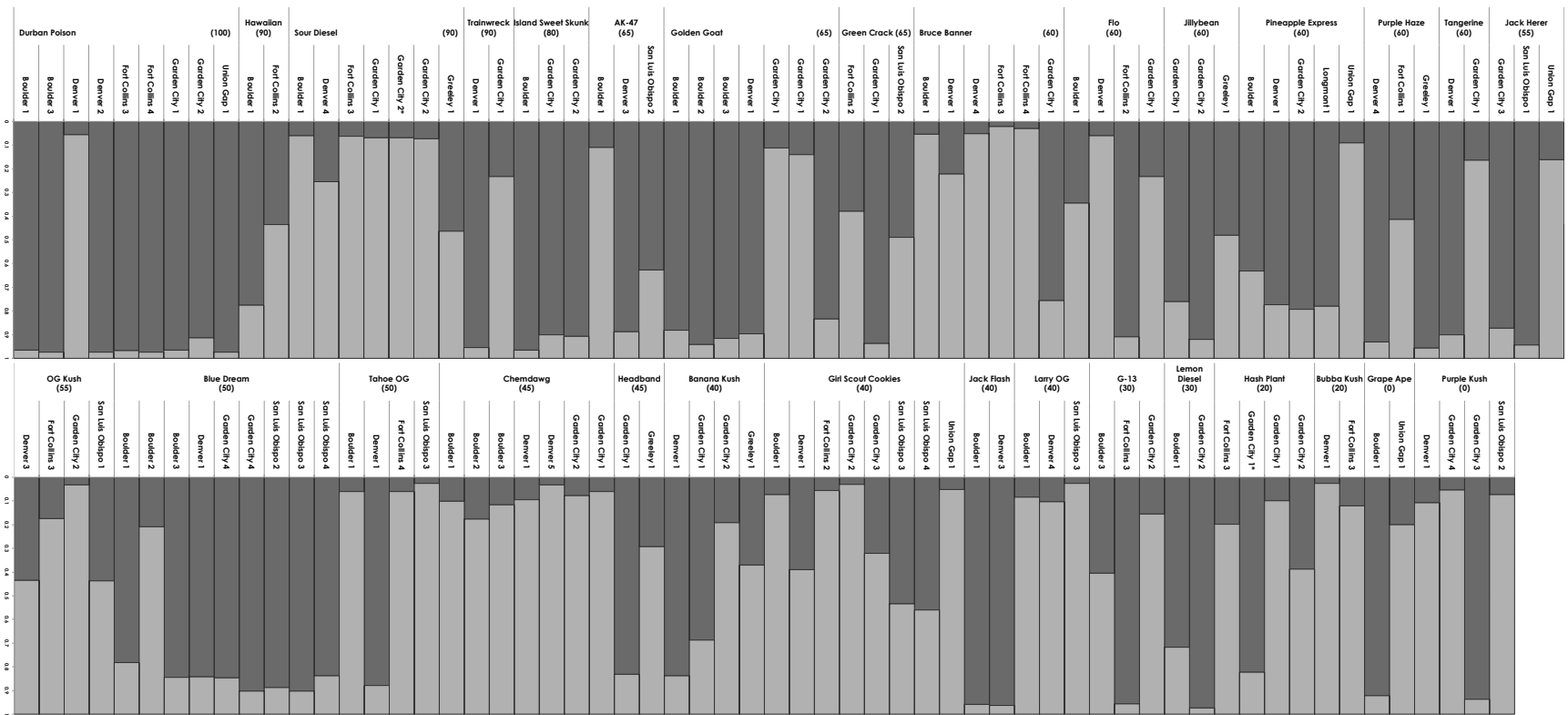


Fig. 1

