University of Northern Colorado
Scholarship & Creative Works @ Digital UNC

Dissertations

Student Research

5-2019

A Multifaceted Approach to Address Variation in Cannabis Sativa

Anna Louise Schwabe

Follow this and additional works at: https://digscholarship.unco.edu/dissertations

Recommended Citation

Schwabe, Anna Louise, "A Multifaceted Approach to Address Variation in Cannabis Sativa" (2019). *Dissertations*. 554.
https://digscholarship.unco.edu/dissertations/554

This Text is brought to you for free and open access by the Student Research at Scholarship & Creative Works @ Digital UNC. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholarship & Creative Works @ Digital UNC. For more information, please contact Jane.Monson@unco.edu.

© 2019

ANNA LOUISE SCHWABE

ALL RIGHTS RESERVED

UNIVERSITY OF NORTHERN COLORADO
Greeley, Colorado

The Graduate School

A MULTIFACETED APPROACH TO ADDRESS VARIATION IN *CANNABIS SATIVA*

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Anna Louise Schwabe

College of Natural and Health Sciences
School of Biological Sciences
Biological Education

May 2019

This Dissertation by: Anna Louise Schwabe

Entitled: *A Multifaceted Approach to Address Variation in* Cannabis sativa
has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Natural and Health Sciences in School of Biological Sciences, Program of Biological Education.

Accepted by the Doctoral Committee

____

Mitchell McGlaughlin, Ph.D., Chair

____

Robert Reinsvold, Ph.D., Committee Member

____

Nolan Kane, Ph.D., Committee Member

____

Richard Hyslop, Ph.D., Committee Member

Date of Dissertation Defense ______March 1, 2019___________

Accepted by the Graduate School

____

Linda L. Black, Ed.D., LPC
Associate Provost and Dean
Graduate School and International Admissions
Research and Sponsored Projects

ABSTRACT

Schwabe, Anna Louise. *A Multifaceted Approach to Address Variation in* Cannabis sativa. Published Doctor of Philosophy Dissertation, University of Northern Colorado, 2019.

For thousands of years, humans have cultivated and dispersed *Cannabis sativa* L. across the globe. Although *Cannabis* has been largely illegal worldwide for decades, public perceptions and attitudes are changing. Increasing interest in potential *Cannabis* usage worldwide and nationwide is leading to less restrictions to make way for an expanding and lucrative industry with numerous applications. Although only one species is formally recognized in the *Cannabis* genus, thousands of years of artificial selection for diverse phenotypes and uses have resulted in two major usage groups; hemp-types which are defined worldwide as having very low levels of THC (< 1.0%), and drug-types which exceed a specified level of THC that varies among nations. The drug-type category includes three commonly used subcategories including Sativa, Indica and Hybrid types, and newly developed high CBD varieties that have more THC than hemp-types but are not bred for high THC. The quality of federally produced *Cannabis* for medical studies in the U.S. has recently been brought into question, and we included samples to determine the genetic relationship to these groups.

Phenotypic variation in *Cannabis* gives rise to commonly referenced categories, but sources of variation are unclear and understudied. Phenotypes are observable characteristics that results from a combination of both genotype and the environment.

The preferred method of propagation for *Cannabis* is cloning, and therefore variation within varietals should be from differences in environmental factors. Ten microsatellite markers were developed *de-novo* to investigate four aims: (1) genetic variation within strains, (2) genetic relationships among the common categories, (3) if genetic variation is detectable through olfactory sensation, and (4) how genetic variation is reflected in phytochemical levels. This dissertation includes four manuscript chapters representing each aim and uses a genetic basis for a multifaceted approach to investigate variation in *Cannabis sativa.* Substantial genetic variation was found within strains from obtained from different facilities. Genetic divergence between hemp and drug-types was genetically supported, but the Sativa, Indica, and Hybrid subcategories were not genetically well defined. The high CBD strains appear to bridge the genetic gap between hemp and drug-types, and federally grown research grade marijuana was genetically more similar to hemp than *Cannabis* available through the legal cannabis market. Genetic imposters within a strain had measurable aromatic differences, but there was considerable variation in aromas among samples with identical genetic identity. Analyses of both terpene and cannabinoid profiles among individuals with identical genotypes acquired from different sources varied considerably indicating environmental variation has a substantial impact on phenotype in *Cannabis.*

Together these results show a need for the *Cannabis* industry to implement regulatory checks in the form of genetic testing in order to provide consistency, especially for medical applications. These results demonstrate the need for genotyping in order for phenotypic consistency to be achieved if standard growing conditions can be established. When genetic verification and standard protocols are established, deviations

in phenotypic changes can be identified and disclosed to consumers so they are aware that there may be abnormal effects. This investigation highlights the need for additional research to provide consistent products, which is especially important for medical marijuana flower products. In order to provide consumers consistent products, it is imperative to understand sources of variation. Consumers deserve to be provided with quality consistent products as the industry continues to grow on a global scale.

ACKNOWLEDGEMENTS

I got this crazy idea to pursue a PhD after the legalization of adult recreational *Cannabis* use and possession in Colorado. Following discussions of variations in strains obtained from different dispensaries, it dawned on me that I could research why that would be given my extensive background in genetic relationships among closely species and plant populations of the same species. However, I knew little about *Cannabis,* marijuana, hemp, strains, the history of the plant etc.- so I had to learn everything from the ground up (literally), and there have been so many people who have helped me through this journey of the pursuit of knowledge, and I would like to try to thank each and every one of you.

First, I would like to thank Dr. M (you know who you are) for bringing the issue of strain variation to my attention and giving me such a kickass idea and the beginning point for my research. At such a pivotal point in *Cannabis* reform, I was able to pitch the idea and build a case for my doctoral research to Mitchell McGlaughlin, who has been my advisor and mentor throughout this process. So, most importantly, I need to give my biggest chunk of the gratitude wheel to him. Mit has been the best research advisor and mentor I have ever had- although I have nothing to compare him to, because he was also my advisor for my master's thesis. However, I wouldn't have been able to follow such a crazy idea researching something that was still taboo and largely illegal around the world without support from Mit. He has been supportive and encouraging; he has been my "that's a great idea, but let's think about this" guy; he gave me the idea for the salad

spinner (worked like a charm); he was always there when I called “Mit! Mit! Mit!” outside his office door for help; he was there to help field media interviews to make sure I knew what I was talking about (turns out I do); he let me use his lab and supplies, and he even paid for some stuff (Botany 2018). I also thank Susan Keenan, who not only let me come back for a third degree at UNC but also advocated for my *Cannabis* research.

Thank you to my committee members, Rob Reinsvold, Richard Hyslop and Nolan Kane. Rob, I know you are probably sick of me, but I promise this time I am leaving ;). Richard, thank you for holding my hand and having the patience to walk me through the chemistry stuff since it has been about 12 years since I last sat in a chemistry class. Nolan Kane, I still can’t even believe you said yes to sitting on my committee. When I first learned about Nolan and the *Cannabis* research in his lab at the University of Colorado, I was star-struck. Sending the email to request him as a committee member felt about as promising as asking Snoop to perform at my birthday party, but he said YES! Nolan is amazing- smart and kind, and so funny- I don’t actually think he knows how funny is. Nolan encouraged me to have confidence, and not to let all the dude-bro’s in the industry talk down to me, because apparently I know what I’m talking about. Seriously though, thank you so much ☺

Thank you to Daniela Vergara who I first met at the Cannabis Research Institute Conference. I was so nervous to be presenting my work, but she followed my presentation and announced to her audience (standing room only), that we had found the same things using different tools, which was so nice of her, and helped give me (a nobody at the time) legitimacy. Since then, she has encouraged me to get tough, because the *Cannabis* industry is fraught with self-proclaimed experts who are more than willing

to share how much more they know than you (even when they actually don't). Don't worry Daniela, we got this.

Avery Gilbert, the scent scientist. Thank you for reaching out to me and offering a wonderful collaboration study. What we have done together is pretty groundbreaking, and you were the catalyst. I have learned so much from you about the perception of smell and am so very glad to have met you. I'm sure we will stay in touch and meet up from time to time over a beer as we discuss the aromatic descriptions of beers, wines, and marijuana, that are sometimes hilarious.

Sincerest thanks to Connor Hansen, my superstar undergraduate researcher. It started with a small project looking at relationships along the *Cannabis* spectrum with a few samples from each group. Just some hemp, some drug-types, some high CBD samples, and a couple of NIDA samples from the University of Mississippi. Well, that turned into a pretty fantastic piece of work, and I couldn't ask for a better research assistant. I'm sure you will be using your chemistry degree in the future, but you also have a pretty amazing bit of population genetics in your back pocket now too. I would also like to thank my lab-mates Ryan Fuller (\m/), Brandee Anderson (love you and your granny blanket to pieces), Nathan Redecker (badass Viking conservationist), and Emily Schumacher (my official statistics consultant), and last but not least, my dearest Samamami. Meow =^..^=

To my friends- I love you all!! I couldn't possibly name you all, but I couldn't have done any of this without the love and support I have from all the wonderful people who I am humbled to call my friends. Thank you to Tyler Sherman- we've been traveling this road for a good long while and you have always been there to discuss interesting

things with and have “lab meetings” (except when you had yoga!), Kristy “The White Tornado” Kappenman- my biggest cheerleader and craziest coolest friend, Heather Rudolph- thanks for letting me fix the A&P labs and I love you, Jonna Jackson- I miss your face, maybe one day we can work together on the effects of *Cannabis* in layer V of the prefrontal cortex, Tom McCabe- BFFs, Laura Heiker- my friend and sister, and finally for those who are named, Vanessa Johnson- A&P guru, YouTube video star, and member of the Friday Foxes (Michele, Jonna, Kristy, Vanessa, Randi and sometimes Lucas).

To all the graduate students in the biology department, thank you. It was a pleasure getting to know the most amazing group of scientists who study so many different fields and who produce fabulous research. I’m going to miss you when they actually make me leave.

I have to give recognition and gratitude to all the *Cannabis* experts who have taught me so much over the last few years. Some whom I have met in person, some I have not. Caren Kershner who I respect and admire for her love of the plant and also for providing hemp research material, Matt Kahl who has tirelessly been advocating for veterans with PTSD and bringing education about *Cannabis* therapy to the forefront, Joel “Hemp Hero” Bedard who knows probably the most things about *Cannabis*- we have a lot of work to do my friend, Frank Conrad who I got to know by attending his “Frank Talks” and learned so much about how little we know, Cindy Blair who keeps Frank under control and is just lovely, Greg Duran who I always enjoyed chatting with about problems and how to start figuring out solutions, Teri Robnett who is relentless in her medical marijuana patient advocacy- I’ll fight along with you to get these patients what

they need, Carter Baird (aka "The Pot Scientist")- get back on YouTube and go get that doctoral degree! I just adore you. Big shout out and gigantic thank you to Jason Lopez from NANO Rx who is not only super smart and driven like a force 5 hurricane, he dedicated time to talk to me, time to fly out and meet me, and also donated funds so I can get my research published and out to the people. Finally, Cheysser Harding who tracked down seeds for one of my projects and connected me with the scientists at Mile High Labs so I could analyze my samples- I don't think I would have finished when I did without you.

I saved the best for last. Thank you to my family. My parents who, although they were anti-marijuana when I began this journey, they supported me and my crazy weed project. They live on the other side of the world in Uruguay (oh, the irony), but they showed up at least once a year to give me all their love and support. Ten years ago, they were grateful that I finally decided to go back to school to finish my bachelor's degree. I seriously doubt they ever envisioned me in graduate school, let alone pursuing a doctoral degree in *Cannabis* genetics (of all the things) and biological education. My mother quotes me as saying "I don't want to spend my whole life in a lab coat!". Whoops. I guess I do want to do that after all. My dad has become more interested in the global landscape of *Cannabis*, and always keeps an eye out for me. I love you guys! My kiddo Gavin, who has to put up with me all the time, and vice versa. I love you so much. Thank you for always being there and listening as well as teaching me. You knew more about *Cannabis* than I did when I got into this project! Finally, I can't end this string of emotional outpouring without thanking my other half, Brett. He has stuck by my side through two graduate degrees and has been unconditionally loving and supportive. Thank

you for your patience as I sat on the internet reading articles, graded assignments, constructed PowerPoint lectures, and stressed about comps. Thank you for traveling with me to conferences and letting me practice my talks with you. Thank you for your ideas and feedback, thank you for calming me down when I got all riled up, and thank you for the music

TABLE OF CONTENTS

CHAPTER
I. INTRODUCTION AND REVIEW OF THE LITERATURE……………1

Introduction……………………………………………………………1
Overview of *Cannabis sativa*……………………………………………3
Cannabis Breeding System………………………………………………5
Intellectual Property Protection for Plant Varieties……………………7
Cannabis Consistency……………………………………………………9
Cannabinoids and Terpenes……………………………………………10
History of Prohibition……………………………………………………12
Current Status……………………………………………………………13
Medical Applications……………………………………………………17
Genetic Research…………………………………………………………18
Summary…………………………………………………………………21

II. GENETIC VARIATION PART 1: GENETIC TOOLS WEED OUT MISCONCEPTIONS OF STRAIN RELIABILITY IN CANNABIS SATIVA: IMPLICATIONS FOR A BUDDING INDUSTRY…………23

Abstract…………………………………………………………………24
Introduction………………………………………………………………26
Methods…………………………………………………………………33
Genetic Material…………………………………………………………33
Microsatellite Development ……………………………………………34
Polymerase Chain Reaction and Data Scoring…………………………40
Genetic Statistical Analyses……………………………………………40
Results……………………………………………………………………41
Discussion………………………………………………………………54
Conclusion………………………………………………………………61

III. GENETIC VARIATION PART 2: RESEARCH GRADE MARIJUANA SUPPLIED BY THE NATIONAL INSTITUTE ON DRUG ABUSEIS GENETICALLY DIVERGENT FROM COMMERCIALLY AVAILABLE CANNABIS…………………63

Abstract…………………………………………………………………64
Introduction………………………………………………………………65
Methods…………………………………………………………………69
Sampling…………………………………………………………………69

Deoxyribonucleic Acid Extraction……………………………………72
Statistical Analyses……………………………………………………72
Results……………………………………………………………………74
Genetic Differentiation…………………………………………………74
Clustering Analysis……………………………………………………75
Genetic Relatedness Network…………………………………………82
Discussion………………………………………………………………84
Conclusion………………………………………………………………87

IV. SENSORY VARATION: HUMAN OLFACTORY DISCRIMINATION OF GENETIC VARIATION IN CANNABIS STRAINS……………………………………………………………………88

Abstract…………………………………………………………………89
Introduction……………………………………………………………90
Methods…………………………………………………………………94
Genetic Methods………………………………………………………94
Strain Selection…………………………………………………………94
Genetic Material………………………………………………………95
Deoxyribonucleic Acid Extraction, Polymerase Chain Reaction, and Fragment Analysis……………97
Genetic Statistical Analysis……………………………………………97
Sample Selection………………………………………………………98
Sensory Methods………………………………………………………99
Odor Stimuli……………………………………………………………99
Odor Descriptors………………………………………………………100
Rating Scale and Presentation…………………………………………100
Ethics Statement………………………………………………………100
Participants……………………………………………………………101
Sensory Statistical Analyses…………………………………………101
Genetic Results…………………………………………………………104
Genetic Relatedness……………………………………………………104
Clustering Analyses……………………………………………………105
Sample Selection for the Sensory Study………………………………110
Sensory Results…………………………………………………………111
Subject Demographics…………………………………………………111
Hierarchical Cluster Analysis…………………………………………111
Within Strain Descriptor Profiles………………………………………113
Additional Within-Strain Scent Profiling Analyses……………………119
Within Strain Scent Profiling: Perceived Shared Character Categories……………………119
Within Strain Scent Profiling: Cannabis Lexicon Categories……………………………121
Discussion………………………………………………………………122
Conclusion………………………………………………………………130

V. PHYTOCHEMICAL VARIATION……………………………..……..132

Abstract……………………………………………………………133
Introduction…………………………………………………………..134
Overview……………………………………………………..134
Chemical Constituents of Cannabis……………………….135
Cannabinoids………………………………………136
Terpenes………………….…………..……138
Monoterpenes…………………………...139
Sesquiterpenes…………………………..140
Summary……………………………………………..141
Methods……………………………………………………………141
Cannabinoids……………………………………………142
Terpenes…………………………………...………152
Results…………………….……………………………………….154
Cannabinoids……………………………………………154
Terpenes…………………………………………………...168
Discussion…………………………………………………………185
Conclusion…………………………………………………...193

VI. CONCLUSION……………………………….…………………….195
Conclusion…………………………………………………………195
Future Directions……..……………………………………………201

References……….……………………………...…………………203

LIST OF TABLES

Table:

1.1. Worldwide and U.S. Legalization History……………………..…..………16

2.1. Twelve popular strains and their described assignment of Sativa and Indica according to six online databases of *Cannabis*……………….31

2.2. *Cannabis* samples (122) from 30 strains…………………………….…...35

2.3. Summary of *Cannabis* samples (122) from 30 strains……………………37

2.4. Primer information…………...………………………...……………...39

2.5. Lynch & Ritland (Lynch and Ritland 1999) pairwise relatedness comparisons of overall *r*-means (Mean) and standard deviations (SD) for samples of 30 strains………………...………..…..52

3.1. Sampling Information..…………………..……………………………..…...70

3.2. Genetic Differentiation..……….……………………..…………………..…..75

4.1. The number of samples of five selected *Cannabis* strains..………………95

4.2. Information for all samples included in the genetic portion of this study………………………………………………………………...96

4.3. Samples used in the olfactory study………………………………………98

4.4. The 40 odor descriptors used to characterize the samples in this study in alphabetical order..…………………………………..……100

4.5. Perceived Shared Character categories..………………………….……….102

4.6. *Cannabis* Lexicon Categories..…………………………………………103

4.7. Results of Friedman’s test on the summed frequencies for each odor descriptor across all samples within a given strain for each *Cannabis* strain……………………...…………………………113

4.8. Frequencies plus mean and standard deviation of the top five pooled odor descriptors for the genetic consensus samples of "Durban Poison", "OG Kush", "Blue Dream" and "Mob Boss"……..………………………………………………….……116

4.9. Shared Character Category Frequency of Detection..……………..……120

4.10. The Frequency of Detection Scores..……………………..…..………121

5.1. HPLC Method Parameters ……………..…..……………….…………142

5.2. Injection Sequence………………………...…………………………143

5.3. Peak Relative retention time (RRT) and Relative Response Factor (RRF) values………………..………...…………………………143

5.4. Retention Times, Peak Height, Peak Area and Response Factor for the composite HPLC-DAD analysis.…………………..……145

5.5. Retention Times, Peak Height, Peak Area and Response for the RL standard HPLC-DAD analysis..……………….…..………147

5.6. Retention Times, Peak Height, Peak Area and Response for the sample 752 (OG Kush 2) HPLC-DAD analysis..……..…………149

5.7. Retention Times, Peak Height, Peak Area and Response for the blank HPLC-DAD analysis..……………….…..………...………150

5.8. Percentages of eight cannabinoid levels in four strains by HPLC-DAD as percent dry weight for eight cannabinoids, and calculated THC_{TOTAL}..……………………………………..…….155

5.9. Average terpene percent distribution by mass (mg) calculated as normalized area percentages of each of the terpenes present in the samples.………………………………………………….173

5.10. Fifteen samples from four strains with the THC (total % dry weight) measured by Mile High Labs (MHL)…………..…191

6.1. Significant pairwise Pearson correlation coefficients (R^2) between terpene levels (Chapter V) and scents detected in the olfactory study (Chapter IV)………………….. ………………....199

LIST OF FIGURES

Figure:

2.1. STRUCTURE HARVESTER……………………………………………42

2.1. STRUCTURE graphs for 122 individuals..………………………………44

2.2. Bar plot graphs generated from STRUCTURE analysis for individuals from twelve popular strains…………………….……………45

2.3. Principal Coordinates Analysis (PCoA) generated in GENALEX using Nei's genetic distance matrix..……………...………47

2.4. A genetic heat map chart of Lynch & Ritland pairwise genetic relatedness (*r*) values for 122 samples..…………………………………49

2.5. Heat maps of six prominent strains using Lynch & Ritland (Lynch and Ritland 1999) pairwise genetic relatedness (*r*) values……..…50

2.6. Iterative removal of pairwise genetic relatedness (*r*) outliers from Table 2.5……………………………..……….………..…..…53

3.1. Principal Coordinates Analysis..…………………………………………76

3.2. PC-Ord group linkage dendrogram..…………………..…………………78

3.3. STRUCTURE HARVESTER graph showing high support for two genetic groups……………………………………………………79

3.4. MavericK 1.0.5 thermodynamic integration evidence estimates normalized to a sum of 1.0..…….………………………..………………80

3.5. Bayesian clustering analysis from STRUCTURE……………...………81

3.6. EDENetworks genetic relationship network with incrementally decreasing stringency of required genetic relatedness among samples in the data set……………………………..……………...…83

4.1. Samples (1 gram) of dried cannabis flower..………………………..…...99

4.2. Lynch & Ritland pairwise genetic relatedness within each strain..…..…102

4.3 Principal Coordinate Analysis generated using ggplot in R Studio……..106

4.4. STRUCTURE HARVESTER……………………………..……………107

4.5. STRUCTURE graph with the proportion of genetic assignment to each of the three genetic groups as indicated by the proportion of each color in each bar representing an individual. ………..…………108

4.6. PC-Ord hierarchical genetic cluster analysis based on genetic distance color coded by strain..………………………….………………109

4.7. Hierarchical genetic cluster analysis of the cannabis samples….………110

4.8. Hierarchical cluster analysis (HCA) for 15 samples included in the sensory study containing two large clusters, Cluster B' and Cluster A'……………………………………………………….…112

4.9. Detection frequency of top-rated odor descriptors for "Durban Poison"………….…………………………………………..…114

4.10. Detection frequency of top-rated descriptors for consensus samples of "Blue Dream", "Mob Boss" and "OG Kush" and the mean for each descriptor………..…………………………………..…115

4.11. Detection frequency of top-rated odor descriptors for "OG Kush" samples…………………………………………………….117

4.12. Detection frequency of top-rated odor descriptors for "Blue Dream" samples…………………………………………….……118

4.13. Detection frequency of top-rated odor descriptors for "Mob Boss" samples………………………..…..……………………….……119

4.14. Shared Character Category Histograms. ……………………….….………120

4.15. Histograms for Frequency of Detection Scores of each *Cannabis* Lexicon category (Sweet, Sour, Spicy and Bitter)..………..…122

4.16. The glandular trichomes of the female flower of *Cannabis sativa* at 40X magnification.…………….………………………...………130

5.1. The glandular trichomes of the female flower of *Cannabis sativa* magnified 40X..……………..……...……………………………………136

5.2. The chemical structures of four monoterpenes..……………...…………140

5.3. The chemical structures of two sesquiterpenes..…………………..……140

5.4. Composite HPLC-DAD chromatogram……………………....…...……144

5.5. RL standard HPLC-DAD chromatogram.………………………………147

5.6. Sample HPLC-DAD chromatogram for sample 752 (OG Kush 2).......…148

5.7. HPLC-DAD chromatogram for the blank…….……………..…………150

5.8. Percent dry weight (mg) of eight cannabinoids measured by UV-HPLC in 15 samples of four commercially available strains…………………………………………………..…..…157

5.9. Chemotype by proportions of detected cannabinoids within the total cannabinoid fraction (% by dry weight) measured by UV-HPLC in 15 samples of four commercially available strains…...…...……………………………………..………158

5.10. Percent dry weight (mg) of THC and THCA (six cannabinoids removed)……..……………………………………………………………159

5.11. Percent dry weight (mg) of six minor cannabinoids (THC and THCA removed).……………………………………………159

5.12. Average levels of THCA (total % dry weight) in the genetic consensus samples compared to the genetic outlier with error bars of the standard deviation.…..………………………………………160

5.13. Average levels of the minor cannabinoids (total % dry weight) in the genetic consensus samples compared to the levels of the genetic outlier with error bars of the standard deviation.………………..161

5.14. Hierarchical clustering dendrogram for 15 samples of four strains based on levels of eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, and THCA).…….……………………………………162

5.15. Hierarchical clustering dendrogram for 15 samples within (A) “Durban Poison”, (B) “Blue Dream”, (C) “Mob Boss”, and (D) “OG Kush” based on levels of eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, and THCA).……..…163

5.16. Hierarchical clustering dendrogram for 15 samples of four strains based on levels of six minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV).……………………………..……….………..164

5.17. Hierarchical clustering dendrogram for 15 samples of (A) “Durban Poison”, (B) “Blue Dream”, (C) “Mob Boss”, and (D) “OG Kush” based on levels of six minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV)………………..……..……165

5.18. Principal Components Analysis of scaled by THCA on Axis 1 (99.002% of variation) and remaining variation on Axes 2 and 3 (0.900% and 0.079% of variation)……….…………………………..…169

5.19. Figure 5.10. Principal Components Analysis of scaled by CBGA on Axis 1 (99.925% of variation) and the remaining variation on Axis 2 (0.075% of variation)…………………….…………….….…170

5.20. Figure 5.10. Principal Components Analysis of scaled by CBGA on Axis 1 (99.925% of variation) and the remaining variation on Axis 2 (0.075% of variation)………………………..…………..…171

5.21. Terpene variation within each accession of four strains (% distribution) for (A) “Durban Poison”, (B) “Blue Dream”, (C) “Mob Boss”, and (D) “OG Kush”. The genetic outlier is indicated with an asterisk…………………….……………………………177-8

5.22. Twenty-one terpenes (A) “Blue Dream”, (B) “Mob Boss”, and (C) “OG Kush”, with the average for the genetic consensus samples and the genetic outlier…………………………………..………179

5.23. Top five terpenes with the highest levels pooled for (A) “Blue Dream”, (B) “Mob Boss”, and (C) “OG Kush”, with the average for the genetic consensus samples and the genetic outlier……………………………………………...………181

5.24. PCA clustering analysis. ………………………………………….………172

5.25. Average terpene levels in seven strains (% distribution by terpene)…….183

5.26. Terpene profiles for seven strains…………………………………….………184

CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

Cannabis sativa L. is one of the most interesting and useful plants with evidence of human cultivation dating back as far as 10,000 years (Abel 2013; Clarke and Merlin 2013; Okazaki et al. 2011; Small 2015a). As humans moved across the globe to various continents, they took *Cannabis* with them for utilitarian purposes such as fiber, food, fuel, and also for the plants medicinal and psychoactive properties (Clarke and Merlin 2013; Small 2015a). Human driven dispersal across the globe may have led to local adaptation, contributing to morphological variants. However, cultivation and selective breeding over thousands of years has arguably been the driving force behind the wide variation of phenotypes observed in modern *Cannabis.* Generally, two broad categories of *Cannabis* are recognized: hemp-types and drug-types. Hemp-types are grown for fiber, seeds, and non-psychoactive phytochemicals produced mainly by the flowers. Drug-types produce more of the psychoactive phytochemical precursor Δ^9-tetrahydrocannabinolic acid (THCA) and are labeled as such when THCA exceeds a predefined threshold of percent of dry weight, which varies by country. However, they are labeled as “THC” content which is the analyte measured due to decarboxylation during the analysis. While the level of THCA produced in the flower is the main distinction between hemp and drug types, usage, morphology, and cultivation methods differ between the two types. There is a third elusive type, *C. ruderalis*, which is smaller and flowers as a function of age rather

than photoperiod (Clarke and Merlin 2013; Small 2015b) but has been suggested that this type may be an escapee from hemp cultivation that has adapted to a weedy lifecycle (Clarke and Merlin 2013; Emboden 1974; Schultes et al. 1974). Cultivation and selective breeding over several millennia are thought to have altered the evolutionary trajectory of *Cannabis* to such an extent that natural, unaltered ancestral populations no longer exist (Small 2017). *Cannabis* cultivators have successfully bred thousands of varietals, and extensive genetic variation has been observed throughout the species (Clarke and Merlin 2016; Lynch et al. 2016; Pisupati et al. 2018; Soler et al. 2017). Recent legalization for medical and recreational consumption has increased access to *Cannabis* and information about *Cannabis*, leading to changes in the demographics of consumers (Han and Palamar 2018). The goal of this work is to bring additional awareness and information about the products consumers have access to and give more context to the information they are provided.

The *Cannabis* samples used in this study were collected from dispensaries, herbaria, and cultivators. Retail samples were legally purchased from dispensaries located in Colorado, Washington, and California. It was important to purchase samples anonymously since I wanted flowers that were representative of product supplied to customers without bias. Disclosing that the samples would be used in a research study could introduce the potential for producers to differentially select samples. The primary purpose of this investigation was to examine variation in *Cannabis* and to determine how genetic variation manifests as differences in phenotypic characters. The chapters herein address (1) genetic variation within strains, (2) genetic variation among categories, (3) if genetic variation is detectable through olfactory sensation, and (4) how genetic variation

is reflected in phytochemical levels. Variation in *Cannabis* can come from multiple sources, some of which are explored here to expose and highlight the need for additional research to provide consistent products, which is especially important for medical marijuana.

Overview of *Cannabis sativa*

Cannabis sativa L. is a member of the Cannabaceae with about 170 species in ten small genera including *Cannabis, Humulus*, and *Celtis* (McPartland 2018), although sources have conflicting information about the current taxonomy (Clarke and Merlin 2013; Integrated Taxonomic Information System 2014; The Plant List 2013). Although the Flora of North America recognizes only *Cannabis sativa* L. (Small 1997), many breeders and botanists support the polytypic taxonomy of *Cannabis* (Anderson 1980; Clarke and Merlin 2013; de Lamarck 1785; Emboden 1974). Whether the genus is comprised of one (*C. sativa*), two (*C. sativa* and *C. indica*) or three species (*C. sativa, C. indica* and *C. ruderalis*) remains a topic of debate. Monotypic *Cannabis sativa* includes narrow and broad leaf drug types, non-drug hemp types, and *C. ruderalis*, which is smaller and flowers as a function of age rather that photoperiod. Drug types are defined as any *Cannabis sativa* plant with total THC (THCA + THC) concentrations above a stated limit. Plants of broad and narrow leaf drug types, as well as hybrid variants, are commonly referred to as marijuana. Low total THCA defines hemp types with the legal limit varying among countries: for example 0.3% by dry weight in the U.S., 0.2% in the U.K., and 1.0% in Western Australia (European Monitoring Centre for Drugs and Drug Addiction 2018; Parliament of Western Australia 2004; United States Department of Agriculture and 113th United States Congress 2018). However, drug types generally have

much higher levels of THC compared to hemp, often reported as 12-25% THC in retail strains (Jikomes and Zoorob 2018; Potter et al. 2008; Vergara et al. 2017). There are also some varieties with relatively low THC that are nonetheless ranked as a drug type due to the THC limits defining hemp. Hemp and marijuana are genetically distinct (e.g. Lynch et al. 2016; Sawler et al. 2015; Schwabe and McGlaughlin 2018; Soler et al. 2017) and are distinguished further by levels of chemical constituents, particularly THC, as well as the products that will be made from the plant (Datwyler and Weiblen 2006; de Meijer et al. 1992; Lynch et al. 2016; Rustichelli et al. 1998).

Terminology to describe *Cannabis* is convoluted and conflicted among scientists, horticulturalists, taxonomists, enthusiasts and consumers. For the purposes of the work here, *Cannabis* and cannabis refer to any variety of *Cannabis sativa.* Drug-type and marijuana refer to varieties with > 0.3% THC. Hemp-type and hemp refer to varieties with low THC (< 0.3% in the U.S.). Sativa is used in the colloquial context to describe narrow-leafed strains with uplifting or energizing psychoactivity. Indica is used in the colloquial context to describe broad-leafed strains with relaxing and sedating psychoactive effects. Hybrid is used to describe varieties with a combination of morphologies and/or reported effects from both Sativa and Indica types. Variety is a group with distinct and uniform characters (physical, chemical, genetic) that are exhibited in all members of the group (Cervantes 2006). The term cultivar refers to a cultivated variety that is developed by a plant breeder through cross breeding, which can also be called a hybrid (Griess 2016). Plants grown from the seeds of a cultivar often will not display characteristics of a single parent since they were produced by crossing two distinct varieties (Griess 2016). True-to-type plants have the same genetics as the parent

(Griess 2016), and in *Cannabis* this is most commonly achieved through cloning. Although the term "strain" is botanically incorrect when applied to varieties of plants, the term is widely used in the cannabis industry to describe selections of cultivars of varieties, in part because many strains of *Cannabis* are not true varieties in that they are not genetically stable and do not breed true (Cervantes 2006). According to Cervantes (2006), strains in many cases do not have described defining characters and are merely hybrids of hybrids that have been given a unique name. Given the recency and exponential growth of the number of available strains, it is likely that many strains lack genetic stability and are unlikely to be produced consistently.

***Cannabis* Breeding System**

Cannabis is predominantly dioecious with separate sex chromosomes, which is rare for plants. Male and female flowers develop on separate plants, although occasionally hermaphrodites are observed (Moliterni et al. 2004). Male and female plants are virtually indistinguishable prior to flowering, although genetic tests for Deoxyribonucleic acid (DNA) polymorphisms and sex chromosomes may assist in early developmental sexing of *Cannabis* plants (Mandolino et al. 1999). Hemp types are commonly grown for their fibrous stem and seeds that are highly nutritious and rich in omega-3 and omega-6 fatty acids (Callaway 2004; Small 2016). Therefore, male and female hemp plants are grown together because when female flowers are fertilized, they produce large quantities of seed. Drug types are commonly grown for the female inflorescence, as it is mainly the female flowers that produce the glandular trichomes where the manufacturing of cannabinoids and aromatic terpenes occurs. Specific cannabinoid molecules bind to receptors in animals (McPartland et al. 2001) and elicit

various effects. Terpenes are produced in many different plants and are the primary constituents of essential oils, which are also thought to have therapeutic properties (Paduch et al. 2007). The dioecious breeding system is problematic for drug type *Cannabis* breeders because they seek to forego seed production and maximize inflorescence growth. Producers of drug type *Cannabis* remove pollen producing male plants from the population to prevent fertilization and seed production (Meier and Mediavilla 1998). The preference for production of unfertilized female flowers has led to the widely practiced artificial vegetative propagation via cloning of popular *Cannabis* drug type strains.

In order to produce novel *Cannabis* varieties, plants are cross- pollinated to produce seeds that have characteristics of both the parents (Cervantes 2006). Plants grown from seeds of the parental cross are called the first filial (F1) generation and are genetic hybrids of the parents (Cervantes 2006). Offspring resulting from the F1 generation seeds are assessed, and individual plants are selected based on desirable phenotypic traits (Cervantes 2006). In order to remove unwanted genetic traits from the lineage, it is necessary to continue crossing offspring of each subsequent generation until the offspring reliably and consistently exhibit the desired phenotype, and at this point the new variety is said to have stable genetics. Other crop and ornamental plants are often subjected to inbreeding for several generations to remove genetic variation, but as this is time consuming and because breeders are often limited by space, this important technique to developing a genetically stable variety may be cursory.

The legal medical and recreational *Cannabis* industries aim to produce consistent products for consumption and maximum yield, which is generally achieved through

vegetative cloning techniques. Female plants are selected based on desirable characters and can be used as mother plants for the cloning process. Cloning methods are used to reproduce desirable strains that are genetically identical. Some *Cannabis* strains reportedly have stable seed genetics (discussed in Chapter II) and should have little variability observed in the phenotype of the offspring. In either case, the genotypes of clonal plants or plants from stable seed should be highly similar. Cloning in *Cannabis* is widely practiced in order to produce consistent products from the F1 generations and can produce hundreds of genetically identical plants. Therefore, breeders may not be invested in creating stable varieties because not only is cloning relatively easy, but also they are not provided protection for their novel varieties. This creates a situation where strains may be marketed with a specific name, but the genetics of plants with the same strain name but from a different source could be quite different.

Intellectual Property Protection for Plant Varieties

In the U.S., intellectual property and commercial exploitation protection for new plant cultivars is afforded under the Plant Variety Protection Act (PVPA) of 1970 (United States Department of Agriculture 1970). Plant growers and breeders can register proprietary varietals with the United States Department of Agriculture (United States Department of Agriculture 1970, 2015). In order to obtain a certificate for a novel variety from the Plant Variety Protection Office, the cultivar must be "new, distinct, uniform, and stable" (United States Department of Agriculture 1970, 2015). The new varietal must have a name that does not conflict with existing names of that crop. The distinctness of a novel variety may be based on one or more identifiable morphological, physiological, genetic, or other characteristics (e.g. baking characteristic for wheat) (United States

Department of Agriculture 1970, 2015). However, the ability to describe, distinguish and certify new cultivars in crop species does not apply to *Cannabis sativa*. The USDA lists ‘Hemp’ as an ineligible commodity (United States Department of Agriculture 1970; United States Department of Agriculture 2015) for protection under the PVPA, therefore, all *Cannabis* varietals are excluded from protection.

There are thousands of described *Cannabis* varieties, and probably thousands more that remain as “backyard” creations. For example, the online database Leafly describes more than 2500 strains, but it is far from a comprehensive list (Leafly 2018b). Moreover, many of these strains have not been stabilized and will only exist as long as there are healthy mothers to produce clones. Because the USDA Plant Variety Protection Office lists ‘Hemp’ (*Cannabis sativa*) as an ineligible commodity (United States Department of Agriculture 2015), there is no official or standardized database describing the different strains and the characteristics defining each strain. As a result, the cannabis industry has no way to verify varieties. Additionally, suppliers are not required to provide confirmation that the strain marked for sale as “Blue Dream”, for example, is in fact “Blue Dream”. The lack of a verification system for *Cannabis* strains is more than likely contributing to the high potential for misidentification and mislabeling. Consumers report that acquiring strains, such as “Blue Dream”, does not always result in the same effects each time (Prichard 2014), which reinforces the likelihood that *Cannabis* strain names are not a reliable identifier for plants and flower material at the present time.

***Cannabis* Consistency**

Public, scientific, and economic interest in *Cannabis* and *Cannabis* products is increasing worldwide. Consumers want to have confidence that the products they buy are

consistent. Consistency is an issue that researchers and industry are beginning to address. Teasing apart some of the elements that could contribute to variation is vital to the *Cannabis* industry, for both recreational and medical consumers. Determining variation within and among *Cannabis* types and how those genetic differences might be reflected in physical characteristics such as phytochemical constituents and aromatic profile, requires examining genetic differences. The phenotype of any organism is a product of genotype and environment. However, environmental variation such as growing conditions, harvesting time, soil, nutrient regimes and water levels are examples of confounding variables that could contribute to phenotypic differences observed in clonal organisms such as commercial *Cannabis*. Where variation is unexpected, examining the genotype to rule out genetic variation, rather than some other variable, is required. Unknown genetic differences leading to variation cannot be remedied by standardizing growing, harvesting, and storage procedures. However, if a grower has a certified and verified variety, any variation among plants can be narrowed down to differences in treatment following germination, and presumably be addressed. Occasional recreational users may not be concerned with variation in products; however, the growing number of medical marijuana patients who seek specific effects from their *Cannabis* need to be provided consistent products. Expecting one set of effects and experiencing another set of effects is unacceptable when it comes to medicine.

Cannabinoids and Terpenes

Cannabis sativa is a chemically complex plant with numerous natural constituents. To date, 565 constituents have been identified (ElSohly et al. 2016) that are classified as cannabinoids or non-cannabinoids (alkaloids, flavonoids, terpenoids, amino

acids and others). Varying levels and combinations of the chemical constituents results in unique chemical profiles, referred to as a chemotype. Constituents include 120 phytocannabinoids, which are a group of C21 terpenophenolic molecules with a ring structure derived from geranyl pyrophosphate (ElSohly et al. 2017). The main psychoactive cannabinoid, and the main reason for *Cannabis* prohibition, is Δ^9-tetrahydrocannabinol (THC). Besides being a psychoactive substance, THC has other known effects, such as analgesic properties (Rahn and Hohmann 2009). Cannabidiol (CBD), the second most abundant cannabinoid, has recently received attention as an antiepileptic and is particularly promising for intractable pediatric epilepsy (United States Food and Drug Administration 2018). However, the bioactive cannabinoids THC and CBD are not produced by the plant. Rather, the acidic forms Δ^9 -tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) are produced in the plant and are converted to their active forms through other mechanisms, such as the addition of heat. Other cannabinoids gaining popularity for various reported effects include: tetrahydrocannabivarin (THCV), which decreases appetite and increases metabolism (Halford and Harrold 2008); cannabichromene (CBC) which has anti-inflammatory, anti-bacterial, and anti-fungal properties (ElSohly et al. 1982); cannabigerol (CBG), which has anti-bacterial properties and reduces blood pressure (Banerjee et al. 1975); and cannabinol (CBN), which is an analgesic (Zygmunt et al. 2002), appetite stimulant (Farrimond et al. 2012), and an effective but mild sedative (Musty et al. 1976). Aromatic terpenes produced in the glandular trichomes of the female flower are a second important group of chemical constituents abundantly produced in *Cannabis*. Terpenes are manufactured in varying combinations and levels and produce distinctive characteristic

odors. *Cannabis* strain aroma descriptions include skunk, diesel, fruit, and cheese (see Chapter IV). Aromatic profiles give rise to descriptive strain names such as “Island Sweet Skunk”, “Sour Diesel”, “Banana Kush”, and “Blue Cheese”.

Cannabis types, such as hemp and drug types, are often defined by the level of THCA (but reported as THC) produced in the plant, validated methods to measure relative amounts of cannabinoids, as well as terpenes, have been developed. Gas chromatography (GC) is widely used for detecting the major cannabinoids because it is simple, fast and sensitive. However, GC cannot distinguish acidic cannabinoids from their decarboxylated forms unless a derivatization is performed. Gas chromatography uses high temperature and will decarboxylate the natural acidic forms of several cannabinoids such as Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) (de Oliveira et al. 2008; ElSohly et al. 2016; Hazekamp et al. 2005; Hillig 2005; Pellegrini et al. 2005). Several validated GC cannabinoid and terpene separation methods are available (ElSohly et al. 2017; Mariotti et al. 2016; Raharjo and Verpoorte 2004). High-performance liquid chromatography (HPLC) is another method to detect *Cannabis* chemical components that does not heat the sample and therefore the natural acidic and neutral cannabinoids are unaffected. The limitation of HPLC is that it may not resolve the full array of cannabinoids due to the complex composition of the plant extracts. Several validated HPLC methods for separation of cannabinoids and terpenes are available (Aizpurua-Olaizola et al. 2014; Brighenti et al. 2017; De Backer et al. 2009; Giese et al. 2015; Gul et al. 2015; Rustichelli et al. 1996; Swift et al. 2013; Xiaoyan et al. 2016).

History of Prohibition

Cannabis has been used for centuries and is notorious for the psychoactive properties produced when THCA is activated through heating, for example when it is smoked. The Single Convention on Narcotic Drugs of 1961 is an international treaty listing specific drugs prohibited worldwide with the exception of medical and research purposes (The United Nations 1961). In this treaty, *Cannabis* drug types, such as hashish and marijuana, are listed as both Schedule I and IV drugs (The United Nations 1961). Schedule I substances have the strictest controls and Schedule IV substances are described as having "particularly dangerous properties" (The United Nations 1961). The most dangerous drugs (includes opium, opioids, coca, cocaine, heroin, and fentanyl), are listed as Schedule IV and are described as having extremely limited and therapeutic value and are thereby subject to the strictest controls under Schedule I (The United Nations 1961). The treaty explicitly states that *Cannabis* used for hemp and fiber is not considered controlled substances, but rather only the fruiting crowns of the plant and derived products are included. The Single Convention requires countries in the treaty to establish a government agency to control cultivation of scheduled drugs from plants such as *Cannabis* and opium. In the U.S. the National Institute on Drug Abuse (NIDA) is that agency. However, since the treaty was written, the discovery of the endocannabinoid system suggests there are therapeutic applications, and mounting evidence suggests a wide variety of medical applications for *Cannabis.*

The Single Convention united countries worldwide in the prohibition of a multitude of substances, but *Cannabis* bans had been introduced long before 1961. One of the first bans was by the Emir of the Joneima in Arabia in 1378 who declared

Cannabis ingestion was punishable by removing all the offender's teeth (Johnson et al. 2010). King Andrianampoinimerina of Madagascar imposed capital punishment for *Cannabis* use in 1787 (Yates 2015). Napoleon banned *Cannabis* use and distribution in 1800 (Booth 2004). Singapore banned *Cannabis* in 1870 (De Padua et al. 1999). Greece banned cultivating, importing, and use in 1890 (Abel 2013). The Ganja Law supported by the Council of Evangelical Churches outlawed *Cannabis* in Jamaica in 1913 (Moyston 2013). Australia banned *Cannabis* in 1926 (Wodak and Owens 1996).

The United States participated in The Single Convention, but already had existing legislation in the form of the Marihuana Tax Act of 1937 (75th Congress of the United States 1937). The Act did not ban *Cannabis* outright, but rather placed a hefty sales tax on *Cannabis* and *Cannabis* products, including hemp, thereby effectively making it difficult to engage in *Cannabis* based business. *Cannabis* was formally criminalized in the United States under the Controlled Substances Act, passed by the 91st United States Congress, and signed into Law by President Richard Nixon (United States Congress 1970).

Current Status

The United States has seen significant changes over the past two decades regarding the legal status of *Cannabis* use for both medical and recreational purposes. State-level legislation has side-stepping the federal *Cannabis* ban, making allowances for medical *Cannabis* use. While *Cannabis* remains federally illegal, national enforcement of *Cannabis* restrictions has shifted to the responsibility of states. This disconnect between federal and state laws creates logistical issues for federally regulated organizations, such as research and financial institutions, as well as creating confusion for federal law

enforcement agencies such as the United States Marshals Service (USMS), the Federal Bureau of Investigation (FBI), and the Drug Enforcement Administration (DEA). Moreover, it is impossible to visually distinguish different *Cannabis* types. Now that hemp is legal in all 50 states (United States Department of Agriculture and 113th United States Congress 2018), law enforcement has no way to tell if someone is hauling 17,000 lbs. of hemp flower for CBD, or if they are smuggling 17,000 lbs. of illicit drugs across state lines (Konopasek 2019). Some states allow *Cannabis* treatment in the cases of serious or debilitating conditions (ProCon 2016b). Other states are more lenient with medical conditions that may appropriately be treated with *Cannabis* and allow medical doctors and authorized healthcare professionals to recommended patients use medical marijuana as treatment (National Conference of State Legislatures 2018). The initiation of statewide changes began with California passing Proposition 215 in 1996, which legalized medical marijuana (National Conference of State Legislatures 2018). From 1996 to 2012, sixteen states and the District of Columbia passed legislation allowing medical marijuana use (Table 1.1). In 2012, Colorado was the first state to legalize recreational marijuana use, followed by Washington later that year. Since then, ten more states have legalized medical marijuana, and eight states plus the District of Columbia legalized recreational use (National Conference of State Legislatures 2018). All in all, the current standing in the U.S. is 33 states allow medical marijuana, and of those ten and the District of Columbia, also allow recreational use (Table 1.1) (ProCon 2018b).

While the United States continues to relax legislation, other countries are following suit (Table 1.1). Over the last two decades, a wave of decriminalization, re-classification, and legalization of *Cannabis* has surged worldwide. Paraguay was one of

the first countries to decriminalize possession of small amounts (10 grams) of *Cannabis* in 1988 (Mostyn et al. 2012). Uruguay was the first country in this modern era to fully legalize *Cannabis*, although the legal purchase of *Cannabis* is limited to registered Uruguayan citizens (Gerner 2015). Canada legalized medical use in 2001, and recently became the second country to legalize recreational use and establish a nationwide marijuana market. There are at least fifteen countries worldwide that have legalized medical marijuana use, the majority of which have made changes in the last five years.

Table 1.1. Worldwide and U.S. Legalization History. Legalization year of medical and recreational *Cannabis* both worldwide and in the United States (by state).

Year	Medical		Recreational	
	Country	U.S. State	Country	U.S. State
1990	Israel			
1996		California		
1998		Alaska		
		Oregon Washington		
1999		Maine		
2000		Colorado Hawaii		
2001	Canada			
2004	Chile			
2007		Montana New Mexico Rhode Island Vermont		
2008		Michigan		
2009		New Jersey		
2010		Arizona District of Columbia		
2011		Delaware		
2012		Massachusetts Connecticut		Colorado Washington
2013	Czech Republic Uruguay	Illinois New Hampshire	Uruguay	
2014		Minnesota New York Utah Maryland		Alaska District of Columbia Oregon
2015	Columbia Croatia Italy Puerto Rico	Georgia Louisiana		
2016	Argentina Australia Macedonia Turkey	Florida North Dakota Ohio Pennsylvania Nevada Arkansas		California Maine Massachusetts Nevada
2017	Germany Mexico Philippines Poland Malta Peru Greece	West Virginia		
2018	Georgia New Zealand Luxembourg Portugal	Oklahoma Missouri	Canada Georgia South Africa	Vermont Michigan

Medical Applications

Research on potential medical applications for the treatment of a wide array of medical conditions is abundant and ongoing. However, the short and long-term health effects of *Cannabis* consumption remain largely unknown. The National Academies of Sciences, Engineering, and Medicine published a comprehensive review of medical *Cannabis* and conclude that the effects of *Cannabis* are understudied, and limitations need to be addressed and prioritized (Cousijn et al. 2018; National Academies of Sciences 2017).

The recent surge of legalization of medical marijuana in the U.S. and worldwide suggests there is enough evidence to support the claims that *Cannabis* is effective in treating certain medical conditions. The U.S. had an estimated 2.2 million registered medical marijuana patients in legal medical states in 2016 (Leafly 2018a), and there are a wide range of conditions for which *Cannabis* treatment is being investigated, including chronic pain, multiple sclerosis, cancer, and epilepsy. Chronic pain is persistent pain that results from injury, disease, or can be a disease in itself (The American Academy of Pain Medicine 2019). The American Academy of Pain Medicine estimates there are more than 100 million Americans suffering from chronic pain with associated costs > $600 billion annually (The American Academy of Pain Medicine 2019). Reviews of research on *Cannabis* treatment of chronic pain have found mixed results and suggest further large-scale clinical trials are necessary (Baron 2018; Hill 2015; Jensen et al. 2015). Recently, a large-scale clinical study in Israel administered four strains and reported 93.7% of the 2,736 elderly patients in the study reported significant pain reductions (Abuhasira et al. 2018). Cannabinoid agonists found in *Cannabis,* medicinal isolates such as Sativex®,

and synthetic cannabinoid receptor agonists such as nabilone, reduce pain by acting on CB1 and CB2 receptors located in the central and peripheral nervous systems (Costa et al. 2007; Guindon and Hohmann 2008; Pertwee 2001, 2005, 2009). Studies conducted in the United Kingdom found convincing evidence that Sativex® is an effective treatment for tremors and spasticity (Alexander 2016) associated with multiple sclerosis, for example. Sativex® was also shown to be effective in reducing symptoms related to chemotherapy and was well tolerated by patients and resulted in minimal adverse side effects (Duran et al. 2010). Anti-tumoral actions are associated with several cannabinoids including THC, but CBD has been found to be the most effective cannabinoid in reducing tumor cell growth (Ligresti et al. 2006). Additionally, CBD induces apoptosis in human myleoblastic cells but has no effect on healthy mononuclear cells (Gallily et al. 2003; McKallip et al. 2006; Vaccani et al. 2005). The U.S. Food and Drug Administration (FDA) has not approved *Cannabis* to treat medical conditions but has approved three cannabinoids for medical use (United States Food and Drug Administration 2018; United States Department of Health and Human Services et al. 2018). These are Epidiolex® (CBD) for the treatment of two rare forms of epilepsy, dronabinol (synthetic THC) for nausea and weight loss associated with cancer and AIDS, and nabilone (synthetic THC) for nausea associated with cancer treatments (United States Food and Drug Administration 2018; United States Department of Health and Human Services et al. 2018).

Genetic Research

Genetic research on *Cannabis* is complicated because it is primarily dioecious, highly heterozygous, considerably variable, and extraordinarily plastic in response to

varying environmental conditions (Onofri and Mandolino 2017). Genetic studies on *Cannabis* have focused on evolutionary history (Booth 2004; Clarke and Merlin 2016; Hillig 2005; Russo 2007; Small et al. 1976; Sytsma et al. 2002; Yang et al. 2013), speciation (Emboden 1981; Hillig 2005; McPartland and Guy 2017; Sawler et al. 2015), geographic origins (Alghanim and Almirall 2003; Coyle et al. 2003; Gao et al. 2014; Gilmore et al. 2003), distribution (Hillig 2005; Piluzza et al. 2013; Piomelli and Russo 2016), identification of sex chromosomes (Faux et al. 2014; Faux et al. 2016; Mandolino et al. 1999; Moliterni et al. 2004; Peil et al. 2003; Razumova et al. 2016; Techen et al. 2010), genetic contribution to the variation of chemotypes among varietals (Aizpurua-Olaizola et al. 2016; de Meijer et al. 2009a; de Meijer et al. 2009b; Desjardins 2008; Pacifico et al. 2006; Staginnus et al. 2014; Welling et al. 2016), and analyses to aid law enforcement and forensic investigations (Dufresnes et al. 2017; Houston et al. 2016; Kojoma et al. 2006; Onofri and Mandolino 2017).

Studies with accessions of both hemp and drug types have clearly and consistently shown genetic distinction between the two types using clustering analyses such as PCA, UPGMA, and STRUCTURE (Datwyler and Weiblen 2006; Gilmore and Peakall 2003; Gilmore et al. 2003; Grassa et al. 2018; Hillig 2005; Kojoma et al. 2006; Lynch et al. 2016; Pacifico et al. 2006; Sawler et al. 2015). Genetic evidence using traditional genetic techniques such as allozymes, Random Amplification of Polymorphic DNA (RAPD) and Short Tandem Repeat (STR) markers give little support to the differentiation of Sativa narrow-leaf drug type and the Indica broad-leaf drug type (Hillig 2005; Knight et al. 2010; Lynch et al. 2016; Sawler et al. 2015). However, new genetic tools using next-generation sequencing (NGS) techniques such as single nucleotide polymorphisms

(SNPs), whole genome shotgun sequencing (WGS), and Restriction site Associated DNA Sequencing (RAD-Seq) create datasets for large portions of the genome compared to previous techniques. These tools may be able to better distinguish the genetic difference driving the reported Sativa and Indica phenotypic differences. Recent studies using NGS have had more success distinguishing the two types (Henry 2015; Lynch et al. 2016; Sawler et al. 2015). With the information gathered thus far, it appears that the Sativa and Indica types shared a common ancestor that likely diverged via natural selection when populations established in regions of India (warm, low, wet) and Afghanistan (cool, high, dry) leading to phenotypic and genotypic differences (McPartland 2017). However, the human relationship with *Cannabis,* including cultivation, breeding and selection, has blurred the line differentiating what were presumably two distinct species (McPartland 2017; McPartland and Guy 2017).

Entire genomic sequences allow researchers to not only explore relationships among different *Cannabis* types, but also uncover information about genes controlling characters of interest, such as cannabinoid and terpene synthesis, as well as flowering time, and flower production. The nuclear (van Bakel et al. 2011; Vergara et al. 2016), chloroplast (Oh et al. 2015; Vergara et al. 2015), and mitochondrial (White et al. 2016) genomes, as well as transcriptomes for “Purple Kush” (drug type) and “Finola” (hemp type) (van Bakel et al. 2011) have been published. Despite the full sequencing of the nuclear genome, it is complex and highly repetitive, and has yet to be assembled in entirety. However, there are several researchers focused on the complete assembly and annotation of the entire genome. Of great interest are the genes responsible for synthesizing cannabinoids (Grassa et al. 2018; Laverty et al. 2019). Recently the genes

responsible for THCA and CBDA production were found to be located on chromosome 9, and analyses suggest these genes have been targets for selection by breeding for drug type strains (Grassa et al. 2018). However, the genes responsible for controlling the relative abundance of cannabinoids have not yet been discovered (Grassa et al. 2018).