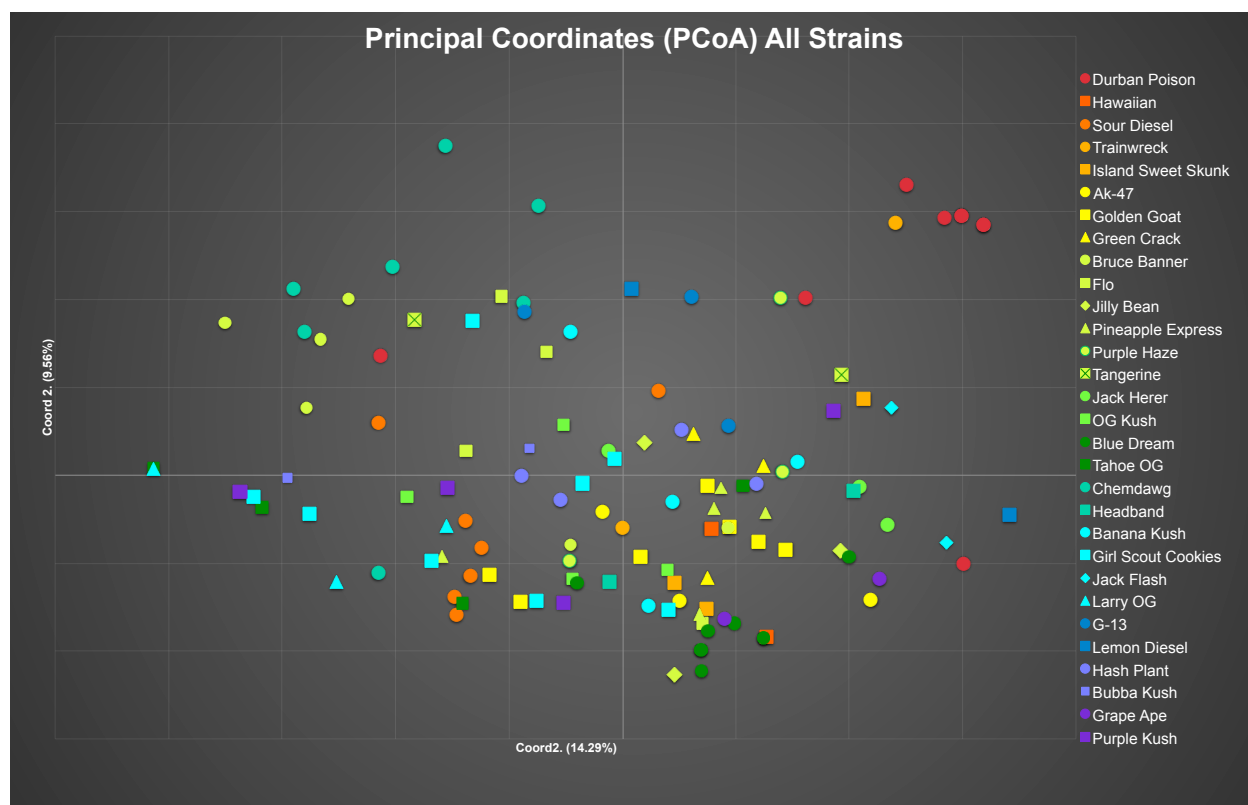


Fig. 2

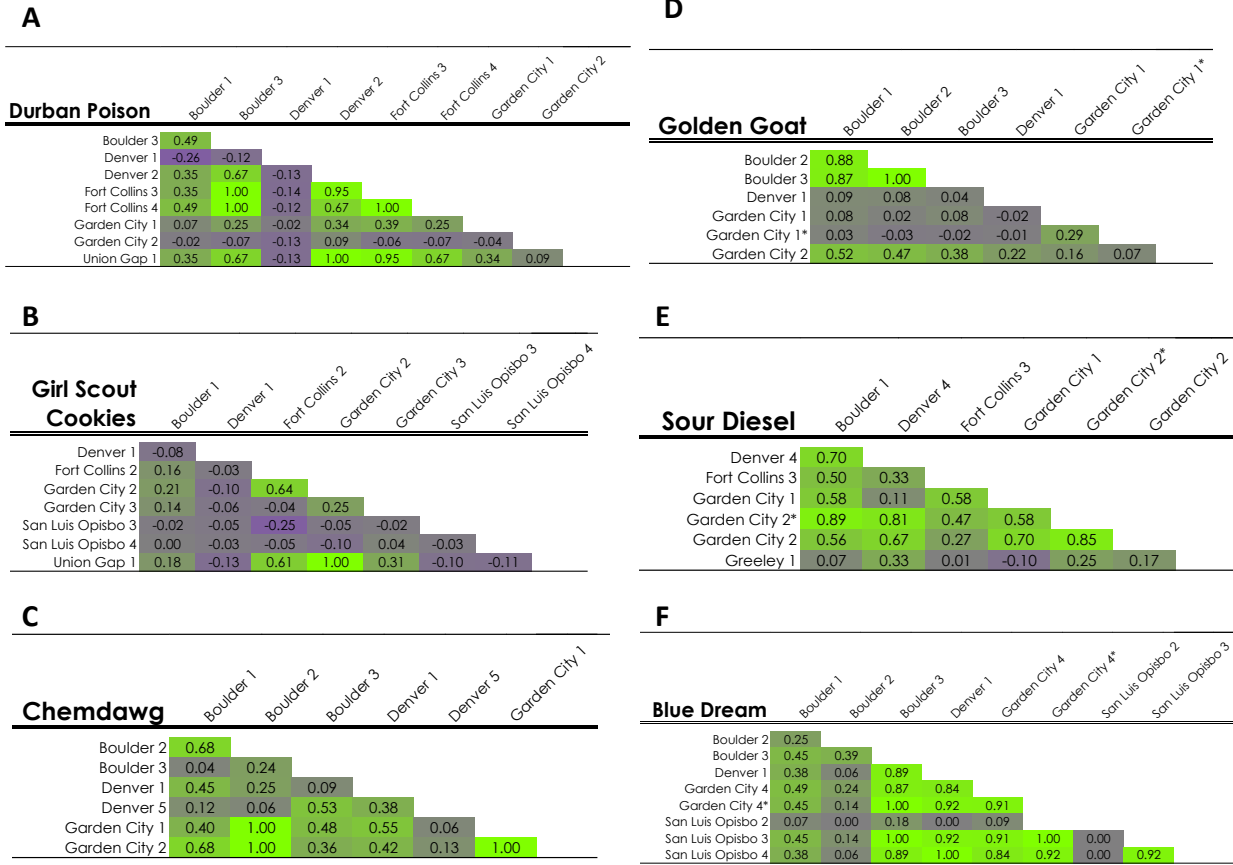


Fig. 3

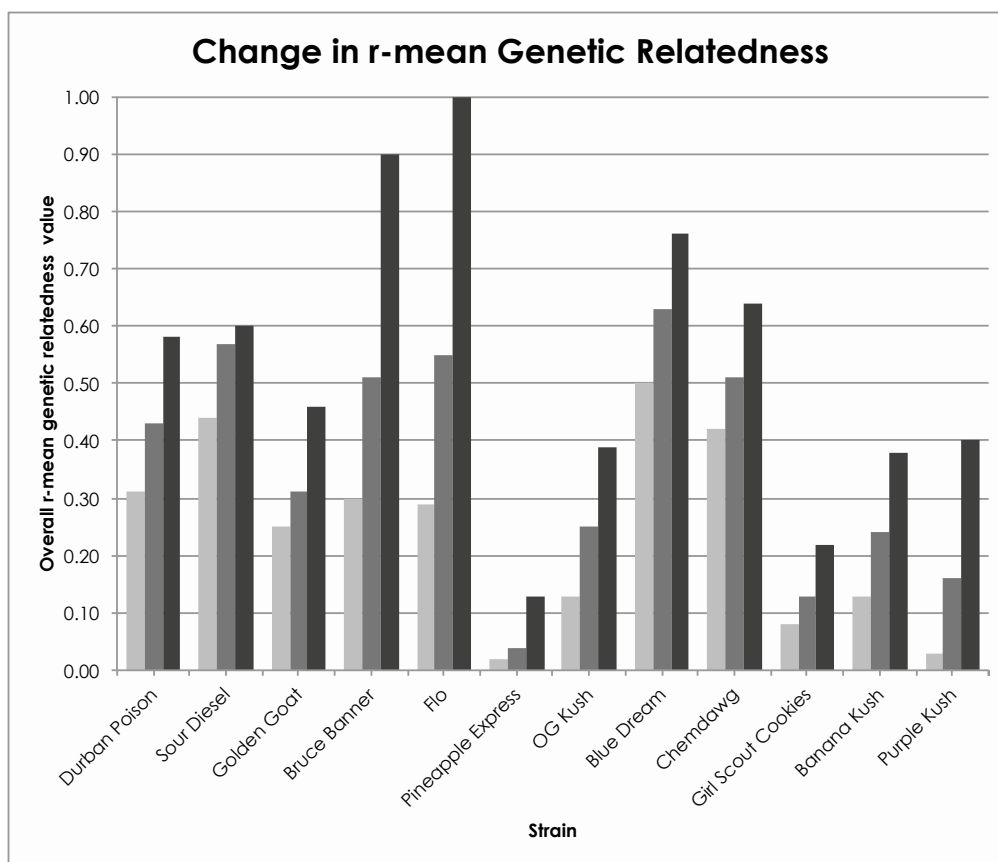


Fig. 4