Summary

Historical criminalization of *Cannabis* has severely hindered scientific research on this prominent plant. Research on *Cannabis'* phytochemicals including hundreds of cannabinoids and terpenes is growing, but the details of the genetic contribution to the abundance and combination of these compounds is in its research infancy. Relatively few genetic studies have been conducted and the origins and genetic identities of most *Cannabis* varieties are largely unknown. Additionally, there are few *Cannabis* studies researching genetic and chemical aspects together. While a lack of research on such an economically important plant is problematic for the *Cannabis* industry, it can be argued that a larger problem is the lack of a regulation and verification system to accurately identify or verify the thousands of strains that have been described. Chemical constituents of strains are the dominant focus of the *Cannabis* industry (Hillig and Mahlberg 2004; Pacifico et al. 2006; Fischedick et al. 2010; Hazekamp and Fischedick 2012; Elzinga et al. 2015; Aizpurua-Olaizola et al. 2016). However, the cannabinoid and terpenes present in the plant, as well as the levels of gene expression can vary widely. *Cannabis* chemotypes are variable (plastic) and are therefore unreliable to identify strains. Research on plastic traits such as chemical constituents needs to be juxtaposed with genetic data in order to more accurately describe and verify *Cannabis* strains.

The research herein addresses genetic variation in *Cannabis* and how genetic variation relates to phenotype. This dissertation includes: (1) a genetic investigation to determine if strains with the same name purchased from *Cannabis* dispensaries are genetically similar (Chapter II), (2) an examination of the genetic relationship among various types of *Cannabis* (Chapter III), (3) an investigation of human perception of aromas in four *Cannabis* strains and if genetic anomalies are detectable through olfaction (Chapter IV), and (4) an examination of cannabinoids and terpenes to determine if chemical profiles of four *Cannabis* strains are similar within strains and if genetic anomalies are reflected in different cannabinoid levels (Chapter V). Taken together, these studies aim to provide valuable information about sources of variation in the recently revived and globally expanding *Cannabis* industry.

CHAPTER II

GENETIC VARIATION PART 1: GENETIC TOOLS WEED OUT MISCONCEPTIONS OF STRAIN RELIABILITY IN *CANNABIS SATIVA:* IMPLICATIONS FOR A BUDDING INDUSTRY

Contributions of Authors and Co-Authors

Manuscript in Chapter II

Author: Anna L. Schwabe

Contributions: Conceived the project, collected samples, developed the microsatellite markers, conducted lab work, data analysis and manuscript preparation

Author: Mitchell E. McGlaughlin

Contributions: Directed the project, contributed some funding, and provided guidance throughout the project and manuscript preparation

Abstract

Unlike other plants, *Cannabis sativa* is excluded from regulation by the United States Department of Agriculture (USDA). Distinctive *Cannabis* varieties are ostracized from registration and therefore nearly impossible to verify. As *Cannabis* has become legal for medical and recreational consumption in many states, consumers have been exposed to a wave of novel *Cannabis* products with many distinctive names. Despite more than 2000 named strains being available to consumers, questions about the consistency of commercially available strains have not been investigated through scientific methodologies. As *Cannabis* legalization and consumption increases, the need to provide consumers with consistent products becomes more pressing. In this research, we examined commercially available, drug-type *Cannabis* strains using genetic methods to determine if the commonly referenced distinctions are supported and if samples with the same strain name are consistent when obtained from different facilities. We developed ten *de-novo* microsatellite markers using the "Purple Kush" genome to investigate potential genetic variation within 30 strains obtained from dispensaries in three states. Samples were examined to determine if there is any genetic distinction separating the commonly referenced Sativa, Indica and Hybrid types and if there is consistent genetic identity found within strain accessions obtained from different facilities. Although there was strong statistical support dividing the samples into two genetic groups, the groups did not correspond to commonly reported Sativa/Hybrid/Indica types. The analyses revealed genetic inconsistencies within strains, with most strains containing at least one genetic outlier. However, after the removal of obvious outliers, many strains showed considerable genetic stability. We failed to find

clear genetic support for common strain descriptions of Sativa, Indica and Hybrid types as described in online databases. Significant genetic differences within samples of the same strain were observed indicating that consumers could be provided inconsistent products. These differences have the potential to lead to phenotypic differences and unexpected effects, which could be surprising for the recreational user, but have more serious implications for patients relying on strains that alleviate specific medical symptoms.

Introduction

Cultivation of *Cannabis sativa* L. dates back thousands of years (Abel 2013) but has been largely illegal worldwide for the best part of the last century. The U.S. Drug Enforcement Agency considers *Cannabis* a Schedule I drug with no "accepted medical use in treatment in the United States" (United States Congress 1970), but laws allowing *Cannabis* for use as hemp, medicine, and some adult recreational use are emerging (ProCon 2018b). Global restrictions have limited *Cannabis* related research, and there are relatively few genetic studies focused on strains (Lynch et al. 2016; Soler et al. 2017), but studies with multiple accessions of a particular strain show variation (Lynch et al. 2016; Sawler et al. 2015; Soler et al. 2017).

Currently, the *Cannabis* industry has no way to verify strains. Consequently, suppliers are unable to provide confirmation of strains, and consumers have to trust the printed name on a label matches the product inside the package. Reports of inconsistencies, along with the history of underground trading and growing in the absence of a verification system, reinforce the likelihood that strain names may be unreliable identifiers for *Cannabis* products at the present time. Without verification systems in place, there is the potential for misidentification and mislabeling of plants, creating names for plants of unknown origin, and even re-naming or re-labeling plants with prominent names for better sale. *Cannabis* taxonomy is complex (Clarke and Merlin 2013, 2015, 2016; Emboden 1974; Hillig 2005; Russo 2007; Schultes et al. 1974; Small 2015b; Small et al. 1976), but given the success of using genetic markers, such as microsatellites, to determine varieties in other crops, we suggest that similar genetic

based approaches should be used to identify *Cannabis* strains in medical and recreational marketplaces.

There are an estimated ~3.5 million medical marijuana patients in the United States (U.S.) (Leafly 2018a) and various levels of recent legalization in many states has led to a surge of new strains (Leafly 2018b; Wikileaf 2018). Breeders are producing new *Cannabis* strains with novel chemical profiles resulting in various psychotropic effects and relief for an array of symptoms associated with medical conditions including (but not limited to): glaucoma (Tomida et al. 2004), Chron's Disease (Naftali et al. 2013), epilepsy (United States Food and Drug Administration 2018; Maa and Figi 2014), chronic pain, depression, anxiety, PTSD, autism, and fibromyalgia (Borgelt et al. 2013; Cousijn et al. 2018; Naftali et al. 2013; Ogborne et al. 2000; ProCon 2016a).

There are primarily two *Cannabis* usage groups, which are well supported by genetic analyses (Dufresnes et al. 2017; Lynch et al. 2016; Sawler et al. 2015; Soler et al. 2017): ***hemp*** defined by a limit of $< 0.3\%$ Δ^9-tetrahydrocannabinol (THC) in the U.S., and ***marijuana*** or ***drug-types*** with moderate to high THC concentrations (always $> 0.3\%$ THC). Within the two major groups *Cannabis* can be further divided into strains (varietals), and particularly for the drug types, strains are assigned to one of three categories: ***Sativa*** which reportedly has uplifting and more psychotropic effects, ***Indica*** which reportedly has more relaxing and sedative effects, and ***Hybrid*** which is the result of breeding Sativa and Indica types resulting in intermediate effects. The colloquial terms **Sativa, Hybrid**, and **Indica** are used throughout this document even though these terms do not align with the current formal botanical taxonomy for *Cannabis sativa* and proposed *Cannabis indica* (McPartland 2017; Piomelli and Russo 2016). We feel the

colloquial terminology is 8necessary here as the approach for this study was from a consumer view, and these are the terms offered as common descriptors for the general public (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018; Marijuana strains database 2019). Genetic analyses have not provided a clear consensus for higher taxonomic distinction among these commonly described *Cannabis* types (Lynch et al. 2016; Sawler et al. 2015), and whether there is a verifiable difference between Sativa and Indica type strains is debated. However, both the recreational and medical *Cannabis* communities claim there are distinct differences in effects between Sativa and Indica type strains (Smith 2012; Leaf Science 2016; Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018; Marijuana strains database 2019).

Female *Cannabis* plants are selected based on desirable characters (mother plants) and are produced through cloning and, in some cases, self-fertilization to produce seeds (Green 2005). Cloning allows *Cannabis* growers to replicate plants, ideally producing consistent products. There are an overwhelming number of *Cannabis* strains that vary widely in appearance, taste, smell and psychotropic effects (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018; Marijuana strains database 2019). Online databases such as Leafly (Leafly 2018b) and Wikileaf (Wikileaf 2018), for example, provide consumers with information about strains but lack scientific merit for the *Cannabis* industry to regulate the consistency of strains. Other databases exist (NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Marijuana strains database 2019), but the method of assignment to the three groups is often undisclosed, confounded, or mysterious. Wikileaf reports a numeric percentage of assignment to Sativa and/or Indica

(Wikileaf 2018), which is why we chose it as our scale reference scale of ancestry, although there is some disagreement among online sources (Table 2.1). To our knowledge, there have not been any published scientific studies specifically investigating the genetic consistency of strains at multiple points of sale for *Cannabis* consumers.

Breeders and growers choose *Cannabis* plants with desirable characters (phenotype) related to flowers, cannabinoid profile, and terpene production. Phenotype is a product of genotype and environment. *Cannabis* is considerably variable and extraordinarily plastic in response to varying environmental conditions (Onofri and Mandolino 2017). Therefore, determining sources of variation, at the most basic level, requires examining genetic differences. Strains propagated through cloning should have minimal genetic variation. Eight of the strains examined in this study are reportedly clone only strains indicating there should be little to no genetic variation within these strains. That being said, it is possible for mutations to accumulate over multiple generations of cloning (Gabriel et al. 1993; Hojsgaard and Horandl 2015), but these should not be widespread. Self-fertilization and subsequent seed production may also be used to grow a particular strain. With most commercial plant products growers go through multiple generations of self-fertilization and backcrossing to remove genetic variability within a strain and provide a consistent product (Riggs 1988). However, for many *Cannabis* strains, the extent of genetic variability stabilization is uncertain. It has been observed that novel *Cannabis* strains developed through crossing are often phenotypically variable (Green 2005), which could be the result of seed producers growing seeds that are not stabilized enough to produce a consistent phenotype. Soler et al. (Soler et al. 2017)

examined the genetic diversity and structure of *Cannabis* cultivars grown from seed and found considerable variation, suggesting that seed lots are not consistent.

Table 2.1. Twelve popular strains and their described assignment of Sativa and Indica according to six online databases of *Cannabis* strain information (Leafly 2018b).

Strain	% Sativa	Leafly	Strainfinder	NCSM	PotGuide	Seedfinder
Durban Poison*	100	Pure sativa	Sativa dominant	Pure Sativa	Sativa	7 breeders: Pure Sativa (4), Mostly Sativa (2), Sativa/Indica (1)
Sour Diesel*	90	Sativa dominant	Sativa dominant (70%)	Mostly Sativa	Sativa Dominant Hybrid	19 breeders: Mostly Sativa (12), Sativa/Indica (5), mostly Indica (2)
Golden Goat*[v]	65	Sativa dominant hybrid	Not Found	Sativa Dominant	Sativa Dominant Hybrid	3 breeders: Mostly Sativa (2), Sativa/Indica (1)
Bruce Banner*	60	Sativa effects	*Sativa dominant (65%)	Sativa Dominant (60%)	*Sativa Dominant Hybrid	10 breeders: Mostly Sativa (6), Sativa/Indica (1), mostly Indica (3)
Flo*	60	Hybrid	Sativa dominant (60%)	Not Found	Not Found	1 breeder: Mostly Sativa
Pineapple Express*	60	Sativa dominant hybrid	Indica dominant (70%)	Sativa Dominant (60%)	Not Found	6 breeders: Mostly Sativa (2), Sativa/Indica (1), mostly Indica (1), Ruderalis/Sativa/Indica (2)
OG Kush*[v]	55	Hybrid	Indica dominant (75%)	Not Found	Sativa Dominant Hybrid	25 breeders: Pure Indica (3), mostly Sativa (1), mostly Indica (15), Sativa/Indica (6)
Blue Dream*[v]	50	Sativa dominant hybrid	Sativa dominant (70%)	Sativa Dominant (60%)	Sativa Dominant Hybrid	10 breeders: Mostly Sativa (9), mostly Indica (1)
Chemdawg*	45	Hybrid	Indica dominant (60%)	Indica dominant (55%)	Sativa Dominant Hybrid	2 breeders: Mostly Sativa (2) (Chem Dog and Chemdawg)
Banana Kush*	40	Hybrid	Not Found	Not Found	Indica Dominant Hybrid	3 breeders: Mostly Indica (1), Sativa/Indica (2)
Girl Scout Cookies*[v]	40	Hybrid	Indica dominant (60%)	Indica dominant (60%)	Hybrid	17 breeders: Mostly Sativa (2), Sativa/Indica (7), mostly Indica (7), Ruderalis/Sativa/Indica (1)
Purple Kush*[v]	0	Pure Indica	Indica dominant (75%)	Pure Indica	Indica Dominant Hybrid	5 breeders: Mostly Indica (5)

Given the uncertainties surrounding named *Cannabis* strains, genetic data provide an ideal path to examine how widespread genetic inconsistencies might be.

In the U.S., protection against commercial exploitation, trademarking, and recognition of intellectual property for developers of new plant cultivars is provided through the United States Department of Agriculture (USDA) and The Plant Variety Protection Act of 1970 (United States Department of Agriculture 1970). Traditionally, morphological characters were used to define new varieties in crops such as grapes (*Vitis vinifera* L.), olives (*Olea europea* L.) and apples (*Malus domestica* Borkh.). With the rapid development of new varieties in these types of crops, morphological characters have become increasingly difficult to distinguish. Currently, quantitative and/or molecular characters are often used to demonstrate uniqueness among varieties. Microsatellite genotyping enables growers and breeders of new cultivars to demonstrate uniqueness through variable genetic profiles (Rongwen et al. 1995). Microsatellite genotyping has been used to distinguish cultivars and hybrid varieties of multiple crop varietals within species (Baldoni et al. 2009; Belaj et al. 2004; Cipriani et al. 2002; Costantini et al. 2005; Guilford et al. 1997; Hokanson et al. 1998; Muzzalupo et al. 2009; Pellerone et al. 2001; Poljuha et al. 2008; Rongwen et al. 1995; Sarri et al. 2006; Stajner et al. 2011). Generally, 3-12 microsatellite loci are sufficient to accurately identify varietals and detect misidentified individuals (Baldoni et al. 2009; Belaj et al. 2004; Cipriani et al. 2002; Muzzalupo et al. 2009; Poljuha et al. 2008; Sarri et al. 2006). *Cannabis* varieties however, are not afforded any legal protections, as the USDA considers it an “ineligible commodity” (United States Department of Agriculture 2014)

but genetic variety identification systems provide a model by which *Cannabis* strains could be developed, identified, registered, and protected.

We used a well-established genetic technique to compare commercially available *C. sativa* strains to determine if products with the same name purchased from different sources have genetic congruence. This study is highly unique in that we approached sample acquisition as a common retail consumer by purchasing flower samples from dispensaries based on what was available at the time of purchase. All strains were purchased as-is, with no additional information provided by the facility, other than the identifying label. This study aimed to determine if: (1) any genetic distinction separates the common perception of Sativa, Indica and Hybrid types; (2) consistent genetic identity is found within a variety of different strain accessions obtained from different facilities; (3) there is evidence of misidentification or mislabeling.

Methods

Genetic Material

Cannabis samples for 30 strains were acquired from 20 dispensaries or donors in three states (Table 2.2). All samples used in this study were obtained legally from either retail (Colorado and Washington), medical (California) dispensaries, or as a donation from legally obtained samples (Greeley 1). DNA was extracted using a modified CTAB extraction protocol (Doyle 1987) with 0.035-0.100 grams of dried flower tissue per extraction. Although several databases exist with various descriptive Sativa and Indica assignments for thousands of strains (Table 2.1 & 2.2), proportions of Sativa and Indica phenotypes from Wikileaf (Wikileaf 2018) were used for this study. Analyses were performed on the full 122-sample data set (Table 2.2). The 30 strains were assigned a

proportion of Sativa according to online information (Table 2.2). Twelve of the 30 strains were designated as ‘popular’ due to higher availability among the dispensaries as well as online information reporting the most popular strains (Table 2.3) (Escondido 2014; Rahn 2016a; Rahn 2016b; Rahn et al. 2016). Results from popular strains are highlighted to show levels of variation in strains that are more widely available or that are in higher demand.

Table 2.2. *Cannabis* samples (122) from 30 strains. Reported proportion of Sativa from Wikileaf (Wikileaf 2018) and the city location and state where each sample was acquired are included. (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
Pineapple Express	60	Boulder 1	CO	Hash Plant (Australian)	20	Garden City 1	CO

Table 2.2. *continued*

Name	Sativa	City	State	Name	Sativa	City	State
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				

* Strain proportion of "Chemdawg" variants not listed on Wikileaf

● Strain proportion of "Tangerine" not listed on Wikileaf; proportion listed is of "Tangerine Dream"

Table 2.3. Summary of *Cannabis* samples (122) from 30 strains. The reported proportion of Sativa retrieved from Wikileaf (Wikileaf 2018). Abbreviations used for Lynch & Ritland (Lynch and Ritland 1999) relatedness statistics are included, and the proportions of membership for genotype 1 and genotype 2 from the STRUCTURE (Fig. 2.1) expressed as a percentage.

Strain	Abbr	# Samples	Sativa %	Genotype 1 (% average)	Genotype 2 (% average)	Standard Deviation
Durban Poison*	DuPo	9	100	86	14	9.9
Hawaiian	Hawa	2	90	61	39	27.58
Sour Diesel*	SoDi	7	90	14	86	53.74
Trainwreck	TrWr	2	90	59	41	21.92
Island Sweet Skunk	ISS	3	80	93	7	9.19
AK-47	AK47	3	65	55	45	7.07
Golden Goat*❖	GoGo	7	65	68	32	2.12
Green Crack❖	GrCr	3	65	60	40	3.54
Bruce Banner*	BrBa	6	60	19	81	28.99
Flo*	Flo	4	60	38	62	15.56
Jillybean	JiBe	3	60	73	27	9.19
Pineapple Express*	PiEx	5	60	62	38	1.41
Purple Haze	PuHa	3	60	77	23	12.02
Tangerine	Tang	2	60	53	47	4.95
Jack Herer	JaHe	3	55	66	34	7.78
OG Kush*❖	OGKu	4	55	28	72	19.09
Blue Dream*❖	BlDr	9	50	80	20	21.21
Tahoe OG	TaOG	4	50	26	74	16.97
Chemdawg*	ChDa	7	45	9	91	25.46
Headband	HeBa	2	45	57	43	8.49
Banana Kush*	BaKu	4	40	52	48	8.49
Girl Scout Cookies*❖	GSC	8	40	25	75	10.61
Jack Flash	JaFl	2	40	96	4	39.6
Larry OG	LaOG	3	40	7	93	23.33
G-13	G13	3	30	50	50	14.14
Lemon Diesel❖	LeDi	2	30	85	15	38.89
Hash Plant	HaPl	4	20	37	63	12.02
Pre98-Bubba Kush	PBK	2	15	7	93	5.66
Grape Ape	GrAp	2	0	55	45	38.89
Purple Kush*❖	PuKu	4	0	29	71	20.51

* Twelve popular strains
❖ Clone only strains (SeedFinder 2018)

Microsatellite Development

The *Cannabis* draft genome from “Purple Kush” (GenBank accession AGQN00000000.1) was scanned for microsatellite repeat regions using MSATCOMMANDER-1.0.8-beta (Faircloth 2008). Primers were developed *de-novo* flanking microsatellites with 3-6 nucleotide repeat units (Table 2.4). Seven of the microsatellites were trinucleotide motifs with >10 repeating units. There was two hexanucleotide motifs with 20 and 30 repeating units. Finally, one tetranucleotide motif with 10 repeating units was included (Table 2.4). One primer in each pair was tagged with a 5’ universal sequence (M13 or T7) so that a matching sequence with a fluorochrome tag could be incorporated via Polymerase Chain Reaction (PCR) (Schwabe et al. 2015). Ten primer pairs produced consistent peaks within the predicted size range and were used for the genetic analyses herein (Table 2.4).

Table 2.4. Primer information. Includes the multiplex assignment, primer name, microsatellite repeat and number of units repeated in the "Purple Kush" draft genome (National Center for Biotechnology Information, accession AGQN00000000.1), forward and reverse sequences (asterisk denotes the sequence to which the tag is attached), the universal tag (sequence revealed at the bottom of the table), dye (VIC, FAM, PET), optimized annealing temperature, MgCl uL volume, amplified fragment size range, and the number of alleles in the data set.

Multiplex	Primer	Repeat Unit	Forward Sequence	Reverse Sequence	Tag	Dye	Anneal Temp	Magnesium	Fragment Size	Na
1	Casa_002	(GGAATT)$_{20}$	GTTAGACAATGCTGCCGGTG	*TTCCGATCCAATCCGCAC	M13	FAM	57.4	MgCl 2uL	270-324	8
1	Casa_022	(AGAT)$_{10}$	TCCACAGCCAGAGGAGAATC	*GGATCATTGGACAGCCATTC	T7	VIC	63.0	MgCl 2uL	190-208	6
1	Casa_027	(GTT)$_{20}$	CATCTCCCAGCCCTTTCATA	*GCTAGGGTTTTTGCCAAC	M13	FAM	57.4	MgCl 2uL	184-196	9
1	Casa_028	(AAT)$_{13}$	*TGCACATTGCTCTCCTTTTG	GAATGTGGTCCAATAAACACTCC	M13	PET	55.1	MgCl 2uL	173-190	8
1	Casa_030	(CAA)$_{19}$	CAATCCACACAACAGCTCCT	*TGCAGCAAGTTTAGGTGGTC	M13	VIC	55.1	MgCl 6uL	271-300	8
2	Casa_006	(TTTCTC)$_{30}$	*TTCTTCTCTCGACAGAACCC	TAGAACCAAGCAAGAAGGGC	M13	FAM	55.1	MgCl 1uL	410-422	6
2	Casa_014	(TAG)$_{13}$	*ATCGTGTTGCATGTTTGTGG	TGTGCTCCCTCTTGTATGATTC	M13	FAM	63.0	MgCl 2uL	270-290	8
2	Casa_018	(ATT)$_{28}$	*TCATAACCCCAAAAGCAAAG	GGGTAAATATAGCTGGCAAAGC	T7	VIC	55.1	MgCl 3uL	182-221	10
2	Casa_026	(CTT)$_{13}$	*CCATTTCGACCCTTGTAGGT	CTGGGGAAGATGAACGAAAG	M13	FAM	57.4	MgCl 1uL	201-206	9
2	Casa_029	(ACC)$_{11}$	CCCTCTCAGTCCCAAATTCA	*GATGGTGATGAGGAGGAGGA	M13	PET	55.1	MgCl 2uL	183-192	5

M13: AGGAAACAGCTATGACCAT

T7: GCTAGTTATTGCTCAGCGG

Polymerase Chain Reaction and Data Scoring

Microsatellite loci (Table 2.4) were amplified in 12 µL reactions using 1.0 µL DNA (10-20 ng/ µL), 0.6 µL fluorescent tag (5 µM; FAM, VIC, or PET), 0.6 µL non-tagged primer (5 µM), 0.6 µL tagged primer (0.5 µM), 0.7 µL dNTP mix (2.5mM), 2.4 µL GoTaq Flexi Buffer (Promega, Madison, WI, USA), 0.06 µL GoFlexi taq polymerase (Promega), 0.06 µL BSA (Bovine Serum Albumin 100X), 0.5 - 6.0 µL MgCl or $MgSO_4$, and 0.48 - 4.98 µL dH_2O. An initial 5-minute denaturing step was followed by thirty-five amplification cycles with a 1-minute denaturing at 95º C, 1-minute annealing at primer-specific temperatures and 1-minute extension at 72ºC. Two multiplexes (Table 2.4) based on fragment size and fluorescent tag were assembled and 2 µL of each PCR product were combined into multiplexes up to a total volume of 10 µL. From the multiplexed product, 2 µL was added to Hi-Di formamide and LIZ 500 size standard (Applied Biosystems, Foster City, CA, USA) for electrophoresis on a 3730 Genetic Analyzer (Applied Biosystems) at the Arizona State University DNA Lab. Fragments were sized using GENEIOUS 8.1.8 (Biomatters Ltd).

Genetic Statistical Analyses

GENALEX ver. 6.4.1 (Peakall and Smouse 2006; Peakall and Smouse 2012) was used to calculate deviation from Hardy–Weinberg equilibrium (HWE) and number of alleles for each locus (Table 2.4). Linkage disequilibrium was tested using GENEPOP ver. 4.0.10 (Raymond and Rousset 1995; Rousset 2008). Presence of null alleles was assessed using MICRO-CHECKER (Van Oosterhout et al. 2004). Genotypes were analyzed using the Bayesian cluster analysis program STRUCTURE ver. 2.4.2 (Pritchard et al. 2000). Burn-in and run-lengths of 50,000 generations were used with ten

independent replicates for each STRUCTURE analysis. STRUCTURE HARVESTER (Earl and vonHoldt 2012) was used to determine the *K* value to best describe the likely number of genetic groups for the data set. GENALEX produced a Principal Coordinate Analysis (PCoA) to examine variation in the data set. Lynch & Ritland (Lynch and Ritland 1999) mean pairwise relatedness (*r*) statistics were calculated between all 122 samples resulting in 7381 pairwise *r*-values showing degrees of relatedness. For all strains the *r*-mean and standard deviation (SD) was calculated averaging among all samples. Obvious outliers were determined by calculating the lowest *r*-mean and iteratively removing those samples to determine the relatedness among the remaining samples in the subset. A graph was generated for 12 popular strains (Table 2.3) to show how the *r*-mean value change within a strain when outliers were removed.

Results

The microsatellite analyses show genetic inconsistencies in *Cannabis* strains acquired from different facilities. While popular strains were widely available, some strains were found only at two dispensaries (Table 2.2). Since the aim of the research was not to identify specific locations where strain inconsistencies were found, dispensaries are coded to protect the identity of businesses.

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

genome. The number of alleles ranged from 5-10 across the ten loci (Table 2.4). There was no evidence of null alleles due to scoring errors.

STRUCTURE HARVESTER calculated high support (ΔK=146.56) for two genetic groups, *K*=2 (Figure 2.1). STRUCTURE assignment is shown in Figure 1 with the strains ordered by the purported proportions of Sativa phenotype (Wikileaf 2018).

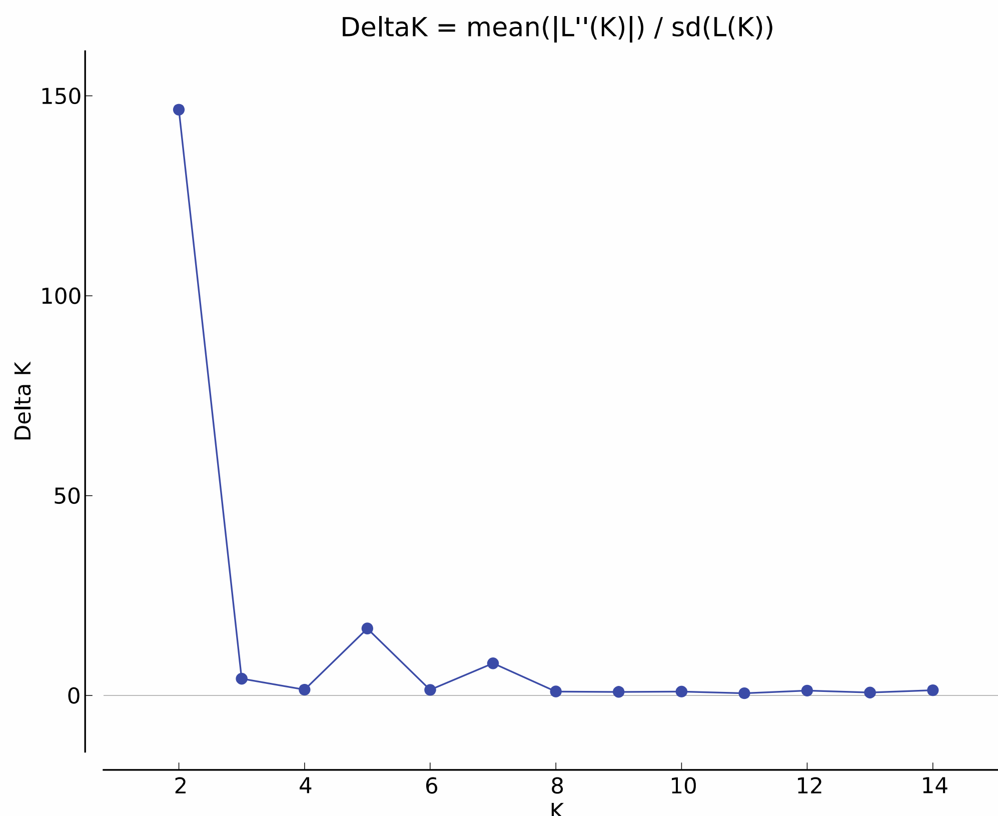


Figure 2.1. STRUCTURE HARVESTER. Graph indicating K=2 is highly support (ΔK=146.56) as the number of genetic groups for this data.

A clear genetic distinction between Sativa and Indica types would assign 100% Sativa strains ("Durban Poison") to one genotype and assign 100% Indica strains ("Purple Kush") to the other genotype (Table 2.3, Figure 2.2, and 2.3). Division into two genetic groups does not support the commonly described Sativa and Indica phenotypes. "Durban Poison" and "Purple Kush" follow what we would expect if there was support for the

Sativa/Indica division. Seven of nine “Durban Poison” (100% Sativa) samples had 96% assignment to genotype 1, and three of four “Purple Kush” (100% Indica) had 89% assignment to genotype 2 (Figure 2.2, 2.3). However, samples of “Hawaiian” (90% Sativa) and “Grape Ape” (100% Indica) do not show consistent patterns of predominant assignment to genotype 1 or 2. Interestingly, two predominantly Sativa strains “Durban Poison” (100% Sativa) and “Sour Diesel” (90% Sativa) have 86% and 14% average assignment to genotype 1, respectively. Hybrid strains such as “Blue Dream” and “Tahoe OG” (50% Sativa) should result in some proportion of shared ancestry, with assignment to both genotype 1 and 2. Eight of nine samples of “Blue Dream” show > 80% assignment to genotype 1, and three of four samples of “Tahoe OG” show < 7% assignment to genotype 1.

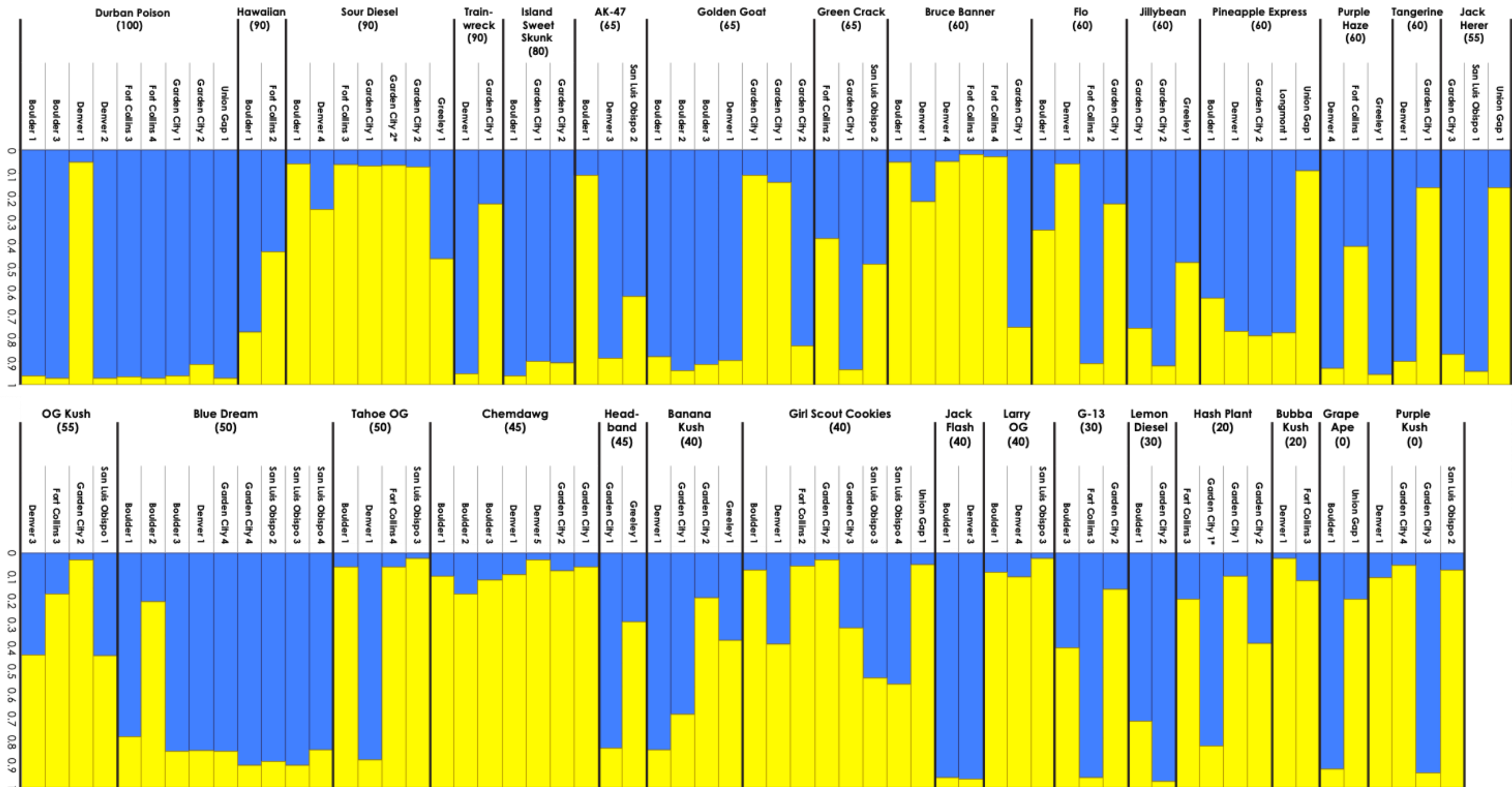


Figure 2.1. STRUCTURE graphs for 122 individuals. Bar plot graphs generated from STRUCTURE analysis for 122 individuals from 30 strains dividing genotypes into two genetic groups, K=2. Samples were arranged by purported proportions from 100% Sativa to 100% Indica and then alphabetically within each strain by city. Each strain includes reported proportion of Sativa in parentheses and each sample includes the coded location and city from where it was acquired. Each bar indicates proportion of assignment to genotype 1 (blue) and genotype 2 (yellow).

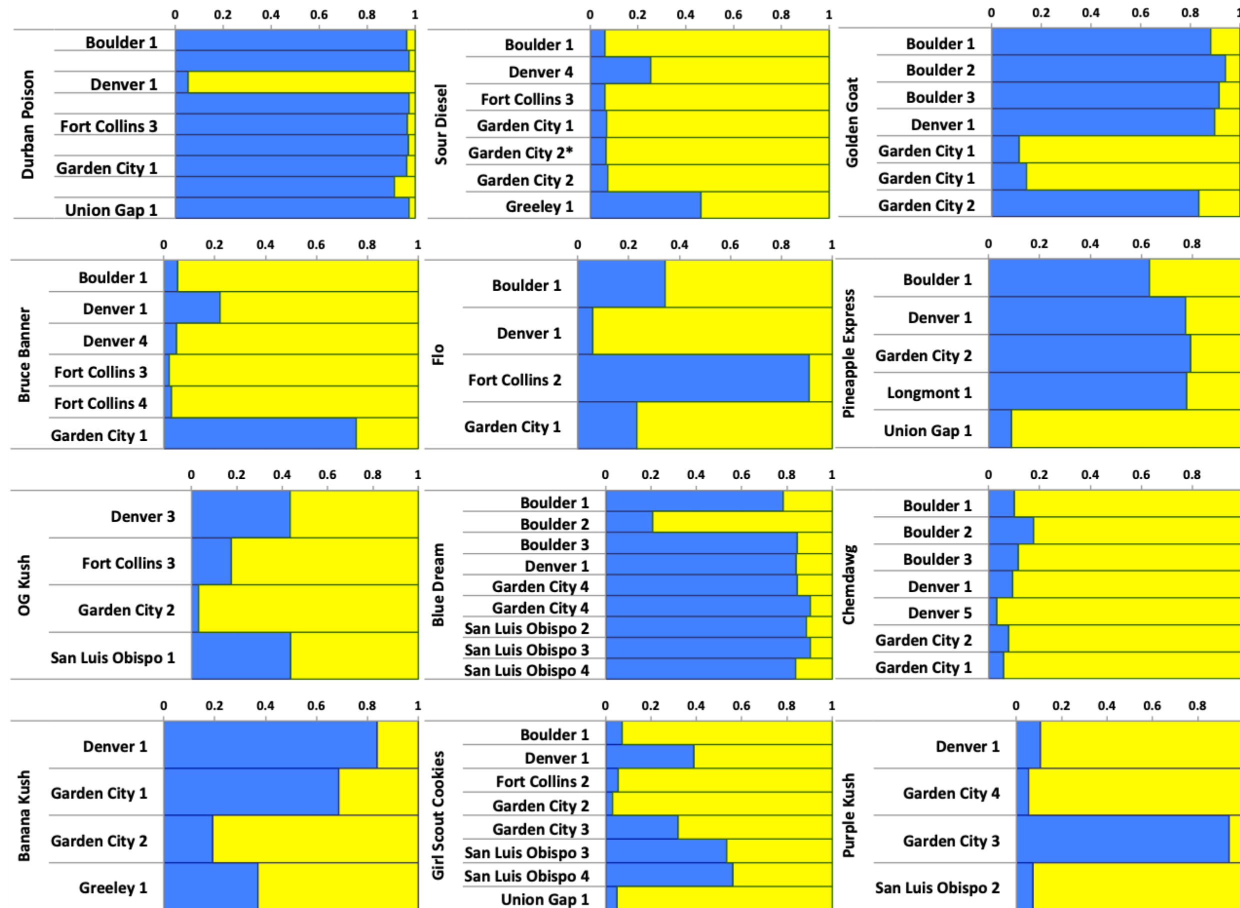


Figure 2.2. Bar plot graphs generated from STRUCTURE analysis for individuals from twelve popular strains (Table 2.3), dividing genotypes into two genetic groups, K=2. Each sample includes the coded location and city from where it was acquired. Each bar indicates proportion of assignment to genotype 1 (blue) and genotype 2 (yellow).

A Principal Coordinate Analyses (PCoA) was conducted using GENALEX (Figure 2.3). Principal Coordinate Analyses (PCoA) is organized by color from 100% Sativa types (red), through all levels of Hybrid types (green 50:50), to 100% Indica types (purple; Figure 2.3). Strain types with the same reported proportions are the same color but have different symbols. The PCoA of all strains represents 14.90% of the variation in the data on coordinate axis 1, 9.56% on axis 2, and 7.07% on axis 3 (not shown).

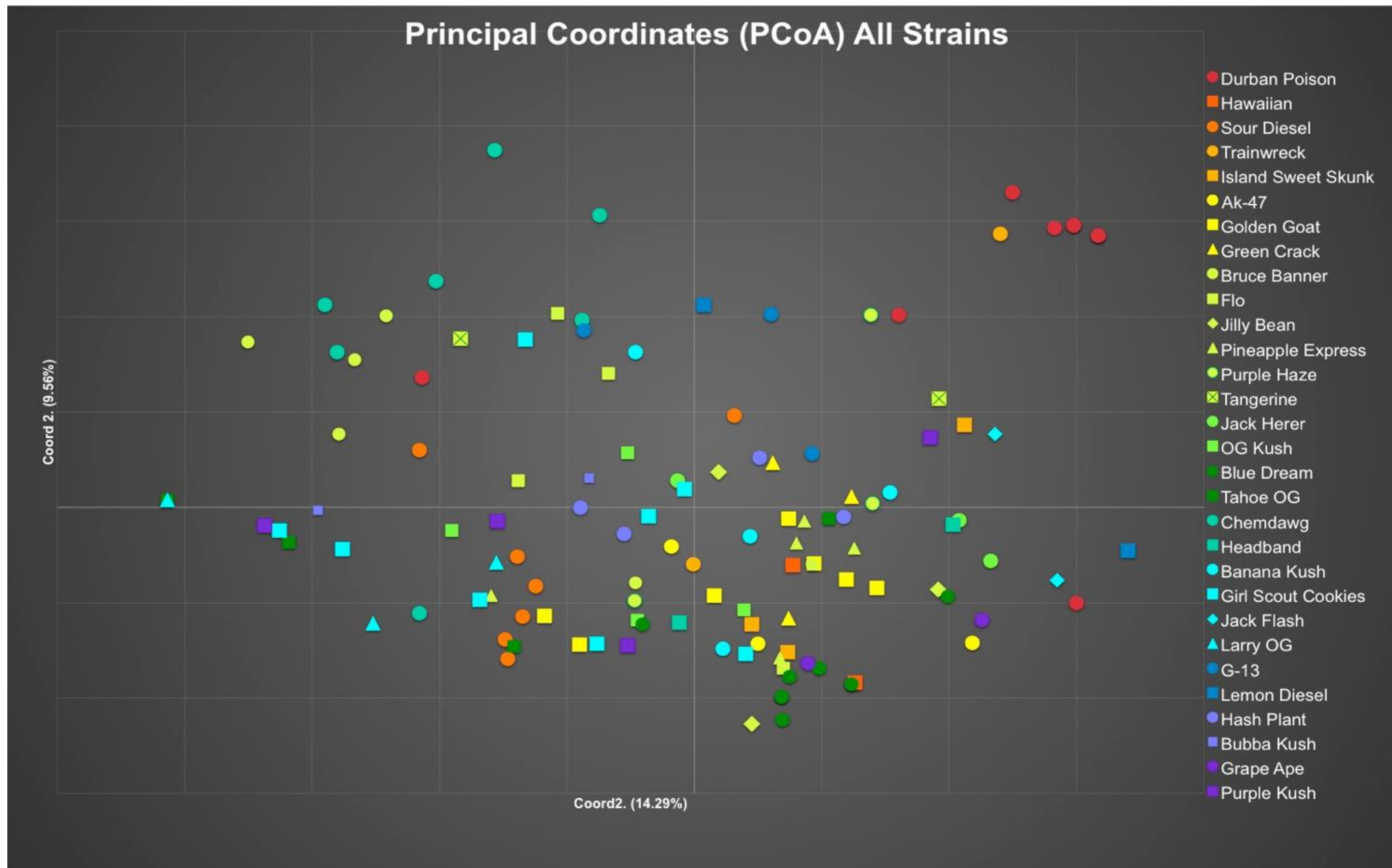


Figure 2.3. Principal Coordinates Analysis (PCoA) generated in GENALEX using Nei's genetic distance matrix. Samples are a color-coded continuum by proportion of Sativa (Table 2.2) with the strain name given for each sample: Sativa type (red: 100% Sativa proportion, Hybrid type (dark green: 50% Sativa 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%.

Lynch & Ritland (Lynch and Ritland 1999) pairwise genetic relatedness (*r*) between all 122 samples was calculated in GENALEX. The resulting 7381 pairwise *r*-values were converted to a heat map using purple to indicate the lowest pairwise relatedness value (-1.09) and green to indicate the highest pairwise relatedness value (1.00; Figure 2.4).

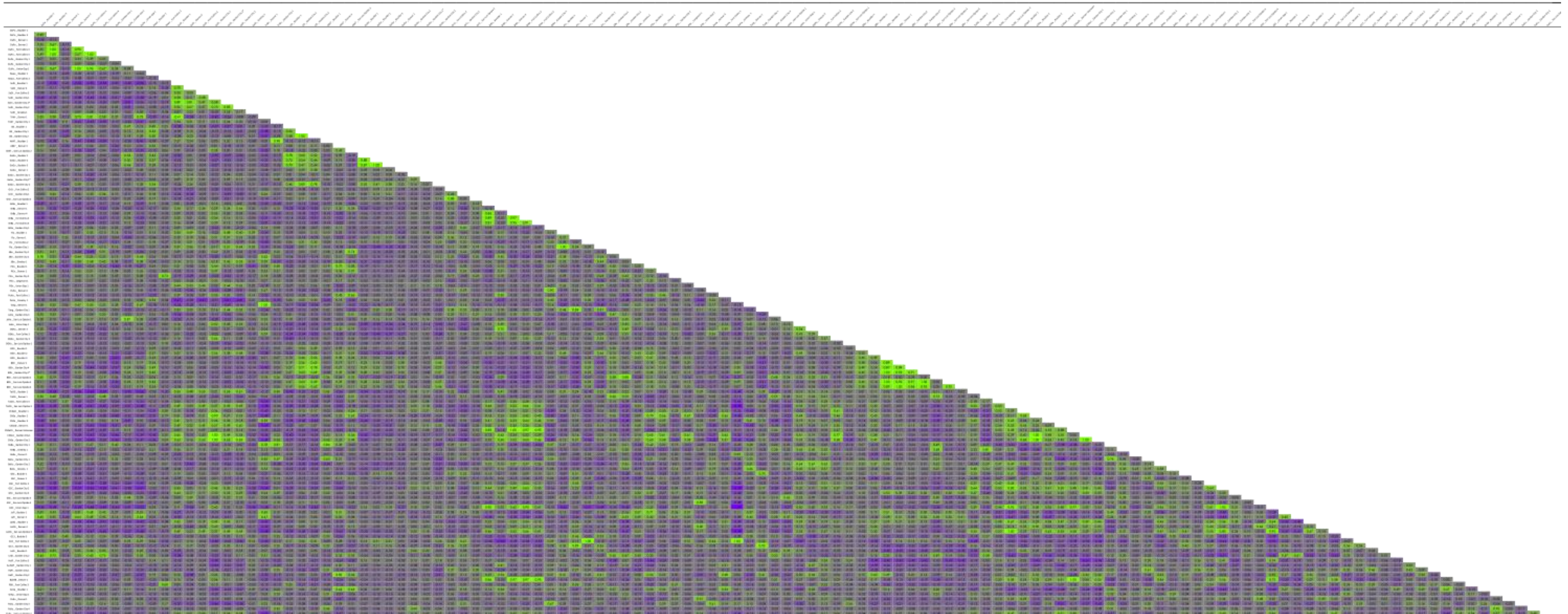
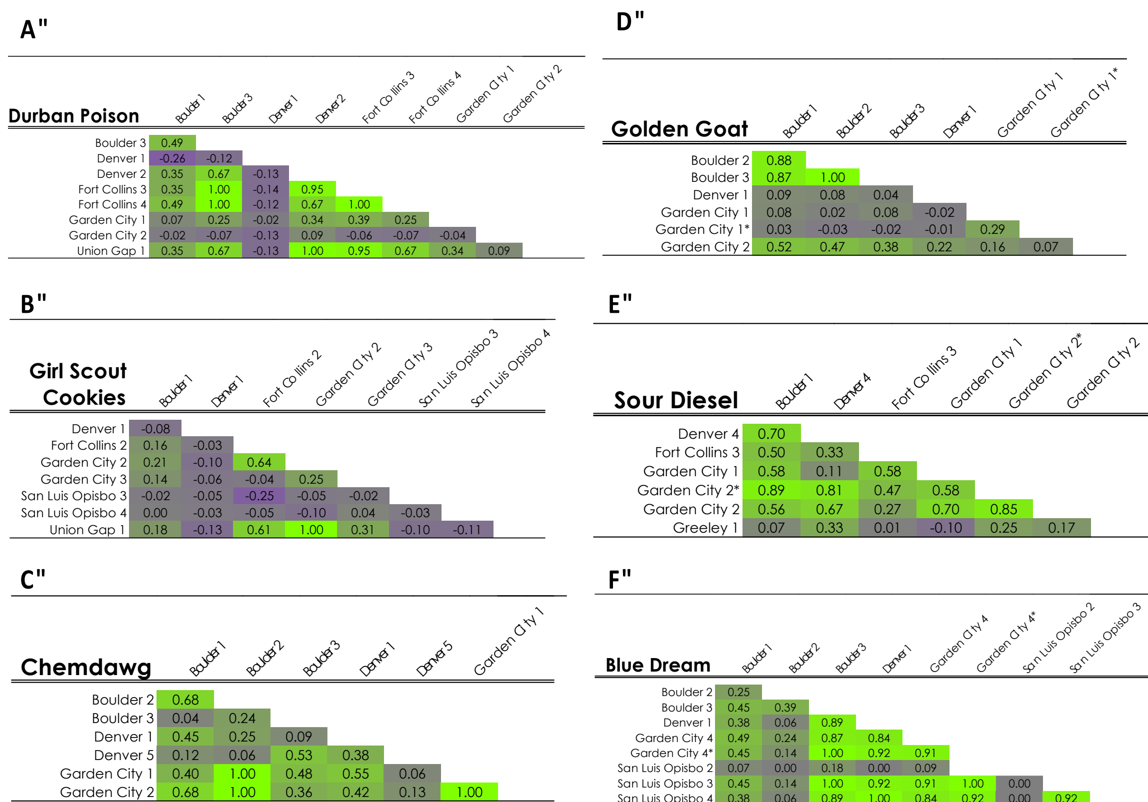


Figure 2.4. A genetic heat map chart of Lynch & Ritland pairwise genetic relatedness (*r*) values for 122 samples. Purple indicates no genetic relatedness (minimum value -1.09) and green indicates a high degree of relatedness (maximum value 1.0). Sample strain names and location of origin are indicated 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.



A"

Durban Poison	Boulder 1	Boulder 3	Denver 1	Denver 2	Fort Collins 3	Fort Collins 4	Garden City 1	Garden City 2
Boulder 3	0.49							
Denver 1	-0.26	-0.12						
Denver 2	0.35	0.67	-0.13					
Fort Collins 3	0.35	1.00	-0.14	0.95				
Fort Collins 4	0.49	1.00	-0.12	0.67	1.00			
Garden City 1	0.07	0.25	-0.02	0.34	0.39	0.25		
Garden City 2	-0.02	-0.07	-0.13	0.09	-0.06	-0.07	-0.04	
Union Gap 1	0.35	0.67	-0.13	1.00	0.95	0.67	0.34	0.09

D"

Golden Goat	Boulder 1	Boulder 2	Boulder 3	Denver 1	Garden City 1	Garden City 1*
Boulder 2	0.88					
Boulder 3	0.87	1.00				
Denver 1	0.09	0.08	0.04			
Garden City 1	0.08	0.02	0.08	-0.02		
Garden City 1*	0.03	-0.03	-0.02	-0.01	0.29	
Garden City 2	0.52	0.47	0.38	0.22	0.16	0.07

B"

Girl Scout Cookies	Boulder 1	Denver 1	Fort Collins 2	Garden City 2	Garden City 3	San Luis Opisbo 3	San Luis Opisbo 4
Denver 1	-0.08						
Fort Collins 2	0.16	-0.03					
Garden City 2	0.21	-0.10	0.64				
Garden City 3	0.14	-0.06	-0.04	0.25			
San Luis Opisbo 3	-0.02	-0.05	-0.25	-0.05	-0.02		
San Luis Opisbo 4	0.00	-0.03	-0.05	-0.10	0.04	-0.03	
Union Gap 1	0.18	-0.13	0.61	1.00	0.31	-0.10	-0.11

E"

Sour Diesel	Boulder 1	Denver 4	Fort Collins 3	Garden City 1	Garden City 2*	Garden City 2
Denver 4	0.70					
Fort Collins 3	0.50	0.33				
Garden City 1	0.58	0.11	0.58			
Garden City 2*	0.89	0.81	0.47	0.58		
Garden City 2	0.56	0.67	0.27	0.70	0.85	
Greeley 1	0.07	0.33	0.01	-0.10	0.25	0.17

C"

Chemdawg	Boulder 1	Boulder 2	Boulder 3	Denver 1	Denver 5	Garden City 1
Boulder 2	0.68					
Boulder 3	0.04	0.24				
Denver 1	0.45	0.25	0.09			
Denver 5	0.12	0.06	0.53	0.38		
Garden City 1	0.40	1.00	0.48	0.55	0.06	
Garden City 2	0.68	1.00	0.36	0.42	0.13	1.00

F"

Blue Dream	Boulder 1	Boulder 2	Boulder 3	Denver 1	Garden City 4	Garden City 4*	San Luis Opisbo 2	San Luis Opisbo 3
Boulder 2	0.25							
Boulder 3	0.45	0.39						
Denver 1	0.38	0.06	0.89					
Garden City 4	0.49	0.24	0.87	0.84				
Garden City 4*	0.45	0.14	1.00	0.92	0.91			
San Luis Opisbo 2	0.07	0.00	0.18	0.00	0.09			
San Luis Opisbo 3	0.45	0.14	1.00	0.92	0.91	1.00	0.00	
San Luis Opisbo 4	0.38	0.06	0.89	1.00	0.84	0.92	0.00	0.92

Figure 2.5. Heat maps of six prominent strains (A-F) using Lynch & Ritland (Lynch and 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 1.00). 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 the cell color reflects the degree to which two individuals are related.

Comparisons are detailed for six popular strains (Figure 2.5) to illustrate the relationship of samples from different sources and the impact of outliers. Values of close to 1.00 indicate a high degree of relatedness (Lynch and Ritland 1999), which could be indicative of clones or seeds from the same mother (Green 2005; SeedFinder 2018). First order relatives (full siblings or mother-daughter) share 50% genetic identity (*r*-value = 0.50), second order relatives (half siblings or cousins) share 25% genetic identity (*r*-value = 0.25), and unrelated individuals are expected to have an *r*-value of 0.00 or lower.

Negative values arise when individuals are less related than expected under normal panmictic conditions (Moura et al. 2013; Norman et al. 2017).

Individual pairwise *r*-values were averaged within strains to calculate the overall *r*-mean as a measure of genetic similarity within strains which ranged from -0.22 ("Tangerine") to 0.68 ("Island Sweet Skunk") (Table 6). Standard deviations ranged from 0.04 ("Jack Herer") to 0.51 ("Bruce Banner"). The strains with higher standard deviation values indicate a wide range of genetic relatedness within a strain, while low values indicate that samples within a strain share similar levels of genetic relatedness. In order to determine how outliers impact the overall relatedness in a strain, the farthest outlier (lowest pairwise *r*-mean value) was removed and the overall *r*-means and SD values within strains were recalculated (Table 2.5).

In all strains, the overall *r*-means increased when outliers were removed. In strains with more than three samples, a second outlier was removed, and the overall *r*-means and SD values were recalculated. Overall *r*-means were used to determine degree of relatedness as clonal (or from stable seed; overall *r*-means > 0.9), first or higher order relatives (overall *r*-means 0.46 – 0.89), second order relatives (overall *r*-means 0.26 - 0.45), low levels of relatedness (overall *r*-means 0.00 - 0.25), and not related (overall *r*-means <0.00). Overall *r*-means are displayed for all 30 strains (Table 2.5), and graphically for 12 popular strains (Figure 2.6).

Table 2.5. Lynch & Ritland (Lynch and 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.

Strain	# Samples	Mean ± SD	Mean ± SD (Outlier 1 removed)	Mean ± SD (Outlier 2 removed)
Durban Poison*	9	0.31 ± 0.4	0.43 ± 0.37	0.58 ± 0.30
Hawaiian	2	-0.115	-	-
Sour Diesel*	7	0.44 ± 0.29	0.57 ± 0.22	0.60 ± 0.18
Trainwreck	2	-0.001	-	-
Island Sweet Skunk	3	0.68	1	-
AK-47	3	0.16	0.45	-
Golden Goat*♥	7	0.25 ± 0.32	0.31 ± 0.36	0.46 ± 0.36
Green Crack♥	3	0.38	0.88	-
Bruce Banner*	6	0.30 ± 0.51	0.51 ± 0.5	0.9 ± 0.05
Flo*	4	0.29 ± 0.38	0.55 ± 0.39	-
Jillybean	3	-0.033	0.039	-
Pineapple Express*	5	0.02 ± 0.16	0.04 ± 0.17	0.13 ± 0.19
Purple Haze	3	0.041	0.26	-
Tangerine	2	-0.22	-	-
Jack Herer	3	0.1	0.13	-
OG Kush*♥	4	0.13 ± 0.19	0.25 ± 0.22	-
Blue Dream*♥	9	0.50 ± 0.39	0.63 ± 0.34	0.76 ± 0.24
Tahoe OG	4	0.21	0.406	0.539
Chemdawg*	7	0.42 ± 0.31	0.51 ± 0.31	0.64 ± 0.28
Headband	2	0.107	-	-
Banana Kush*	4	0.13 ± 0.20	0.24 ± 0.13	-
Girl Scout Cookies*♥	8	0.08 ± 0.27	0.13 ± 0.30	0.22 ± 0.32
Jack Flash	2	0.62	-	-
Larry OG	3	0.32	0.67	-
G-13	3	0.29	0.562	-
Lemon Diesel♥	2	0.1	-	-
Hash Plant	4	0.25	0.25	0.43
Pre98-Bubba Kush	2	-0.02	-	-
Grape Ape	2	-0.05	-	-
Purple Kush*♥	4	0.03 ± 0.21	0.16 ± 0.22	-

* Twelve popular strains

♥ Clone only strains (SeedFinder 2018)

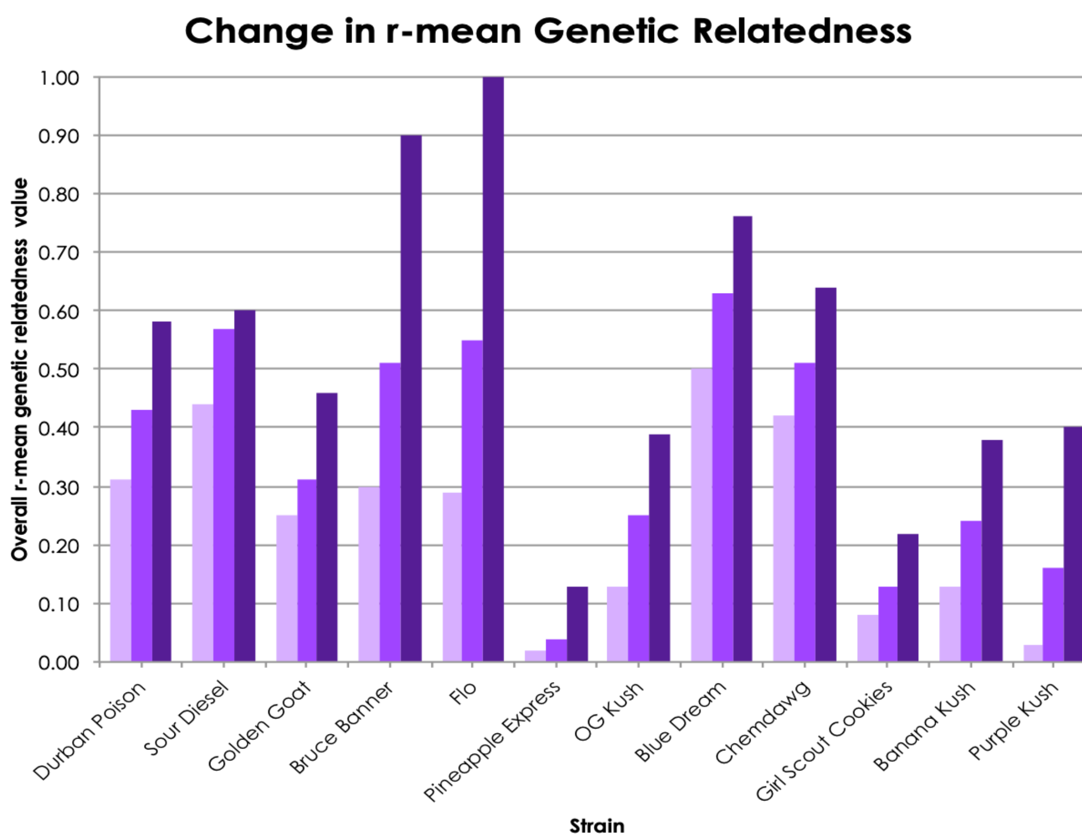


Figure 2.6. Iterative removal of pairwise genetic relatedness *(r)* outliers from Table 2.5. This graph indicates the mean pairwise genetic relatedness (*r*) initially (light purple), and after the removal of one (medium purple) or two (dark purple) outlying samples in 12 popular strains.

Initial overall *r*-means indicate only three strains are first or higher order relatives (Table 2.5). Removing first or second outliers, depending on sample size, revealed that the remaining samples for an additional ten strains are first or higher order relatives (0.46 – 1.00), three strains are second order relatives (*r*-means 0.26 - 0.45), ten strains show low levels of relatedness (*r*-means 0.00 - 0.25; Table 2.5), and five strains are not related (*r*-means <0.00). The impact of outliers can be clearly seen in the heat map for "Durban Poison" which shows the relatedness for 36 comparisons (Figure 2.5A), six of which are nearly identical (*r*-value 0.90 - 1.0), while 13 are not related (*r*-value <0.00). However,

removal of two outliers, Denver 1 and Garden City 2, reduces the number of comparisons ranked as not related from 13 to zero.

Discussion

Cannabis is becoming an ever-increasing topic of discussion, so it is important that scientists and the public can discuss *Cannabis* in a similar manner. Currently, not only are Sativa and Indica types disputed (Clarke and Merlin 2013, 2015, 2016; De Meijer and Keizer 1996; Emboden 1974; Hillig 2005; McPartland 2017; Piomelli and Russo 2016; Russo 2007; Small 2015a), but experts also are at odds about nomenclature for *Cannabis* (Clarke and Merlin 2013, 2015, 2016; De Meijer and Keizer 1996; Emboden 1974; Hillig 2005; McPartland 2017; Piomelli and Russo 2016; Russo 2007; Small 2015a). We postulated that genetic profiles from samples with the same strain identifying name should have identical, or at least, highly similar genotypes no matter the source of origin. The multiple genetic analyses used here address paramount questions for the medical *Cannabis* community and bring empirical evidence to support claims that inconsistent products are being distributed. An important element for this study is that samples were acquired from multiple locations to maximize the potential for variation among samples. Maintenance of the genetic integrity through genotyping is possible only following evaluation of genetic consistency and continuing to overlook this aspect will promote genetic variability and phenotypic variation within *Cannabis*. Addressing strain variability at the molecular level is of the utmost importance while the industry is still relatively new.

Genetic analyses have consistently found genetic distinction between hemp and marijuana, but no clear distinction has been shown between the common description of Sativa and Indica types (De Meijer and Keizer 1996; Dufresnes et al. 2017; Lynch et al.

2016; Sawler et al. 2015; Soler et al. 2017). We found high support for two genetic groups in the data (Figure 2.1) but no discernable distinction or pattern between the described Sativa and Indica strains. The color-coding of strains in the PCoA for all 122 samples allows for visualization of clustering among similar phenotypes by color: Sativa (red/orange), Indica (blue/purple) and Hybrid (green) type strains (Figure 2.4). If genetic differentiation of the commonly perceived Sativa and Indica types previously existed, it is no longer detectable in the neutral genetic markers used here. Extensive hybridization and selection have presumably created a homogenizing effect and erased evidence of potentially divergent historical genotypes.

Wikileaf maintains that the proportions of Sativa and Indica reported for strains are largely based on genetics and lineage (Nelson 2016), although online databases do not give scientific evidence for their categorization other than parentage information from breeders and expert opinions. This has seemingly become convoluted over time (Clarke and Merlin 2013; Russo 2007; Small 2015b; Small 2016). Our results show that commonly reported levels of Sativa, Indica and Hybrid type strains are often not reflected in the average genotype. For example, two described Sativa type strains “Durban Poison” and “Sour Diesel”, have contradicting genetic assignments (Figure 2.1, 2.2 & Table 2.1). This analysis indicates strains with similar reported proportions of Sativa or Indica may have differing genetic assignments. Further illustrating this point is that “Bruce Banner”, “Flo”, “Jillybean”, “Pineapple Express”, “Purple Haze”, and “Tangerine” are all reported to be 60/40 Hybrid type strains, but they clearly have differing levels of admixture both within and among these reportedly similar strains (Figure 2.1, Table 2.1). From these results, we can conclude that reported ratios or differences between Sativa and Indica phenotypes are not discernable using these genetic markers. Given the lack of genetic

distinction between Indica and Sativa types, it is not surprising that reported ancestry proportions are also not supported.

To accurately address reported variation within strains, samples were purchased from various locations, as a customer, with no information of strains other than publicly available online information. Evidence for genetic inconsistencies is apparent within many strains and supported by multiple genetic analyses. Soler et al. (Soler et al. 2017) found genetic variability among seeds from the same strain supplied from a single source, indicating genotypes within strains are variable. When examining the STRUCTURE genotype assignments, it is clear that many strains contained one or more divergent samples with a difference of > 0.10 genotype assignment (e.g. "Durban Poison" – Denver 1; Figure 2.1, 2.2, 2.5A). Of the 30 strains examined, only four strains had consistent STRUCTURE genotype assignment and admixture among all samples. The number of strains with consistent STRUCTURE assignments increased to 11 and 15 when one or two samples were ignored, respectively. These results indicate that half of the included strains showed relatively stable genetic identity among most samples. Six strains had only two samples, both of which were different (e.g., "Trainwreck" and "Headband"). The remaining nine strains in the analysis had more than one divergent sample (e.g., "Sour Diesel") or had no consistent genetic pattern among the samples within the strain (e.g., "Girl Scout Cookies"; Table 2.5, Figure 2.2, 2.2, 2.5B). It is noteworthy that many of the strains used here fell into a range of genetic relatedness indicative of first order siblings (*see Lynch & Ritland analysis*) when samples with high genetic divergence were removed from the data set (Table 2.5; Figure 2.6). Eight of the 30 strains examined are identified as clone only (Table 2.3). All eight of the strains described as clone only show differentiation of at least one sample within the strain (Figure 2.1). For example, one

sample of “Blue Dream” is clearly differentiated from the remaining eight, and “Girl Scout Cookies” has little genetic cohesiveness among the eight samples (Figure 2.1, 2.2). Other genetic studies have similarly found genetic inconsistencies across samples within the same strain (Lynch et al. 2016; Sawler et al. 2015; Soler et al. 2017). These results lend support to the idea that unstable genetic lines are being used to produce seed.

A pairwise genetic heat map based on Lynch & Ritland (Lynch and Ritland 1999) pairwise genetic relatedness (*r*-values) was generated to visualize genetic relatedness throughout the data set (Figure 2.3). Values of 1.00 (or close to) are assumed to be clones or plants from self-fertilized seed. Six examples of within-strain pairwise comparison heat maps were examined to illustrate common patterns (Figure 2.4). The heat map shows that many strains contain samples that are first order relatives or higher (*r*-value > 0.49). For example, “Sour Diesel” (Figure 2.4) has 12 comparisons of first order or above, and six have low/no relationship. There are also values that could be indicative of clones or plants from a stable seed source such as “Blue Dream” (Figure 2.2, 2.5F), which has 10 nearly identical comparisons (*r*-value 0.90-1.00), and no comparisons in “Blue Dream” have negative values. While “Blue Dream” has an initial overall *r*-mean indicating first order relatedness within the samples (Table 2.5, Figure 2.5F), it still contains more variation than would be expected from a clone only strain (SeedFinder 2018). Other clone-only strains (SeedFinder 2018) e.g., “Girl Scout Cookies” (Table 2.2, Figure 2.5B) and “Golden Goat” (Table 3, Figure 6D), have a high degree of genetic variation resulting in low overall relatedness values. Outliers were calculated and removed iteratively to demonstrate how they affected the overall *r*- mean within the 12 popular strains (Table 2.2, Figure 2.6). In all cases, removing outliers increased the mean *r*-value, as illustrated by “Bruce Banner”, which increased substantially, from 0.3 to 0.9

when samples with two outlying genotypes were removed. There are unexpected areas in the entire data set heat map that indicate high degrees of relatedness between different strains (Figure 2.4). For example, comparisons between “Golden Goat” and “Island Sweet Skunk” (overall *r*- mean 0.37) are higher than within samples of “Sour Diesel”. Interestingly, “Golden Goat” is reported to be a hybrid descendant of “Island Sweet Skunk” (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018) which could explain the high genetic relatedness between these strains. However, most of the between strain overall *r*- mean are negative (e.g., “Golden Goat” to “Durban Poison” -0.03 and “Chemdawg” to “Durban Poison” -0.22; Figure 2.5), indicative of limited recent genetic relationship.

While collecting samples from various dispensaries, it was noted that strains of “Chemdawg” had various different spellings of the strain name, as well as numbers and/or letters attached to the name. Without knowledge of the history of “Chemdawg”, the assumption was that these were local variations. These were acquired to include in the study to determine if and how these variants were related. Upon investigation of possible origins of “Chemdawg”, an interesting history was uncovered, especially in light of the results. Legend has it that someone named “Chemdog” (a person) grew the variations (“Chem Dog”, “Chem Dog D”, “Chem Dog 4”) from seeds he found in a single bag of *Cannabis* purchased at a Grateful Dead concert (Danko 2016). However, sampling suggests dispensaries use variations of the name, and more often the “Chemdawg” form of the name is used, albeit incorrectly (Danko 2016). The STRUCTURE analysis indicates only one “Chemdawg” individual has > 0.10 genetic divergence compared to the other six samples (Figure 2.1, 2.5C). Five of seven “Chemdawg” samples cluster in the PCoA (Figure 4), and six of seven “Chemdawg” samples are first order relatives (*r*-

value >0.50; Table 2.2, Figure 2.5C). The history of "Chem Dog" is currently unverifiable, but the analysis supports that these variations could be from seeds of the same plant. This illustrates how *Cannabis* strains may have come to market in a non-traditional manner. Genetic analyses can add scientific support to the stories behind vintage strains and possibly help clarify the history of specific strains.

Genetic inconsistencies may come from both suppliers and growers of *Cannabis* clones and stable seed, because currently they can only assume the strains they possess are true to name. There is a chain of events from seed to sale that relies heavily on the supplier, grower, and dispensary to provide the correct product, but there is currently no reliable way to verify *Cannabis* strains. The possibility exists for errors in plant labeling, misplacement, misspelling (e.g., "Chem Dog" vs. "Chemdawg"), and/or relabeling along the entire chain of production. Although the expectation is that plants are labeled carefully and not re-labeled with a more desirable name for a quick sale, these misgivings must be considered. Identification by genetic markers has largely eliminated these types of mistakes in other widely cultivated crops such as grapes, olives and apples. Modern genetic applications can accurately identify varieties and can clarify ambiguity in closely related and hybrid species (Costantini et al. 2005; Guilford et al. 1997; Hokanson et al. 1998; Sarri et al. 2006; United States Department of Agriculture 2014).

Matching genotypes within the same strains were expected, but highly similar genotypes between samples of different strains could be the result of mislabeling or misidentification, especially when acquired from the same source. The pairwise genetic relatedness *r*-values were examined for incidence of possible mislabeling or re-labeling. There were instances in which different strains had *r*-values = 1.0 (Figure 2.4, 2.5), indicating clonal genetic relationships. Two samples with matching genotypes were

obtained from the same location ("Larry OG" and "Tahoe OG" from San Luis Obispo 3, Figure 2.1). This could be evidence for mislabeling or misidentification because these two samples have similar names. It is unlikely that these samples from reportedly different strains have identical genotypes, and more likely that these samples were mislabeled at some point. Misspelling may also be a source of error, especially when facilities are handwriting labels. An example of possible misspelling may have occurred in the sample labeled "Chemdog 1" from Garden City 1. "Chemdawg 1", a described strain, could have easily been misspelled, but it is unclear whether this instance is evidence for mislabeling or renaming a local variant. Inadvertent mistakes may carry through to scientific investigation where strains are spelled or labeled incorrectly. For example, Vergara et al. (Vergara et al. 2016) reports genome assemblies for "Chemdog" and "Chemdog 91" as they are reported in GenBank (GCA_001509995.1), but neither of these labels are recognized strain names. "Chemdawg" and "Chemdawg 91" are recognized strains (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018; Marijuana strains database 2019), but according to the original source, the strain name "Chemdawg" is incorrect, and it should be "Chem Dog" (Danko 2016), but the name has clearly evolved among growers since it emerged in 1991 (Danko 2016). Another example that may lead to confusion is how information is reported in public databases. For example, data are available for the reported monoisolate of "Pineapple Banana Bubba Kush" in GenBank (SAMN06546749), and while "Pineapple Kush", "Banana Kush" and "Bubba Kush" are known strains (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Seedfinder 2018; Wikileaf 2018; Marijuana strains database 2019), the only record we found of "Pineapple Banana Bubba Kush" is in GenBank. This study has highlighted several possible sources of error and how genotyping can serve to

uncover sources of variation. Although this study was unable to confirm sources of error, it is important that producers, growers and consumers are aware that there are errors and they should be documented and corrected whenever possible.

Conclusions

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

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

CHAPTER III

GENETIC VARIATION PART 2: RESEARCH GRADE MARIJUANA SUPPLIED BY THE NATIONAL INSTITUTE ON DRUG ABUSE IS GENETICALLY DIVERGENT FROM COMMERCIALLY AVAILABLE *CANNABIS*

Contributions of Authors and Co-Authors

Manuscript in Chapter IV

Author: Anna Schwabe

Contributions: Conceived the project, collected samples, conducted DNA extractions, designed and optimized microsatellite primers, compiled and analyzed data, and drafted manuscript content.

Author: Connor Hansen

Contributions: Conducted DNA extractions, compiled and analyzed data, and prepared the first draft of the manuscript.

Author: Dr. Richard Hyslop

Contributions: Provided DNA from NIDA samples, contributed manuscript revisions.

Author: Dr. Mitchell McGlaughlin

Contributions: Directed the project, provided some funding, contributed statistical analysis and manuscript revisions.

Abstract

Public comfort with *Cannabis* (marijuana and hemp) has recently increased, resulting in revisions of previously strict *Cannabis* regulations to now allow for hemp cultivation, medical use, and in some states, recreational consumption. There is a growing interest in the potential medical benefits of the various chemical constituents produced by the *Cannabis* plant. Currently, the University of Mississippi, funded through the National Institutes of Health/National Institute on Drug Abuse (NIH/NIDA), is the sole Drug Enforcement Agency (DEA) licensed facility to cultivate *Cannabis* for research purposes. Hence, most federally funded research where participants consume *Cannabis* for medicinal purposes relies on NIDA supplied product. Previous research found that cannabinoid levels in research grade marijuana supplied by NIDA did not align with commercially available *Cannabis* from Colorado, Washington and California. Given NIDA chemotypes were found to misaligned with commercial *Cannabis,* we sought to investigate where NIDA's research grade marijuana falls on the genetic spectrum of *Cannabis* groups. NIDA research grade marijuana was found to genetically group with Hemp samples along with a small subset of commercial drug-type *Cannabis*. A majority of commercially available drug-type *Cannabis* was genetically very distinct from NIDA samples. These results suggested that subjects consuming NIDA research grade marijuana may have different effects than average consumers

Introduction

Humans have a long history with *Cannabis sativa* (marijuana and hemp), with evidence of cultivation dating back as far as 10,000 years ago (Abel 2013). The World Health Organization proclaims *Cannabis* as the most widely cultivated, trafficked and abused illicit drug, and reports over half of worldwide drug seizures are of *Cannabis* (World Health Organization 2018). Phytochemicals of interest in *Cannabis* are primarily Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA), both of which require a decarboxylation conversion to the biologically active forms, THC and CBD, respectively. The United States is currently experiencing drastic changes in patterns of *Cannabis* use associated with widespread relaxation of laws that previously limited both medical and recreational marijuana consumption (Cousijn et al. 2018) and hemp cultivation. This has led to a need for extensive research into the basic biology and taxonomy of *Cannabis sativa* (Clarke and Merlin 2013; Hillig 2005; Lynch et al. 2016; Small 2017; Vergara et al. 2016), and the possible benefits and threats from *Cannabis* consumption (Baron 2018; Cousijn et al. 2018).

Although *Cannabis sativa* is the only described species in the genus *Cannabis* (Cannabaceae), there are several commonly described subcategories of *Cannabis* that are widely recognized. There are two primary *Cannabis* usage groups, which are well supported by genetic analyses (Dufresnes et al. 2017; Lynch et al. 2016; Sawler et al. 2015; Soler et al. 2017): ***hemp*** is defined by a lack of THC (< 0.3% THC in the U.S.), and ***marijuana*** or ***drug-types*** have moderate to high THC concentrations (> 0.3% THC in the U.S.). Hemp-type *Cannabis* tends to have higher concentrations of CBD than drug-types (de Meijer et al. 1992). Drug-type *Cannabis* usually contains > 12% THC and

averages ~ 10-23% THC in commercially available dispensaries (Jikomes and Zoorob 2018; Potter et al. 2008; Vergara et al. 2017). Within the two major usage groups, *Cannabis* can be further divided into varietals, which are referred to as strains. The drug-type strains are commonly categorized further: ***Sativa*** strains reportedly have uplifting and more psychedelic effects, ***Indica*** strains reportedly have more relaxing and sedative effects, and ***Hybrid*** strains, which result from breeding Sativa and Indica strains, have a spectrum of intermediate effects. There is extensive debate among experts surrounding the appropriate taxonomic treatment of *Cannabis* groups, which is confounded by colloquial usage of these terms versus what researchers suggest is more appropriate nomenclature (Clarke and Merlin 2013, 2015; Emboden 1977, 1981; McPartland 2017; McPartland and Guy 2017; Small 2015b; Small 2016; Small et al. 1976). Commercially available drug-type strains for medical or recreational consumption are labeled with a strain name, as well as the levels of THC and often CBD as a percent of the dry weight. Genetic analyses have not shown clear and consistent differentiation among the three commonly described drug-type strains (Lynch et al. 2016; Sawler et al. 2015), but both the recreational and medical *Cannabis* communities maintain there are distinct differences in effects between Sativa and Indica strains (Smith 2012; Leaf Science 2016; Leafly 2018b).

Although *Cannabis* has been federally controlled since 1937 (1937), many states now allow regulated medical (33 states and the District of Columbia) and recreational use (10 states and the District of Columbia) (ProCon 2018b). There were > 3.5 million registered medical marijuana patients reported as of May 2018 (ProCon 2018a). However, the United States Drug Enforcement Agency (DEA) lists *Cannabis sativa* as a

Schedule 1 Substance (United States Congress 1970), and as such, research on all aspects of this plant has been limited. U.S. Surgeon General Jerome Adams recently expressed concern that the current scheduling in the most restrictive category is inhibiting research on *Cannabis* as a potentially therapeutic plant (Jaeger 2018). A Schedule 1 substance is described as a drug with no accepted medical use and a high potential for abuse (United States Congress 1970). The University of Mississippi, funded through the National Institutes of Health/National Institute on Drug Abuse (NIH/NIDA), currently holds the single license issued by the DEA for the cultivation of *Cannabis* for research purposes (National Institute on Drug Abuse 2018). As such, NIDA serves as the sole legal provider of *Cannabis* for federally funded medical research in the United States. Bulk research grade marijuana supplied by NIDA is characterized by the level of THC and CBD. They offer *Cannabis* for research with four levels of THC: ***low*** (< 1%), ***medium*** (1-5 %), ***high*** (5-10 %) and ***very high*** (>10%), with the additional option of four levels of CBD: ***low*** (< 1%), ***medium*** (1-5%), ***high*** (5-10%) and ***very high*** (> 10%).

The National Institute on Drug Abuse funds a wide range of research on drug-type *Cannabis*, including long and short-term effects on behavior, pain, mental illness, brain development, use and abuse, and impacts of policy changes related to marijuana (National Institute of Health and National Institute on Drug Abuse 2018a; National Institute of Health and National Institute on Drug Abuse 2018b). Additionally, the NIH provides support for researching cannabinoids as separate constituents. Funding for CBD related research is reported as $36M (2015 - 2017) and projected to be $36M for 2018 - 2019 (National Institute of Health and National Institute on Drug Abuse 2018b), while cannabinoid related research is reported as $366M from 2015 - 2017 and projected to be

$292M for 2018 - 2019 (National Institute of Health and National Institute on Drug Abuse 2018a).

Recent research has documented that NIDA provided *Cannabis* has distinctly different cannabinoid profiles than commercially available *Cannabis* (Vergara et al. 2017). Specifically, Vergara et al. (2017) found that NIDA samples contained only 27% of the amount of THC and 48% of CBD levels of commercially available *Cannabis*. The substantial chemical differences between NIDA and commercially available *Cannabis* raises significant questions about whether research conducted with federal *Cannabis* is indicative of the experience consumers are having.

Medical research on *Cannabis* primarily focuses on THC and CBD (Baron 2018; Borgelt et al. 2013; Citti et al. 2018; Cousijn et al. 2018; Maa and Figi 2014; Minkin 2014; National Institute of Health and National Institute on Drug Abuse 2018a, 2018b), but there are hundreds of other chemical constituents in *Cannabis* (ElSohly 2007), including cannabinoids and terpenes, which have largely been ignored (Baron 2018). There is evidence to suggest that chemical constituents in various combinations and abundances work in concert in various ways to create the suite of physiological effects reported (Baron 2018). The chemical makeup of each variant of *Cannabis* is influenced by the genetic makeup as well as environmental conditions. Given that previous research has determined the cannabinoid levels of research grade marijuana from NIDA is significantly different from commercially available *Cannabis* (Vergara et al. 2017), genetic investigations are warranted to determine if NIDA *Cannabis* is genetical distinct from other sources.

In the current study we investigated the genetic relationship of NIDA provided *Cannabis* to commercially available drug-type strains, as well as feral and cultivated hemp. Ten variable nuclear microsatellite regions were used to examine genetic differentiation among our samples. Sampling included NIDA (High THC and High THC/CBD), high THC drug-type, low THC/high CBD drug-type, wild growing hemp (presumed escapees from cultivation), and commercial hemp. This study aimed to investigate where research grade marijuana supplied by NIDA falls on the genetic spectrum of *Cannabis* groups.

Methods

Sampling

A total of 49 *Cannabis* samples were used in this research (Table 7), including: wild hemp (5), cultivated hemp (4), NIDA strains (2), high CBD drug-type strains (3), and drug-types strains (35). Drug-type strains were further subdivided into three commonly used categories: Sativa (11), Hybrid (14), and Indica (10) based on information available online (Leafly 2018b; Wikileaf 2018). The drug-type strains were randomly chosen from a much larger pool of samples. Duplicate accessions within strains were not included.

Table 3.1. Sampling Information. Sample names, ID code, accession number/ strain name, and the suppliers name and location

Name	ID Code	Accession/ Strain Name	Supplier Origin	City	State
Wild Hemp 1	1019	Hemp 1019	DBG Herbarium	Denver	Colorado
Wild Hemp 2	24845	Hemp 24845 Male	DBG Herbarium	Denver	Colorado
Wild Hemp 3	25572	Hemp 25572 Male	DBG Herbarium	Denver	Colorado
Wild Hemp 4	22831M	Hemp 22831 Male	UNC Herbarium	Greeley	Colorado
Wild Hemp 5	28381M	Hemp 28381 Male	DBG Herbarium	Denver	Colorado
Wild Hemp 6	UnkM	Hemp Unknown Male	Cannabis Genomic Research Initiative	Boulder	Colorado
Wild Hemp 7	Cara#2_4	Hemp Cara#2	Cannabis Genomic Research Initiative	Boulder	Colorado
Wild Hemp 8	Carm	Hemp Carmagnola	Colorado Seed, Caren Kershner	Co. Springs	Colorado
Wild Hemp 9	CoGo	Hemp Colorado Gold	Colorado Seed, Caren Kershner	Co. Springs	Colorado
NIDA THC	NIDA_THC	NIDA High THC	Univ. of Mississippi	Mississippi	Colorado
NIDA THC/CBD	NIDA_THC-CBD	NIDA THC/CBD	Univ. of Mississippi	Mississippi	Colorado
Otto (High CBD)	Otto1	Otto (High CBD)	Centennial Seeds	Co. Springs	Colorado
Juanita La Lagrimosa	JLL_2	Juanita La Lagrimosa (High CBD)	Nature's Herbs and Wellness	Garden City	Colorado
Fuck Cancer	FuCa	Fuck Cancer	Matt Kahl	Co. Springs	Colorado
Durban Poison	DuPo_19	Durban Poison	The Kind Room	Denver	Colorado
El Dorado	ElDo_1	El Dorado	Smokey's 420	Garden City	Colorado
Hawaiian	Hawa_9	Hawaiian	Best Colorado Meds	Fort Collins	Colorado
Sour Diesel	SoDi_2a	Sour Diesel	Nature's Herbs and Wellness	Garden City	Colorado
Island Sweet Skunk	ISS_1	Island Sweet Skunk	Smokey's 420	Garden City	Colorado
Agent Orange	AgOr_1	Agent Orange	Nature's Herbs and Wellness	Garden City	Colorado
Cinderella 99	Cin99_1	Cinderella 99	Smokey's 420	Garden City	Colorado
AK-47	AK47_21	AK-47	Herbal Alternative	Denver	Colorado
Gorilla Glue #4	GoGl#4_20	Gorilla Glue #4	Colorado Wellness	Denver	Colorado
Golden Goat	GoGo_19	Golden Goat	The Kind Room	Denver	Colorado
Green Crack	GrCr_2b	Green Crack	Nature's Herbs and Wellness	Garden City	Colorado
Bruce Banner	BrBa_19	Bruce Banner	The Kind Room	Denver	Colorado
Flo	Flo_9	Flo	Best Colorado Meds	Fort Collins	Colorado

Table 3.1. *continued*

Name	ID Code	Accession/ Strain Name	Supplier Origin	City	State
Pineapple Express	PiEx_2	Pineapple Express	Nature's Herbs and Wellness	Garden City	Colorado
Purple Haze	PuHa_22	Purple Haze	Lucy Sky	Denver	Colorado
White Widow	WhWi_1	White Widow	Smokey's 420	Garden City	Colorado
Jack Herer	JaHe_12	Jack Herer	The Milkman	SLO	California
OG Kush	OGKu_21	OG Kush	Herbal Alternative	Denver	Colorado
Blue Dream	BlDr_19	Blue Dream	The Kind Room	Denver	Colorado
Tahoe OG	TaOG_11	Tahoe OG	KindCare	Fort Collins	Colorado
Chem Dawg	ChDa_8	Chem Dawg	The Station	Boulder	Colorado
Banana Kush	BaKu_2	Banana Kush	Nature's Herbs and Wellness	Garden City	Colorado
Chem Dawg D	ChDaD_19	Chem Dawg D	The Kind Room	Denver	Colorado
Girl Scout Cookie	GSC_14	Girl Scout Cookie	Day & Night	SLO	California
G13	G13_10	G13	Infinite Wellness	Fort Collins	Colorado
Lemon Diesel	LeDi_2	Lemon Diesel	Nature's Herbs and Wellness	Garden City	Colorado
Hash Plant	HaPl_1	Hash Plant	Smokey's 420	Garden City	Colorado
Australian Hash Plant	HaPlAu_1	Australian Hash Plant	Smokey's 420	Garden City	Colorado
Bubba Kush 98	Bub98_19	Bubba Kush	The Kind Room	Denver	Colorado
Mother of Berries	MoBe_2	Mother of Berries	Nature's Herbs and Wellness	Garden City	Colorado
Northern Lights	NoLi_15	Northern Lights	CannaExpress	SLO	California
Grape Ape	GrAp_16	Grape Ape	Slow Burn	Union Gap	Washington
Purple Kush	PuKu_19	Purple Kush	The Kind Room	Denver	Colorado
Toro Bora	ToBo_4	Toro Bora	Cannabis Genomic Research Initiative	Boulder	Colorado

Deoxyribonucleic Acid Extraction

Deoxyribonucleic acid was extracted using a modified CTAB extraction protocol (Doyle 1987) with 0.035- 0.100 grams of dried flower tissue per extraction. The *Cannabis* draft genome from 'Purple Kush' (GenBank accession AGQN00000000.1) was scanned for microsatellite repeat regions using MSATCOMMANDER-1.0.8-beta (Faircloth 2008). Primers were developed *de-novo* flanking thirty microsatellites with 3-6 nucleotide repeat units and optimized for temperature and magnesium concentration. One primer in each pair was tagged with a 5' universal sequence (M13, CAGT or T7) so that a matching sequence with a fluorochrome tag could be incorporated via PCR following the protocol of Schwabe et al. (2013). Microsatellite primers were optimized (Table 2.4) (Schwabe and McGlaughlin 2018), and ten loci were amplified and analyzed using the Microsatellite Analysis External Plugin ver. 1.4.5 (Biomatters Ltd.) in GENEIOUS ver. 8.1.8 (Biomatters Ltd.).

Statistical Analyses

GENALEX ver. 6.4.1 (Peakall and Smouse 2006; Peakall and Smouse 2012) was used to calculate pairwise genetic differentiation (F_{ST}) and Nei's genetic distance (D) between each of the six groups. PCoA eigenvalues calculated in GENALEX were used to plot the PCoA in RStudio with the ggplot package (R Studio Team 2015) with 95% confidence intervals ellipses. GENALEX was also used to generate a pairwise genetic distance square matrix which was then used to generate a hierarchical cluster analysis dendrogram with Ward's method and Euclidean Genetic distance parameters in PC-ORD (McCune and Mefford 1999).

Genotypes were analyzed using the Bayesian cluster analysis program STRUCTURE ver. 2.4.2 (Pritchard et al. 2000). Burn-in and run-lengths of 50,000 generations were used with ten independent replicates for each STRUCTURE analysis. The number of genetic groups for the data set was determined by STRUCTURE HARVESTER (Earl and vonHoldt 2012), which implements the Evanno et al. method (Evanno et al. 2005).

Maverick v1.0.5 (Verity and Nichols 2016) was used as an additional verification of Bayesian clustering analysis using thermodynamic integration to determine the appropriate number of genetic groups. The following parameters were used: admixture parameter (alpha) of 0.03 with a standard deviation (alphaPropSD) of 0.008, 10 replicates (mainRepeats), 1,000 Burn-in iterations (mainBurnin), 5,000 sample iterations (mainRepeats), 100 TI rungs (thermodynamicRungs), 500 TI Burn-in iterations (thermodynamicBurnin), and 1,000 TI iterations (thermodynamicSamples).

EDENetworks ver. 2.18 (Kivela et al. 2015) was used to construct a web of genetic relationships using the Linear Manhattan distance measure. Auxiliary data were imported to maintain the spatial coordinates and to color individuals by group assignment. The automatic percolation threshold was first derived, and threshold was set to 8.1. Networks were generated for subsequent iterative threshold intervals of 0.5. Increasing the threshold lowers the stringency for genetic relationships, and as the threshold increases, more relationships are formed in the network. EDENetworks diagrams were constructed for the percolation threshold of 8.1, 8.5, 13.7 and 16.9. These are the values that connect: NIDA samples to each other, but not to any other samples in the data set (8.5), connect a single NIDA sample to the larger network (13.7), and finally

connect all samples in the network (16.9). The size of each node is proportionate to the number of relationship connections to other members in the network. The line color and width indicated the strength of the relationship between two individuals- lighter thicker lines indicate stronger genetic relationships, while the darker thinner lines indicate weaker genetic relationships.

Results

Our analyses examined the genetic differentiation and structure of samples from six groups: 1) **NIDA** – research grade marijuana samples obtained from NIDA classified as High THC or High THC/CBD; 2) **Hemp** – *Cannabis* obtained from hemp cultivators and feral collected hemp; 3) **High CBD** – drug-type *Cannabis* with relatively high levels of CBD and low levels of THC; and commercially available drug-type *Cannabis* described as 4) **Sativa**, 5) **Hybrid**, or 6) **Indica** strains. Analyses were also performed on samples at the individual level to control for biases that might arise due to the potential artificial nature of named groups and varying group sample sizes.

Genetic Differentiation

Pairwise genetic differentiation (Fst and Nei's D) calculated in GENALEX ver. 6.4.1 (Peakall & Smouse 2006, Peakall & Smouse 2012) found the highest level of divergence between hemp and high CBD drug-type strains (Fst = 0.215) and between hemp and Sativa drug-type strains (Nei's D = 0.614) (Table 8). The least divergence was observed among the drug-type strains (Fst = 0.023-0.04; Nei's D = 0.66-0.109) (Table 3.2).

Table 3.2. Genetic Differentiation. Pairwise Fst values (below the diagonal) and Nei's D (above the diagonal) for major *Cannabis* groups.

	NIDA	Hemp	High CBD	Sativa	Hybrid	Indica
NIDA		0.519	0.527	0.553	0.48	0.441
Hemp	0.120		0.489	0.614	0.585	0.459
High CBD	0.166	0.215		0.329	0.310	0.281
Sativa	0.114	0.160	0.137		0.098	0.109
Hybrid	0.117	0.149	0.135	0.04		0.066
Indica	0.078	0.124	0.121	0.035	0.023	

Clustering Analysis

Principal Coordinate Analysis (PCoA) was conducted in GENALEX and plotted in R Studio with the ggplot package (R Studio Team 2015) with 95% confidence interval ellipses around the major groups (Figure 8). No confidence intervals were drawn for NIDA (n = 2) or High CBD (n = 3) due to small sample size. Coordinate 1 explains 13.26% of the genetic variation and an additional 11.39% of the genetic variation is explained by coordinate 2. The drug-type strains (Indica, Sativa, Hybrid, and High CBD) all occupy the same character space. There is clear separation of hemp samples from the drug-types, with NIDA samples clustering within the hemp confidence interval.

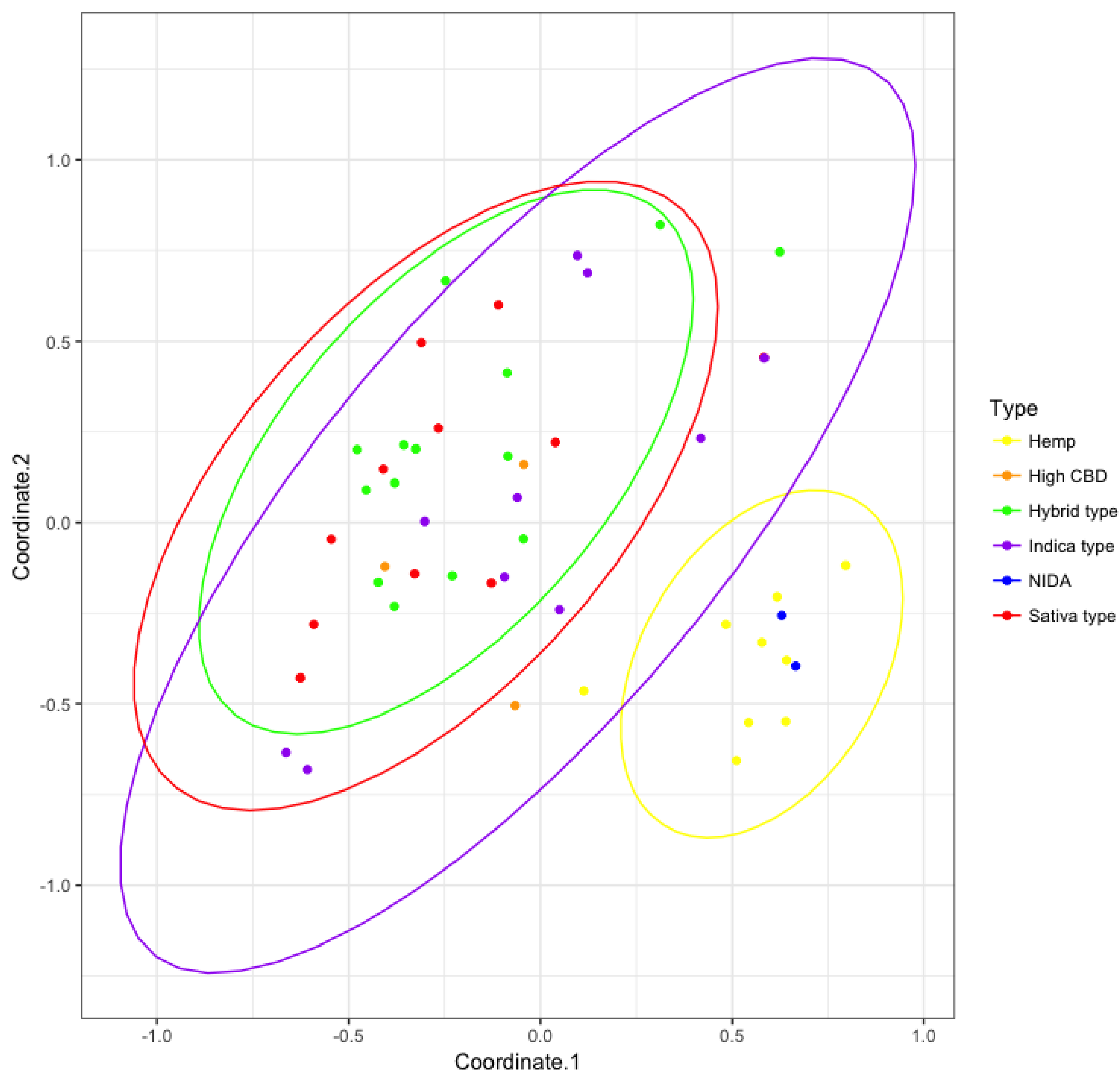


Figure 3.1. Principal Coordinates Analysis. 95% confidence intervals are around the major groups (hemp = yellow, NIDA = blue, High CBD = orange, Sativa = red, Hybrid = green, Indica = purple). Approximately 25% of the genetic variation in these groups is shown (coordinate 1= 13.26% and coordinate 2 = 11.39%). No confidence intervals were drawn for NIDA or High CBD samples due to the small sample size (n = 2 and n = 3, respectively).

PC-Ord version 6 (McCune and Mefford 1999) was used to generate a dendrogram with Ward's method and Euclidean Genetic distance parameters based on pairwise genetic distance values generated in GENALEX (Figure 3.2). The initial branching split the samples into two clusters, A and B. Cluster A contains all but one

hemp sample (88%), as well as the NIDA samples (100%) and two drug-type samples (5%). Cluster B contains the remaining drug-type samples (95%) and one hemp sample (12%). Cluster B further branches into three clusters (C, D, and E), where Sativa, Hybrid and Indica drug type strains are dispersed throughout.

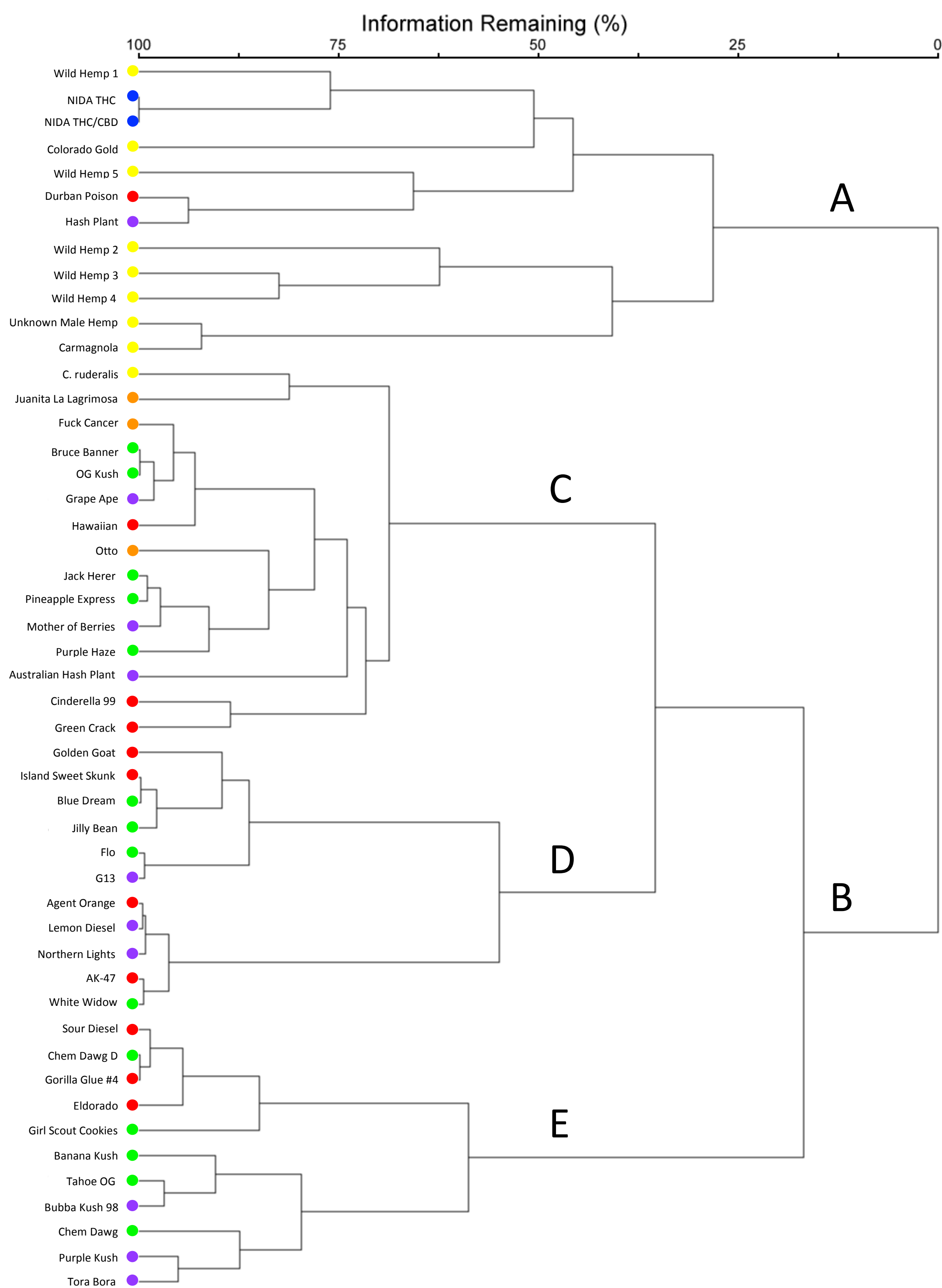


Figure 3.2. PC-Ord group linkage dendrogram. Samples are color-coded (Hemp = yellow, NIDA = blue, High CBD = orange, Sativa = red, Hybrid = green, Indica = purple).

STRUCTURE ver. 2.4.2 (Pritchard et al. 2000) was used to examine sample assignment to genetic groups while allowing admixture. The appropriate number of STRUCTURE groups was validated using STRUCTURE HARVESTER (Earl and vonHoldt 2012), which had high support two genetic groups (K = 2, ΔK = 67.68) and weak support for three genetic groups (K = 2, ΔK = 4.48) (Figure 3.3).

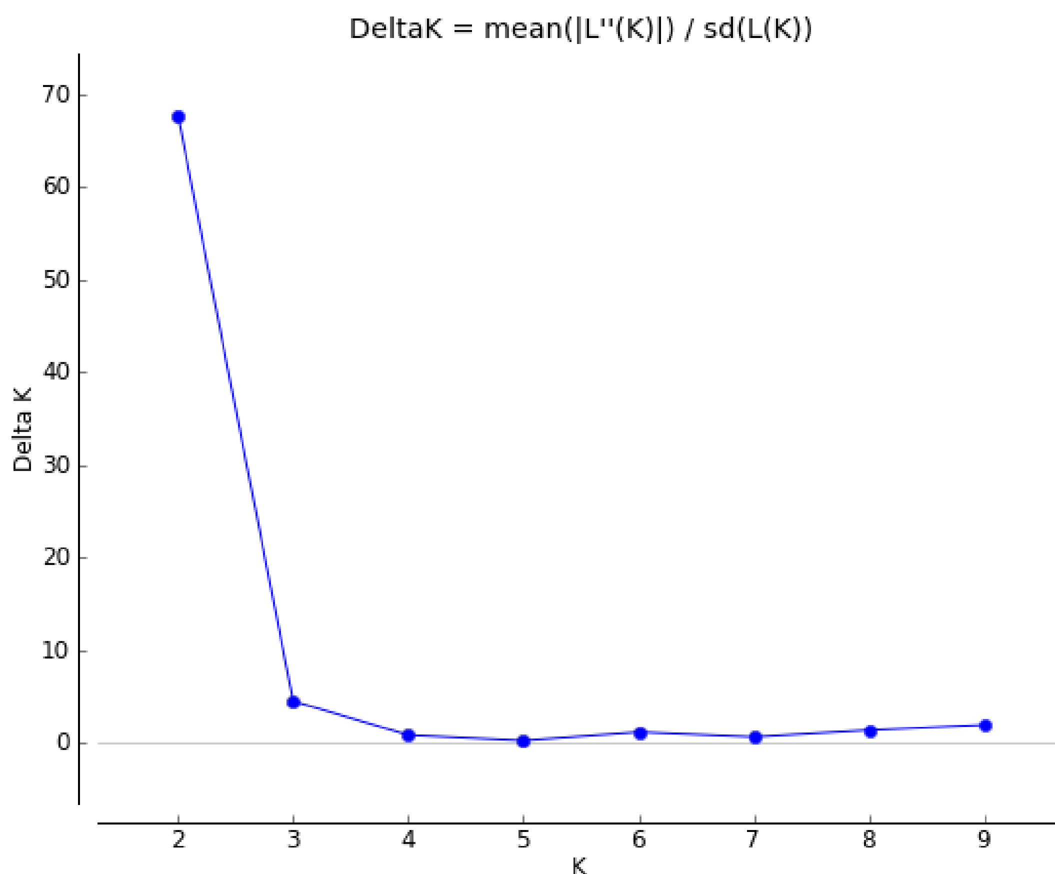


Figure 3.3. STRUCTURE HARVESTER graph showing high support for two genetic groups (K = 2, ΔK = 67.68). There is weak support for three genetic groups (K = 2, ΔK = 4.48).

Additionally, MavericK 1.0.5 (Verity and Nichols 2016) was used to independently test group assignments, which also had strong support for two genetic groups (K = 2,

probability 0.901) and weaker support for three genetic groups (K = 3, probability 0.097) (Figure 3.4), with the sample assignments matching STRUCTURE.

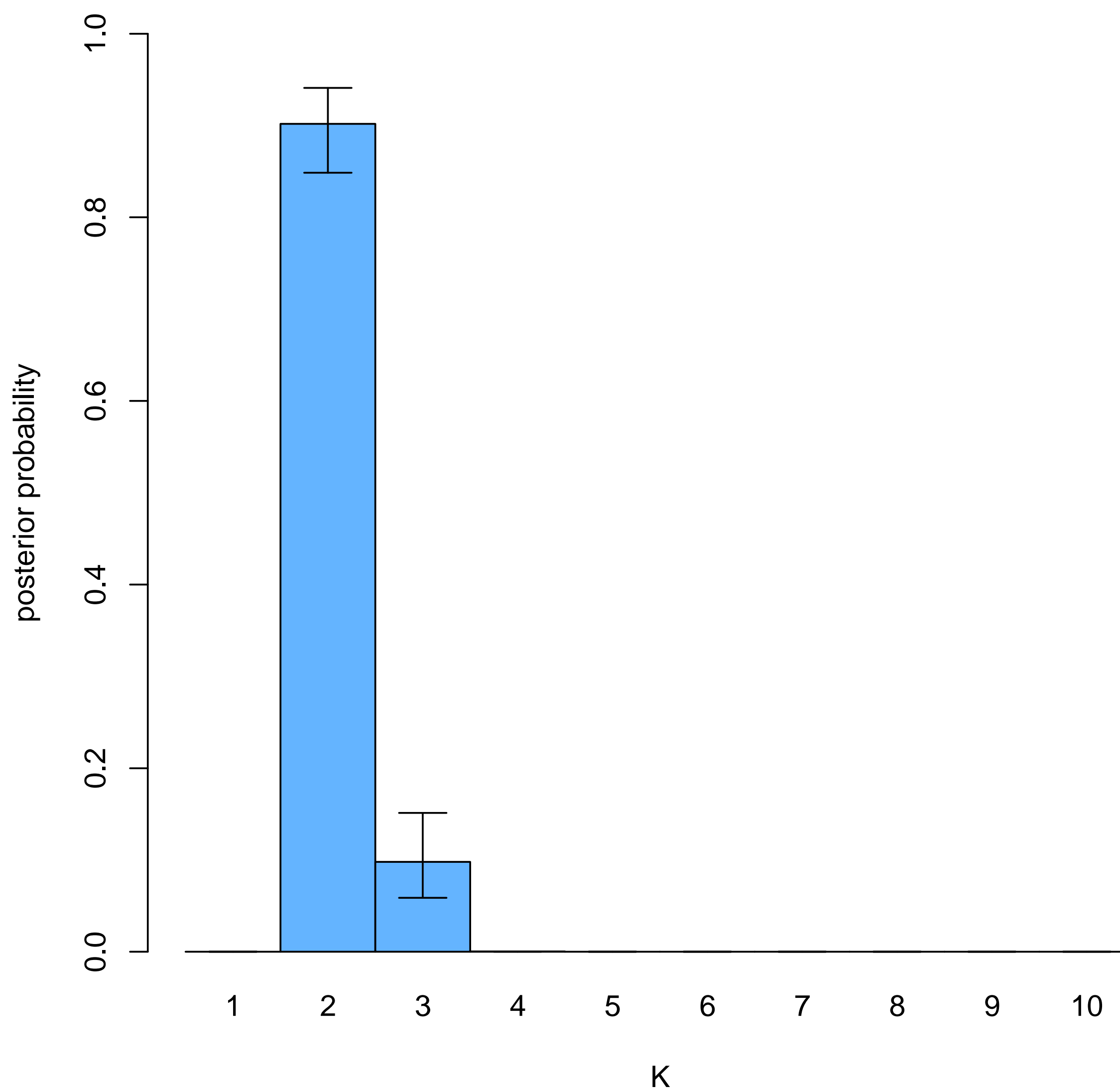


Figure 3.4. MavericK 1.0.5 thermodynamic integration evidence estimates normalized to a sum of 1.0.

The two genetic group STRUCTURE analyses (Figure 3.5) show consistent differentiation between hemp and drug-type strains. All hemp samples were assigned to genetic group 1 (yellow) with a proportion of inferred ancestry (Q) greater than 0.82

(hemp mean group 1, Q = 0.94). Drug-type samples showed some admixture with the majority of the genetic signal of 31 samples (82%) being assigned to genetic group 2 (drug-type mean group 2, Q = 0.72). NIDA samples were assigned to genetic group 1 (NIDA mean group 1, Q = 0.97), demonstrating a strong association with hemp. Although not strongly supported, the three genetic group analysis shows some additional genetic structure among drug-type strains.

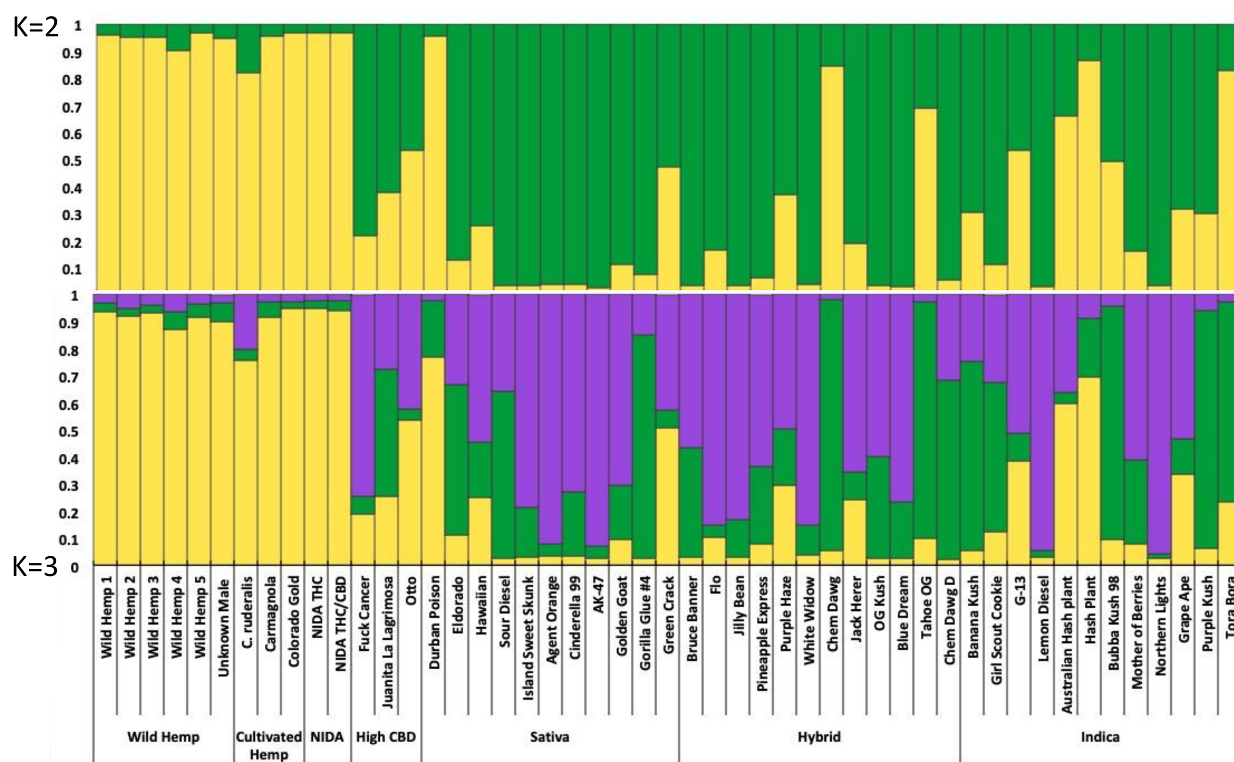


Figure 3.5. Bayesian clustering analysis from STRUCTURE. The proportion of inferred ancestry for two genetic groups (K = 2, top), and for three genetic groups (K = 3, bottom).

Genetic Relatedness Network

EDENetwork ver. 2.18 (Kivela et al. 2015) was used to generate a web of genetic relationship based on pairwise linkages (Figure 3.6). The automatically selected percolation threshold was 8.1 (Figure 3.6A), although not all individuals were connected at this level. The threshold was raised iteratively to connect more divergent samples and explore larger patterns of genetic relationships. The two NIDA samples were united at a threshold of 8.5 (Figure 3.6B). When the threshold was raised to 13.7 (Figure 13C) the NIDA samples become connected to the network via the drug-type sample Eldorado. At a threshold level of 16.9 (Figure 3.6D) all samples in the data set are included in the relationship network.

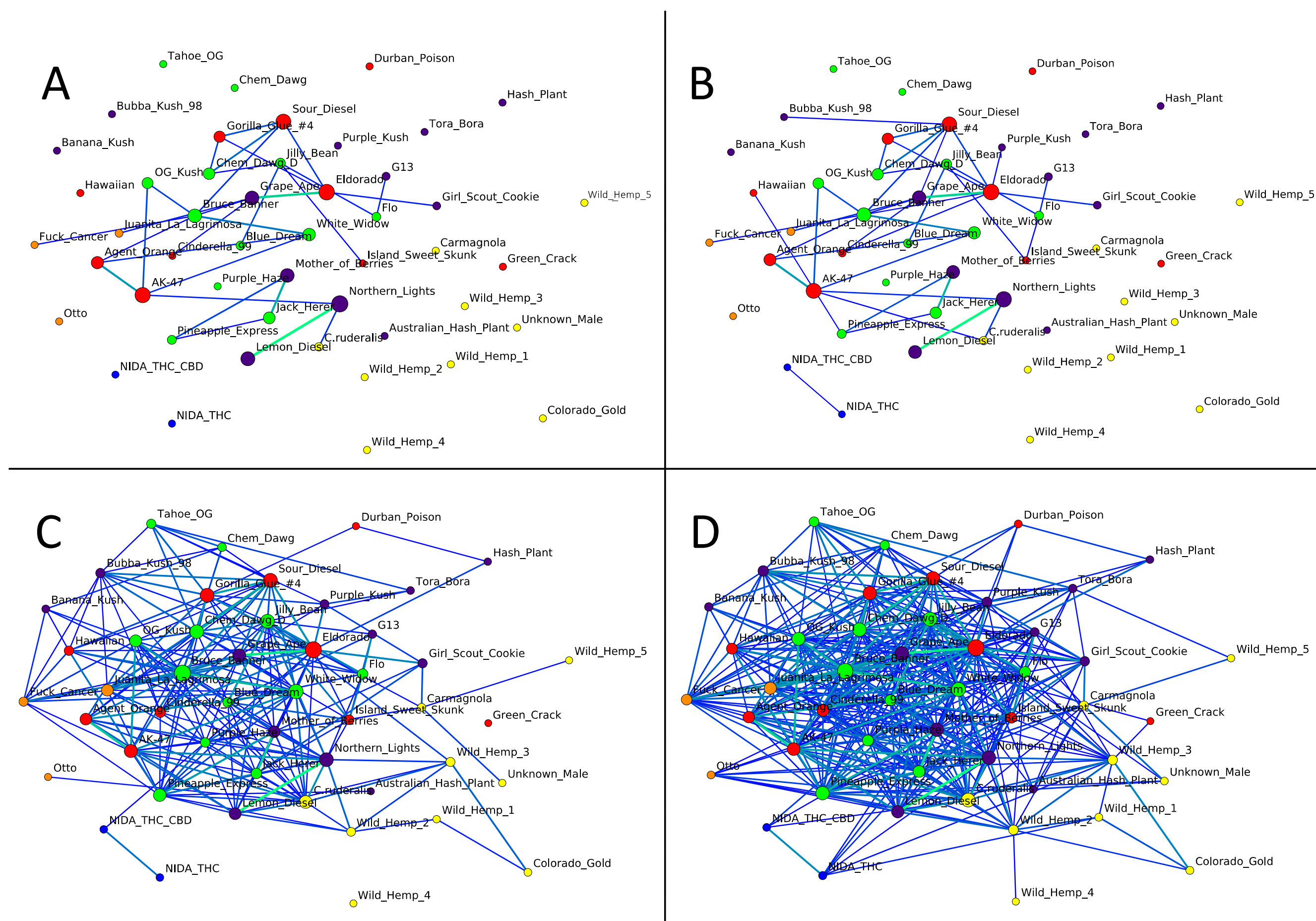


Figure 3.6. EDENetworks genetic relationship network with incrementally decreasing stringency of required genetic relatedness among samples in the data set. (A) Threshold 8.1: the percolation threshold determined by the analysis. (B) Threshold 8.5: the threshold required to connect NIDA samples to each other, but not to any other samples in the data set. (C) Threshold 13.7: the threshold necessary to connect the NIDA sample to the larger network with the connection via the drug-type strain Eldorado. (D) Threshold 16.9: the required threshold to connect all samples in the network. Nodes are colored to indicate group designation (Hemp = yellow, NIDA = blue, High CBD = orange, Sativa = red, Hybrid = green, Indica = purple). Node size is proportionate to the number of connections to that individual within the network. Lines thinner and lighter in color indicate weak genetic relationships, while thicker darker lines indicate stronger relationships.

Discussion

The purpose of this study was to examine the genetic relationship of *Cannabis* samples from the National Institute on Drug Abuse (NIDA) to hemp and drug-type samples. Our results clearly demonstrate that NIDA *Cannabis* samples are substantially different from most commercially available drug-type strains, sharing a genetic affinity with hemp samples in most analyses. Previous research has found that medical and recreational *Cannabis* from California, Colorado and Washington, differs significantly in cannabinoid levels from the research grade marijuana supplied by NIDA (Vergara et al. 2017). Our genetic investigation adds to this previous research, indicating that the genetic makeup of NIDA *Cannabis* is also distinctive from commercially available medical and recreational *Cannabis.*

The genetic data collected in this study indicates that two major genetic groups exist within *Cannabis sativa*. The first group contained a majority of hemp (88 - 100%, depending on analysis) and both NIDA samples (100%), while the second group contained a majority of drug-type samples (82 - 95%). These results contribute to the growing consensus that hemp and drug-type *Cannabis* can be consistently differentiated (Datwyler and Weiblen 2006; Dufresnes et al. 2017; Forapani et al. 2001; Hakki et al. 2007; Lynch et al. 2016; McPartland 2006; Sawler et al. 2015; Soler et al. 2017). This is the first genetic study to include research grade marijuana from NIDA, and its placement with hemp samples was unexpected. However, it is important to note that some drug-type samples (e.g. Durban Poison, Figure 3.2, 3.4) are also placed in the hemp group. Although the sample size of NIDA samples could impact their placement in group-based

analyses such as genetic distances (Table 3.2), all other analyses were carried out at an individual level (Figure 3.1, 3.2, 3.4, 3.5) to avoid this issue.

According to the University of Mississippi National Center for Natural Products Research (NCNPR), which produces research grade marijuana for NIDA, the first experimental plots of *Cannabis* were planted in 1968 with seeds from "Mexico, Panama, Southeast Asia, Korea, India, Afghanistan, Iran, Pakistan, and Lebanon" (Khan 2018; University of Mississippi 2017). Over the next decade, cultivation techniques were standardized, with over 100 varieties planted in 1976 (University of Mississippi 2017). Between the late 1970's and today, the University of Mississippi has continued to be the sole producer of research grade marijuana for NIDA, and it has refined cultivation techniques and extraction procedures, particularly for THC and CBD (Mississippi 2017). The program does not provide variety or strain information when filling *Cannabis* orders, so it is unclear what is currently grown by NCNPR for federally funded marijuana research. The NCNPR director recently stated that "The marijuana project currently stocks 27 plant varieties with different cannabinoid profiles, various CBG potencies, and a wide range of THC levels" (Khan 2018). However, the NCNPR website states that only three *Cannabis* varieties were grown in 2014 (University of Mississippi 2017). Our data suggest that the NIDA *Cannabis* analyzed in this study was sourced from a single strain or two very closely related strains within the NCNPR stock. Without additional information about NCNPR *Cannabis* production, it is difficult to know how many strains are being used in research

This study indicates the need for additional research and refinement of our understanding of *Cannabis* genetic structure and how those difference might impact

Cannabis consumers. Although medicinal research on *Cannabis* has predominantly focused on THC and CBD (Baron 2018; Borgelt et al. 2013; Citti et al. 2018; Cousijn et al. 2018; Maa and Figi 2014; Minkin 2014; National Institute of Health and National Institute on Drug Abuse 2018a, 2018b), it is becoming apparent that other chemical constituents in various combinations and abundances likely have important effects (Baron 2018). If researchers are solely interested in the effects of THC and CBD at know concentrations, then NIDA *Cannabis* could serve as a representative source, although in these cases, isolates of these molecules may be more appropriate. However, given the genetic distinction between NIDA and commercially available *Cannabis,* patients in federally funded *Cannabis* research are likely experiencing effects that are specific to the plant material provided by NIDA. As the interest for medical *Cannabis* increases, it is important that research examining the threats and benefits of *Cannabis* use accurately reflect the experiences of the general public.

Given the rapidly changing landscape of *Cannabis* regulations and consumption (ProCon 2018b), it is not surprising that commercially available *Cannabis* contains a diversity of genetic types. Commercially available *Cannabis* has come to market through non-traditional means leading to many inconsistencies. We have previously documented (Schwabe and McGlaughlin 2018) that there is substantial genetic divergence among samples within named strains, which only exacerbates questions about the impacts of *Cannabis* consumption. This calls to increase regulation and consistency within the *Cannabis* marketplace, and the need for research grade *Cannabis* to accurately represent what consumers have access to.

Conclusion

In conclusion, this study highlights the genetic difference between research grade marijuana provided by NIDA and commercial *Cannabis* available to medical and recreational users. This finding highlights that research conducted with NIDA *Cannabis* may not be indicative of the effects that consumers are experiencing. Additionally, research has found that *Cannabis* distributed by NIDA has lower levels of the principal medicinal cannabinoids (THC and CBD) and higher levels of degradation byproducts of cannabinoids (cannabinol, CBN) (Vergara et al. 2017). Taken together, these results demonstrate the need for there to be greater diversity of *Cannabis* available for medical research and that the genetic provenance of those samples to be established to fully understand the implications of results.

CHAPTER IV

SENSORY VARATION: HUMAN OLFACTORY DISCRIMINATION OF GENETIC VARIATION IN *CANNABIS* STRAINS

Contributions of Authors and Co-Authors

Manuscript in Chapter V

Author: Anna L. Schwabe

Contributions: Conceived the project, provided some funding, recruited participants, collected samples, conducted DNA extractions, designed and optimized microsatellite primers, compiled and analyzed data, and drafted manuscript content.

Author: Samantha K. Naibauer

Contributions: Collected samples, conducted DNA extractions.

Author: Dr. Mitchell E. McGlaughlin

Contributions: Provided some funding, contributed statistical analysis and manuscript revisions.

Author: Dr. Avery N. Gilbert

Contributions: Provided some funding, recruited participants, conducted the sensory data collection, analyzed sensory data, contributed statistical analysis and manuscript revisions

Abstract

There are thousands of *Cannabis* varietals (strains) which are generally described based on psychotropic effects and phytochemical profile. Recent research has found that aroma profiles are distinctive among strains, but also that multiple accessions of the same strain from different sources show genetic inconsistencies. Genetic variation may lead to differences in consumer-relevant phenotypic traits such as terpene content, and therefore differences in aroma. By combining molecular genotyping and olfactory phenotyping techniques, we sought to determine whether genetically inconsistent samples within a commercial strain display inconsistent aroma profiles. We genotyped 42 samples from five strains to determine the consensus genotype as well as genetic outliers (if any) based on 10 variable microsatellite regions. Results were used to select four strains (15 samples) for olfactory testing: “Blue Dream” (5), “OG Kush” (4), “Mob Boss” (3), and “Durban Poison” (3). A genetic outlier sample was included for each strain except “Durban Poison”, which served as a control where all samples had an identical genetic profile. Aroma profiles were produced by 55 untrained sniff panelists (33 men, 22 women) using check-all-that-apply ballots with 40 previously validated odor descriptors. The sensory aroma profile for the “Mob Boss” genetic outlier was at odds with the consensus samples as well as the strain’s previously observed aroma profile. All “OG Kush” samples displayed the strain-typical aroma profile previously described, but the genetic outlier expressed a high-scoring yet atypical “cheese” note. The pungent, chemical, and skunk descriptors were reported far more often in the “Blue Dream” genetic outlier than the for the consensus samples. Although all three samples of “Durban Poison” were genetically identical, the scent profiles do not seem to follow a particular

pattern, which could be due to different growing, curing, storing or age differences among dispensaries. It appears that within-strain differences identified by microsatellite genotyping are associated with differences in aroma profile.

Introduction

Cannabis has been domesticated and cultivated for millennia for fiber, seed and the psychotropic qualities of Δ^9-tetrahydrocannabinol (THC) from the female flower (Clarke and Merlin 2015; Small 2015a; Small 2016, 2017). Recent legalization in many states (National Conference of State Legislatures 2018) and a handful of countries worldwide has created a flourishing retail industry. Marijuana Business Daily estimates retail Cannabis sales in the U.S. may reach as high as $7.3 billion in 2019 (Marijuana Business Daily 2017). Currently, there are thousands of described strains (Leafly 2018b; NCSM 2018; PotGuide.com 2018; Wikileaf 2018) sold as dried flower “buds” that vary in levels of psychotropic cannabinoids and also emit characteristic aromas. The aroma of *Cannabis* is striking and quite unique, and, while some find the odor quite overwhelming and noxious, many people enjoy the aroma and appreciate the subtle nuances among strains. The subtleties underlying the characteristic earthy skunky odor of *Cannabis* often contribute to the multitude of creative strain names. “Sour Diesel” as one can imagine is pungent and possesses a characteristic diesel aroma. Strains with names such as “Cherry Pie”, “Lemon Haze”, “Lavender” and “Banana Kush” lend a suggestion as to their scent.

Aromatic terpene molecules in various combinations and abundances are responsible for creating unique odors associated with *Cannabis*. The terpene profile of each plant is the result of genotype and environmental conditions (Elzinga et al. 2015), is variable over time as the plant matures (Aizpurua-Olaizola et al. 2016) and is also

presumably impacted by differences in curing (drying) techniques, and time in storage. Chemical profile, including terpene profile, has been suggested as a possible mechanism of identification for *Cannabis* cultivars (Casano et al. 2011). Presumably, like other plants such as grapes (de Boubee et al. 2000; Jackson and Lombard 1993) and hops (Patzak et al. 2010; Pavlovic et al. 2012; Sharp et al. 2014), phytochemical production in *Cannabis* is influenced by soil, nutrients, temperature, carbon dioxide, and light, among other environmental factors (Figueiredo et al. 2008). Since chemical profiles change over time, and there are no standard growing conditions yet defined for the *Cannabis* industry, it is unclear if differences in aromatic profiles are due to changes in terpenes over time and/or differences in growing conditions, or perhaps a reflection of genetic variation, or a combination of variables. Steep Hill (Steep Hill Analytics and Research, Berkeley CA) has produced "strain fingerprints" for multiple strains, which are chemical profiles that reportedly characterize ranges of cannabinoids and terpenes specific to strains. The strain fingerprints were initially published on the online strain database Leafly providing details about the levels of seven cannabinoids (THC,CBD, CBN, CBG, THCV, CBC, and CBL) and five terpenes (linalool, β-myrcene, a-pinene, D-limonene, and β-caryophyllene), but they have since been replaced with a set of three icons that describe the dominant flavors of the strain (Leafly 2018b). For example, "Green Crack" flavors are described as earthy, citrus and sweet (Leafly 2018b).

With the legalization of marijuana in many jurisdictions, expert cultivators and connoisseurs are emerging. Cultivators are crossing strains and creating a wide diversity of new strains with an array of aromas and taste profiles. Competitions, such as the Cannabis Cup, allow cultivators to present and connoisseurs to judge aromas, effects, and

quality of novel strains. Retail experts, such as behind-the-counter personnel referred to as “budtenders”, ideally have knowledge about strains and their unique characteristics, especially those in high demand. Scent profiles within a strain should be highly similar, as they are labeled as the same product, presumably with similar genotypes. Several recent genetic studies (Lynch et al. 2016; Sawler et al. 2015; Schwabe and McGlaughlin 2018; Soler et al. 2017) have found genetic differences among samples within *Cannabis* strains, which is interesting as *Cannabis* strains are often produced through cloning methods. Cloning propagation in the legal *Cannabis* industry is often preferred over seed germination for several reasons, arguably the most important being the ability to produce consistent products for consumers. Although genetic variation can result in phenotypic variation, the extent to which genetics might play a role in *Cannabis* strain aroma was previously unknown. Similarities in phenotype, including scent profile, may be a factor leading to misidentification and could be one reason why variation in genetic profiles has been found within strains.

Previous work has identified two dominant aroma groups among a small number of retail strains (Gilbert and DiVerdi 2018). These two groups were described as Cluster A with earthy, woody and herbal aromas, and Cluster B with citrus, lemon, sweet and pungent aromas (Gilbert and DiVerdi 2018). This study included duplicate samples of two strains, “Durban Poison” and “G13”, and both accessions of “G13” fell within Cluster A, while both accessions of “Durban Poison” fell in Cluster B. The “Durban Poison” samples were purchased from different dispensaries and were separated from one another in the cluster. The two “G13” samples were purchased from the same dispensary and had the same harvest date. These samples grouped together as having highly similar

profiles in Cluster A. The researchers were investigating olfactory lexicons to characterize *Cannabis* strains available in the recreational market, but the results raised some interesting questions in light of recent genetic research.

We wondered if genetically anomalous samples labeled as the same strain would have different detectable odors from those that were genetically cohesive. We purposefully identified genetic outliers in a set of otherwise genetically cohesive samples obtained from multiple sources to determine if genetic anomalies are detected through validated sensory methods. In order to assess if genetic anomalies within a strain have different aromas, molecular genotyping and olfactory phenotyping techniques were combined in a two-part study. Based on previous research (Schwabe and McGlaughlin 2018), we determined that in order to maximize our chances of capturing both a consensus and outlying genetic profile, 6-10 samples needed to be collected from different retail facilities. Five strains were chosen based on reported availability at dispensaries and aromatic profile clusters previously described by Gilbert and DiVerdi (2018). Forty-two samples from five strains were genotyped using ten previously published variable short repeating regions of DNA (microsatellites) (Schwabe and McGlaughlin 2018) to determine the consensus genotype and find genetic outliers (if any). For the sensory portion, we predetermined that 15 samples were needed in order to obtain reliable results and not present too many samples in a single setting. These two combined studies aim to demonstrate whether or not genetic anomalies within strains have different odors from samples with a highly similar genotype.

Methods

Genetic Methods

Strain selection. Recreational dispensary strain information was researched online (Weedmaps 2018) to determine which strains were most likely to be available from multiple sources. Online scent profiles (Leafly 2018b) were examined to select a subset of strains with reportedly unique scent profiles in order to minimize aromatic similarity of the strains to include in the olfactory analysis. Fifteen strains were cross-referenced with the results from Gilbert and DiVerdi (2018), and five strains were chosen based on reported availability at dispensaries and aromatic profile (Table 4.1). “Durban Poison” was selected as having an aromatic profile representative of the citrus, lemon, sweet, pungent group (Cluster B) (Gilbert and DiVerdi 2018). “OG Kush” is reportedly unique in both genetic and aromatic profiles (Elzinga et al. 2015; Gilbert and DiVerdi 2018; Leafly 2018b) and was selected as having an aromatic profile representative of the earthy, woody, herbal group (Cluster A) (Gilbert and DiVerdi 2018). “Sour Diesel” was chosen as it was previously observed to have aromatic properties of both groups (Clusters A and B) (Gilbert and DiVerdi 2018) and is described online as earthy, pungent, and diesel (Leafly 2018b). “Blue Dream” had not previously been analyzed for aromatic profile. However, because the online odor descriptors of berry, blueberry, and sweet (Leafly 2018b) indicated this strain might be unique compared to the others, it was chosen for inclusion in the olfactory perception analysis. During sample acquisition, “Mob Boss” was available at many locations, so it was collected in addition to the previously chosen strains. Gilbert and DiVerdi (2018) placed “Mob Boss” in the earthy, woody, herbal group (Cluster A) (Gilbert and DiVerdi 2018). Both “Mob Boss” and “OG

Kush" have the same earthy, pine, and woody descriptors described online (Leafly 2018b).

Table 4.1. The number of samples of five selected *Cannabis* strains. The Scent Cluster assignment (Gilbert & DiVerdi 2018), and the Leafly database scent profile (Leafly 2018b).

Strain	Sample number	Scent Cluster	Leafly
Durban Poison	8	B	earthy, pine, sweet
OG Kush	8	A	earthy, woody, pine
Sour Diesel	10	A/B	diesel, pungent, earthy
Blue Dream	10	unknown	sweet, berry, blueberry
Mob Boss	6	A	pine, pungent, sweet

(A) earthy, woody, herbal (B) citrus, lemon, sweet, pungent

Genetic material. A total of 42 retail *Cannabis* samples were purchased from 25 recreational *Cannabis* dispensaries in six Colorado cities (Table 4.2). The names for each dispensary have been withheld to protect the identity of businesses where genotypes may deviate from the norm. The locations of the dispensaries in this experiment were chosen based solely on the availability of strains. A minimum of six samples of each strain were collected. All samples were purchased legally over-the-counter. The dispensary weighed 2 grams of each sample, and these samples are labeled as 'SN', indicating eligibility for the scent analysis in the olfactory portion of this study. Additional samples were added to the genetic study, labeled as 'GN', in order to capture the variation contained in a strain, without the intention of using them in the olfactory portion of this study. Purchase receipts and original packaging labels were retained for reference.

Table 4.2. Information for all samples included in the genetic portion of this study.

Strain	Location	Date Acquired	Sample ID
Durban Poison	Denver	Aug 7 2018	DuPo_1SN
	Denver	Aug 7 2018	DuPo_2SN
	Denver	Aug 7 2018	DuPo_3SN
	Denver	Aug 7 2018	DuPo_4SN
	Fort Collins	Aug 6 2018	DuPo_5SN
	Garden City	Aug 7 2018	DuPo_6SN
	Breckenridge	May 9 2018	DuPo_7GN
	Garden City	April 28 2018	DuPo_8GN
OG Kush	Denver	Aug 7 2018	OGKu_1SN*
	Denver	Aug 7 2018	OGKu_2SN
	Fort Collins	Aug 6 2018	OGKu_3SN
	Denver	Aug 7 2018	OGKu_4SN*
	Denver	Aug 9 2018	OGKu_5SN*
	Denver	Aug 9 2018	OGKu_6SN
	Denver	Aug 9 2018	OGKu_7SN
	Denver	Aug 7 2018	OGKu_8GN
Sour Diesel	Denver	Aug 7 2018	SoDi_1SN
	Denver	Aug 7 2018	SoDi_2SN
	Denver	Aug 7 2018	SoDi_3SN
	Fort Collins	Aug 6 2018	SoDi_4SN
	Garden City	Aug 6 2018	SoDi_5SN
	Garden City	Aug 7 2018	SoDi_6SN
	Denver	Aug 7 2018	SoDi_7SN
	Frisco	May 9 2018	SoDi_8GN
	Breckenridge	May 9 2018	SoDi_9GN
	Frisco	May 9 2018	SoDi_10SN
Blue Dream	Denver	Aug 7 2018	BlDr_1SN
	Denver	Aug 7 2018	BlDr_2SN
	Denver	Aug 7 2018	BlDr_3SN
	Denver	Aug 7 2018	BlDr_4SN
	Fort Collins	Aug 6 2018	BlDr_5SN
	Garden City	Aug 6 2018	BlDr_6SN
	Denver	Aug 9 2018	BlDr_7SN
	Breckenridge	May 9 2018	BlDr_8GN
	Frisco	May 9 2018	BlDr_9GN
	Breckenridge	May 9 2018	BlDr_10GN
Mob Boss	Denver	Aug 7 2018	MoBo_1SN
	Denver	Aug 7 2018	MoBo_2SN
	Denver	Aug 7 2018	MoBo_3SN
	Denver	Aug 7 2018	MoBo_4SN
	Fort Collins	Aug 7 2018	MoBo_5SN
	Boulder	Aug 8 2018	MoBo_6SN

*Labeled with the same grower and same lot number

Deoxyribonucleic acid extraction, Polymerase Chain Reaction and fragment analysis. Deoxyribonucleic acid was extracted using a modified CTAB extraction protocol (Doyle 1999) with 0.035-0.100 g of dried flower tissue per extraction. Ten primers developed *de-novo* from the 'Purple Kush' genome were used to amplify DNA fragments containing variable microsatellite regions as described in Chapter II (Schwabe and McGlaughlin 2018).

Genetic statistical analysis. GENALEX ver. 6.4.1 (Peakall and Smouse 2006; Peakall and Smouse 2012) was used to calculate Lynch & Ritland (Lynch and Ritland 1999) mean pairwise genetic relatedness (*r*) within each strain. A genetic pairwise relatedness heat map for each strain was generated in Microsoft EXCEL. Samples with identical genotypes share 100% genetic identity (*r*-value = 1.00), first order relatives (full siblings or mother-daughter) share 50% genetic identity (*r*-value = 0.50), second order relatives (half siblings or cousins) share 25% genetic identity (*r*-value = 0.25), and unrelated individuals are expected to have an *r*-value of 0.00 or lower. PCoA eigenvalues were calculated in GENALEX and plotted in RStudio (R Studio Team 2015) with the ggplot package (Wickham 2016) with 95% confidence interval ellipses. GENALEX was also used to generate a pairwise genetic distance square matrix to generate a hierarchical cluster analysis dendrogram with Ward's method and Euclidean Genetic distance parameters in PC-ORD (McCune and Mefford 1999). Genotypes were analyzed using the Bayesian cluster analysis program STRUCTURE ver. 2.4.2 (Pritchard et al. 2000). Burn-in and run-lengths of 100,000 generations were used with ten independent replicates for each STRUCTURE analysis. STRUCTURE HARVESTER (Earl and vonHoldt 2012), which implements the Evanno et al. (2005) method, was used

to determine the *K* value that best describes the number of genetic groups for the data set. Missing data for seven of ten "Blue Dream" samples at one locus did not change the results when removed, therefore all analyses in GENALEX and STRUCTURE were conducted using ten loci.

Sample selection. The samples included in the sensory portion of this study were chosen based on the results of the genetic analysis (see Genetic Results section). The "Sour Diesel" samples obtained for this study did not have enough genetic variation in any of the samples to be considered for the sensory portion. The selected samples were assigned a random identification number for the double-blind olfactory study (Table 4.3). Neither the study conductor nor the participant was provided information to disclose the name of the strain.

Table 4.3. Samples used in the olfactory study. The Sample ID is included with genetic outliers identified with and asterisk and the random Sample Code.

Strain	Sample ID	Sample Code
Durban Poison	DuPo_1SN	245
	DuPo_4SN	351
	DuPo_5SN	403
OG Kush	OGKu_1SN*	584
	OGKu_2SN	752
	OGKu_3SN	781
	OGKu_4SN	437
Blue Dream	BlDr_1SN	925
	BlDr_3SN*	116
	BlDr_4SN	700
	BlDr_5SN	307
	BlDr_6SN	312
Mob Boss	MoBo_1SN	187
	MoBo_3SN	482
	MoBo_5SN*	659

Sensory Methods

Odor Stimuli

Odor stimuli consisted of 15 cannabis samples drawn from four strains: “Durban Poison” (3), “OG Kush” (4), “Blue Dream” (5), “Mob Boss” (3). Each stimulus (1 g of dried cannabis flower) was presented in a wide mouth 118 mL (4 oz) amber glass bottle labeled with a three-digit code (Figure 4.1). Samples were kept in a freezer at -2° C and thawed at room temperature for two hours before testing. The stimuli were exchanged for fresh samples midway through the study. Our use of strain designations provided by the retail dispensaries was a matter of convenience; it does not imply a position regarding the taxonomic validity or botanical derivation of these strains. Our goal was to characterize olfactory variation in commercially available offerings.

Figure 4.1. Samples (1 gram) of dried *Cannabis* flower. Samples were presented to participants in wide mouth 118 mL (4 oz) amber glass bottles labeled with a three-digit code.

Odor Descriptors

Forty odor descriptors were chosen from online sources that describe characteristic scents for strains (Gilbert and DiVerdi 2018; Leafly 2018b) (Table 4.4). Due to the wide variety of strains and descriptive scent characters, the selection aimed to include the majority of previously detected odors in a variety of *Cannabis* strains (Gilbert and DiVerdi 2018).

Table 4.4. The 40 odor descriptors used to characterize the samples in this study in alphabetical order.

Ammonia	Diesel	Mango	Rose
Apricot	Earthy	Menthol	Sage
Berry	Flowery	Mint	Skunk
Blue cheese	Grape	Nutty	Spicy
Butter	Grapefruit	Orange	Sweet
Cheese	Herbal	Peach	Tea
Chemical	Honey	Pepper	Tobacco
Chestnut	Lavender	Pine	Tropical fruit
Citrus	Lemon	Pineapple	Violet
Coffee	Lime	Pungent	Woody

Rating Scales and Presentation

Participants rated each sample using a Check All That Apply (CATA) ballot with 40 descriptors, presented in alphabetical order on a single screen of a touch-screen device (Apple iPad 2). Data were automatically entered into a spreadsheet; scale presentation and data collection were designed using free online services (Google Forms and Google Sheets).

Ethics Statement

This study protocol was approved by the Western Institutional Review Board (Puyallup, Washington) (WIRB Protocol #20170080). All participants provided informed

written consent using a form approved by WIRB. At no time did participants come into direct contact with the *Cannabis* samples. Retail sale of marijuana for recreational use to adults 21 years of age and older has been legal in the state of Colorado since January 1, 2014.

Participants

Test participants were recruited from Fort Collins and vicinity. Participants from a previous study (Gilbert and DiVerdi 2018) who indicated a willingness to participate in further research were re-contacted. A notice (text approved by WIRB) was posted to an online bulletin board for the local community. Printed text emphasized "current, former, and non-users all welcome" and that only sniffing was required ("no touching, no smoking, no eating"). All participants were at least 21 years of age, residents of Colorado, and had a self-reported normal sense of smell. Exclusion criteria included self-reported pregnancy, active nasal allergy, and current head cold. Subjects were paid $20.00 for their participation.

Sensory Statistical Analyses

A Friedman's nonparametric repeated-measures ANOVA was performed on the summed frequencies for each odor descriptor across all samples within a given strain using SPSS Statistics v. 24 (IBM Corp. 2016). A Wilcoxon signed-rank test on all pairwise combinations of samples within each strain (analogous to post-hoc t-tests in ANOVA) was conducted using SPSS Statistics v. 24 (IBM Corp. 2016). Calculations for histograms and tables were conducted using Microsoft Excel.

Within strain scent profiling was analyzed using two measures: Perceived Shared Character Category and *Cannabis* Lexicon Category. For the Perceived Shared Character

Category analyses aroma descriptors were assigned to one of five categories by A. Schwabe. The five Perceived Shared Character Categories were: “Earthy” which included soil, buttery, nutty or roasted aromas, “Spicy” which included spices and dried leafy scents, “Sweet” which included scents associated with fruits, “Floral” scents as fresh plants/flowers, and “Pungent” which included sharp and/or unpleasant aromas (Table 4.5).

Table 4.5. Perceived Shared Character categories. Each of the 40 odor descriptors assigned to one of five categories. The total number of descriptors is in parentheses.

Earthy (7)	Spicy (6)	Sweet (14)	Floral (6)	Pungent (7)
Earthy	Herbal	Apricot	Flowery	Ammonia
Butter	Sage	Berry	Lavender	Skunk
Coffee	Pepper	Citrus	Rose	Cheese
Pine	Spicy	Grape	Violet	Chemical
Woody	Tea	Grapefruit	Mint	Blue cheese
Nutty	Tobacco	Lemon	Menthol	Pungent
Chestnut		Lime		Diesel
		Mango		
		Orange		
		Peach		
		Pineapple		
		Citrus		
		Sweet		
		Tropical fruit		
		Honey		

For the *Cannabis* Lexicon Category analyses, aroma descriptors were assigned to four categories by combining characters of two *Cannabis* lexicons, the Terpene Flavor Wheel © (The Holden Company and Western Cultured 2016) and The Flavor Wheel ™ (Green House Seed Company 2018). The wheels are similar, but a combination of the two was used to capture all 40 scent descriptors used in this study. The four categories in *Cannabis* Lexicon Categories were “Sweet”, “Sour”, “Spicy” and “Bitter” (Table 4.6).

Table 4.6. *Cannabis* Lexicon Categories. Each of the 40 odor descriptors assigned to one of four categories. The total number of descriptors is in parentheses.

Sweet (12)	Sour (8)	Spicy (9)	Bitter (11)
Apricot	Blue cheese	Herbal	Ammonia
Berry	Butter	Lavender	Chemical
Flowery	Cheese	Menthol	Chestnut
Grape	Citrus	Mint	Coffee
Honey	Grapefruit	Pepper	Diesel
Mango	Lemon	Pine	Earthy
Peach	Lime	Sage	Nutty
Pineapple	Orange	Spicy	Pungent
Rose		Woody	Skunk
Sweet			Tea
Tropical fruit			Tobacco
Violet			

The Perceived Shared Character Category and *Cannabis* Lexicon Category data were analyzed separately for the genetically cohesive samples and for the genetic outlier sample. A frequency of detection scale was calculated for each sample by dividing the total number of positive detections by the number of descriptors in the category multiplied by the number of samples in the strain group, multiplied by 55 (the number of possible positive detections) (Equation 1).

$$\frac{\sum \text{positive detections}}{55\,(n_{\text{descriptor}} \times n_{\text{strain samples}})} \qquad \text{Equation 1.}$$

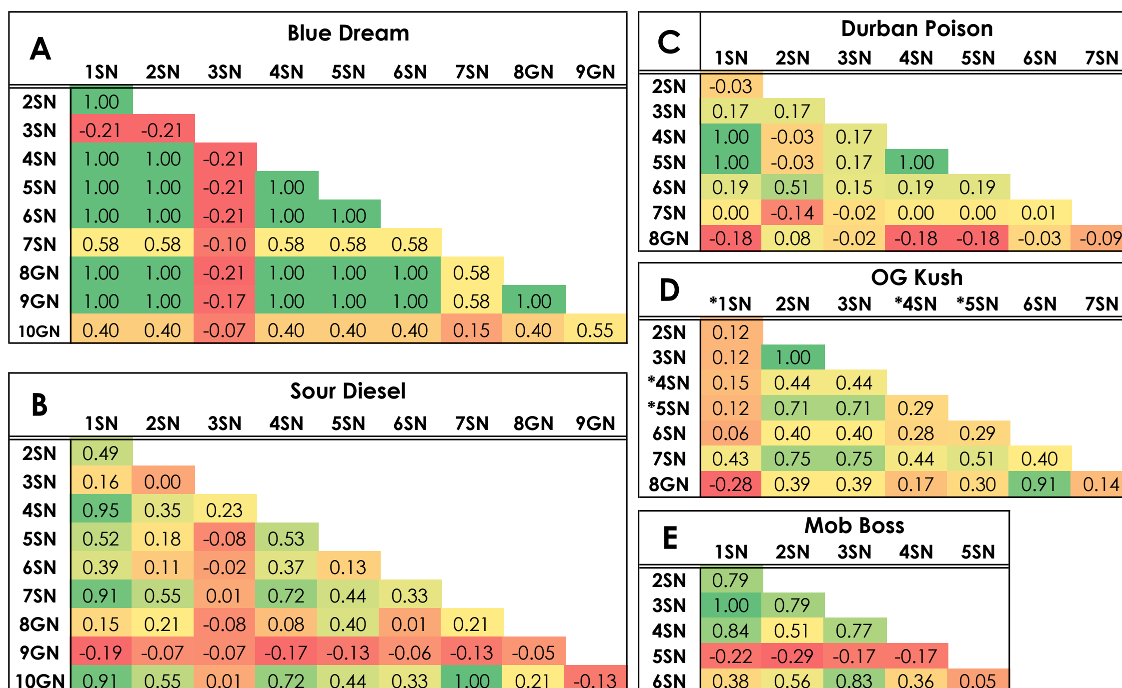
The frequency of detection metric normalizes the data and allows for comparisons across categories with different numbers of scents, as well as across strain groups with different numbers of samples. The range of the frequency of detection scale is 0.00-1.00, where zero means no participants detected any scents in that category, and 1.00 means every participant detected every scent in the category. Given the subjective and personal nature of olfactory ratings, we would not expect to see a value of 1.00. Histograms for the

frequencies were generated in Excel to compare differences in the mean frequencies of the genetic consensus sequences compared and the genetic outlier. As these are frequencies of detection, the data were normalized to demonstrate the scent profile of consensus versus outlier samples in each strain.

Genetic Results

Genetic Relatedness

Lynch & Ritland (Lynch and Ritland 1999) pairwise genetic relatedness within each strain was calculated in GENELEX (Figure 4.2). Values of $r = 1.00$ are indicative of identical individuals as observed in clones. Values of $r < 0$ between two individuals indicate the individuals have a very low level of relatedness. Samples 1SN, 4SN and 5SN in "Durban Poison" were identical ($r = 1.00$) and 8GN was a genetic anomaly compared to the other samples ($r = -0.18$; Figure 4.2C). Samples 2SN and 3SN, and 6SN and 8GN in "OG Kush" had a high level of genetic relatedness ($r = 1.00$ and $r = 0.91$, respectively), and other pairwise relatedness between samples were low to moderate (r =0.06 – 0.75; Figure 4.2D). Samples 1SN, 4SN, 7SN and 10GN in "Sour Diesel" had a high level of genetic relatedness (r =0.91 - 1.00), and the remaining samples had moderate to low genetic relatedness (r =-0.19 - 0.72; Figure 4.2B). Samples 1SN, 2SN, 4SN, 5SN, 6SN, and 8GN in "Blue Dream" were genetically identical ($r = 1.00$), and 3SN had a very low level of relatedness to all identical samples in the set (r = - 0.21; Figure 4.2A). Samples 1SN and 3SN in "Mob Boss" were identical ($r = 1.00$) and 5SN had a very low level of relatedness to all identical samples in the set (r = -0.29 – 0.05; Figure 4.2E).



A — Blue Dream

	1SN	2SN	3SN	4SN	5SN	6SN	7SN	8GN	9GN
2SN	1.00								
3SN	-0.21	-0.21							
4SN	1.00	1.00	-0.21						
5SN	1.00	1.00	-0.21	1.00					
6SN	1.00	1.00	-0.21	1.00	1.00				
7SN	0.58	0.58	-0.10	0.58	0.58	0.58			
8GN	1.00	1.00	-0.21	1.00	1.00	1.00	0.58		
9GN	1.00	1.00	-0.17	1.00	1.00	1.00	0.58	1.00	
10GN	0.40	0.40	-0.07	0.40	0.40	0.40	0.15	0.40	0.55

B — Sour Diesel

	1SN	2SN	3SN	4SN	5SN	6SN	7SN	8GN	9GN
2SN	0.49								
3SN	0.16	0.00							
4SN	0.95	0.35	0.23						
5SN	0.52	0.18	-0.08	0.53					
6SN	0.39	0.11	-0.02	0.37	0.13				
7SN	0.91	0.55	0.01	0.72	0.44	0.33			
8GN	0.15	0.21	-0.08	0.08	0.40	0.01	0.21		
9GN	-0.19	-0.07	-0.07	-0.17	-0.13	-0.06	-0.13	-0.05	
10GN	0.91	0.55	0.01	0.72	0.44	0.33	1.00	0.21	-0.13

C — Durban Poison

	1SN	2SN	3SN	4SN	5SN	6SN	7SN
2SN	-0.03						
3SN	0.17	0.17					
4SN	1.00	-0.03	0.17				
5SN	1.00	-0.03	0.17	1.00			
6SN	0.19	0.51	0.15	0.19	0.19		
7SN	0.00	-0.14	-0.02	0.00	0.00	0.01	
8GN	-0.18	0.08	-0.02	-0.18	-0.18	-0.03	-0.09

D — OG Kush

	*1SN	2SN	3SN	*4SN	*5SN	6SN	7SN
2SN	0.12						
3SN	0.12	1.00					
*4SN	0.15	0.44	0.44				
*5SN	0.12	0.71	0.71	0.29			
6SN	0.06	0.40	0.40	0.28	0.29		
7SN	0.43	0.75	0.75	0.44	0.51	0.40	
8GN	-0.28	0.39	0.39	0.17	0.30	0.91	0.14

E — Mob Boss

	1SN	2SN	3SN	4SN	5SN
2SN	0.79				
3SN	1.00	0.79			
4SN	0.84	0.51	0.77		
5SN	-0.22	-0.29	-0.17	-0.17	
6SN	0.38	0.56	0.83	0.36	0.05

Figure 4.2. Lynch & Ritland pairwise genetic relatedness within each strain. Values of r = 0.50 are indicative identical as observed in clones, r < 0 indicates a low level of genetic relatedness. "OG Kush" samples 1SN, 4SN and 5SN (asterisk) were labeled with the same Retail Marijuana Cultivation Facility and the same lot number even though they were purchased from different dispensaries.

Clustering Analyses

Principal Coordinates Analysis (Figure 4.3) was conducted in GENALEX and plotted using the ggplot package in R Studio with 95% confidence interval ellipses around the major groups (Figure 4.3) (R Studio Team 2015). The samples that fell outside the confidence intervals, "OG Kush" 1 SN, "Blue Dream" 3SN, and "Mob Boss" 5SN were considered genetic outliers (Figure 4.3).

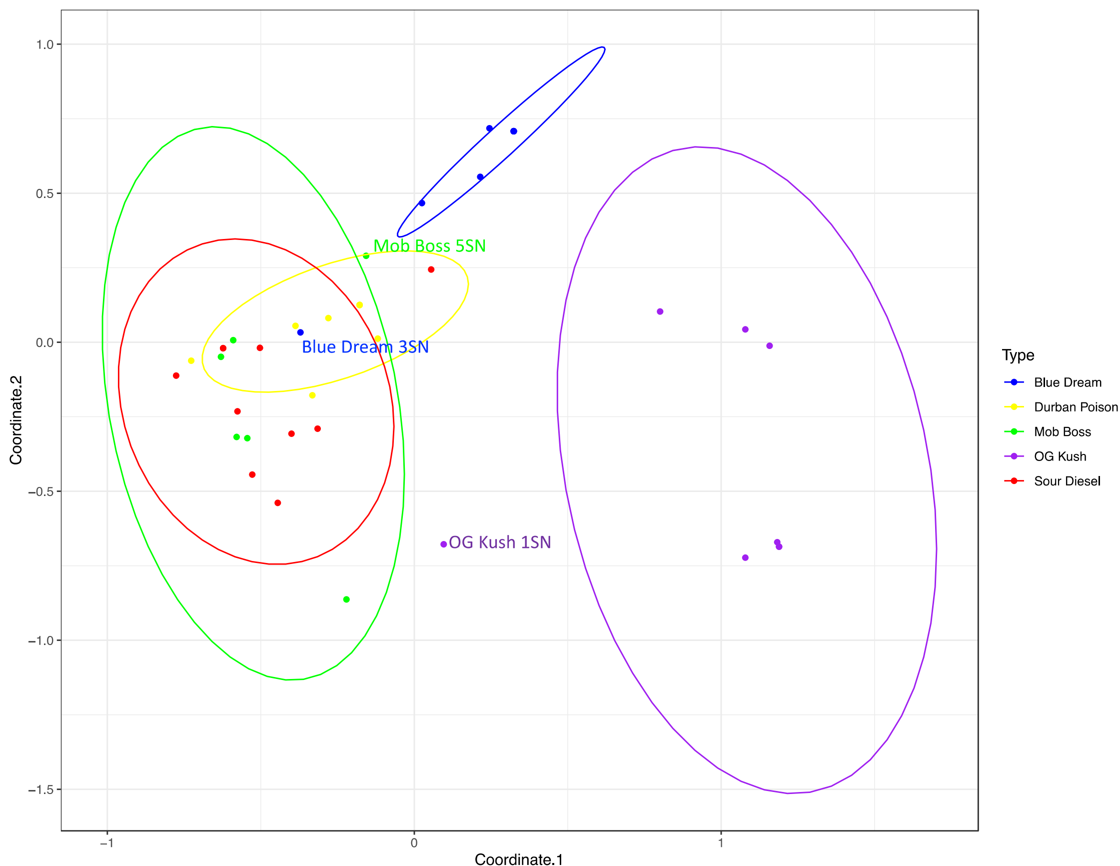


Figure 4.3. Principal Coordinate Analysis generated using ggplot in R Studio (R Studio Team 2015). The samples are colored by strain name, and 95% confidence interval ellipses are drawn around each cluster.

STRUCTURE ver. 2.4.2 (Pritchard et al. 2000) was used to examine individual assignment to genetic groups. STRUCTURE HARVESTER (Earl and vonHoldt 2012) calculated the appropriate number of STRUCTURE groups using the Evanno method (Evanno et al. 2005). This data set had extremely high support for three genetic groups ($K = 3$, $\Delta K = 216.07$) and weak support for two or six genetic groups (Figure 4.4).

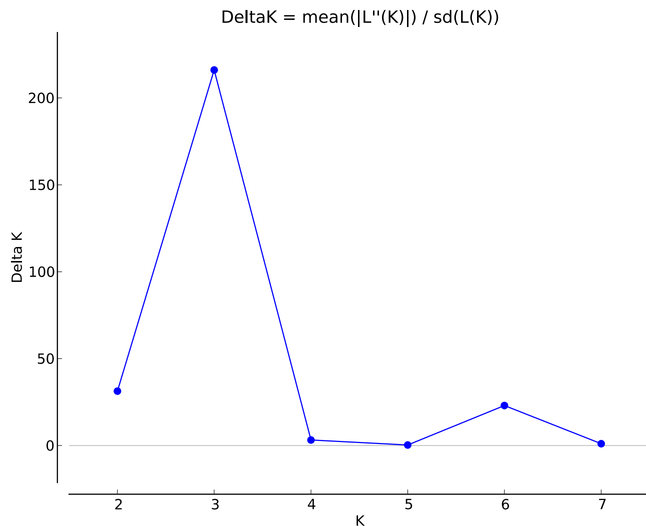


Figure 4.4. STRUCTURE HARVESTER output as calculated using the Evanno method, showing robust support for three genetic groups (K = 3) in this data set (Δ K = 216.07).

The three groups represented in STRUCTURE are color coded as blue (group 1), green (group 2) and yellow (group 3) (Figure 4.5). All but one sample of "OG Kush" was comprised largely of group 1 genetic assignment (blue, 87.1- 98.9 %); sample 1SN had only 20.5 % group 1 genetic assignment and 78.1 % assignment to group 3 (yellow). All but one sample of "Blue Dream" was largely assigned to group 2 (green, 91.2 - 98.6 %); sample 3SN had a 98 % assignment to group 3 (yellow). The remaining samples from "Durban Poison", "Mob Boss", and "Sour Diesel" were assigned to group 3 (66.8 – 98.8 %). "Durban Poison" 8GN was assigned to group 2 (green, 74.8 %), but there was not enough sample for the sensory study. However, "Durban Poison" 2SN and "Mob Boss" 5SN had relatively low assignment to group 3 (66.8 % and 69.5 % respectively) and were considered in this analysis to be comparative outliers. All samples of "Sour Diesel" were assigned to group 3 (yellow, > 95%).

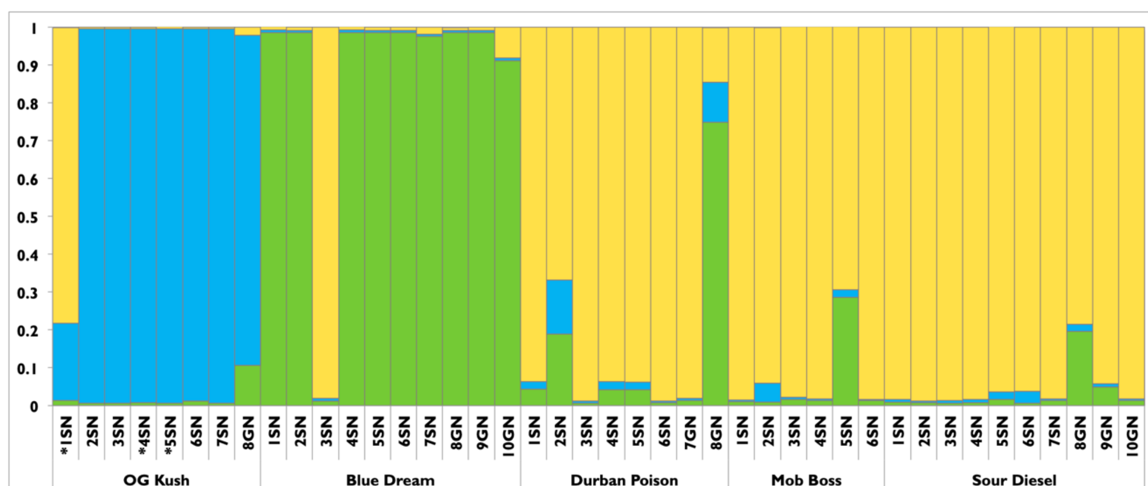


Figure 4.5. STRUCTURE graph with the proportion of genetic assignment to each of the three genetic groups as indicated by the proportion of each color in each bar representing an individual. "OG Kush" samples 1SN, 4SN and 5SN (asterisk) were labeled with the same Retail Marijuana Cultivation Facility and the same lot number even though they were purchased from different dispensaries.

Two dendrograms were created based on pairwise genetic distance values labeled with the sample names and color coded by strain (Figures 4.6, 4.7). The analysis of all 42 samples (Figure 4.6) showed an initial split of "OG Kush" from the remaining samples. Within the remaining strains there was clear groups consisting of seven "Blue Dream", four "Mob Boss", three "Durban Poison", and six "Sour Diesel" samples.

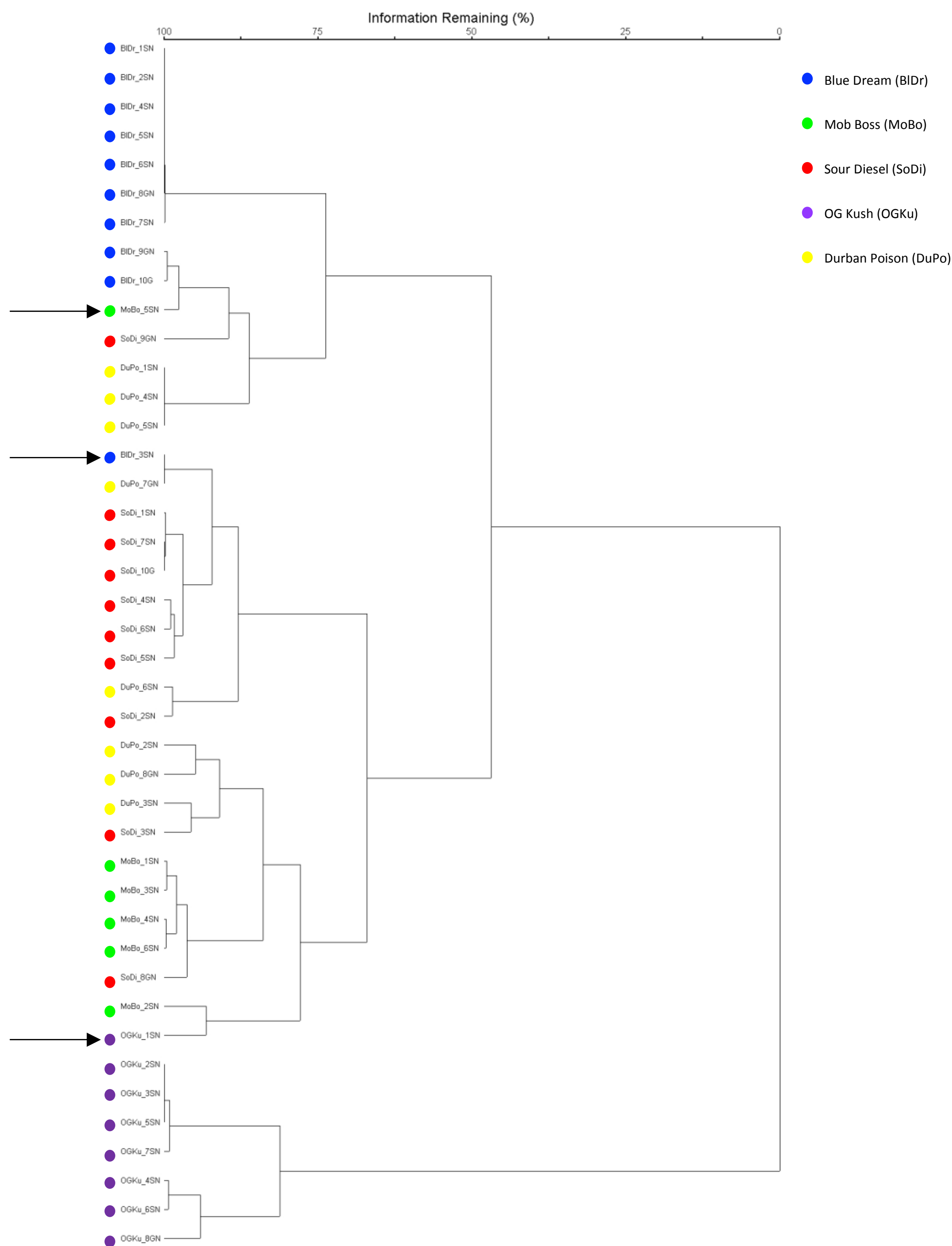


Figure 4.6. PC-Ord hierarchical genetic cluster analysis based on genetic distance color coded by strain. Clear genetic outliers assigned to a conflicting cluster are indicated by the arrows. “Durban Poison” and “Sour Diesel” samples span several clusters therefore no clear genetic outlier is indicated. “Blue Dream” (BlDr) samples are blue, “Mob Boss” (MoBo) samples are green, “Sour Diesel” (SoDi) are red, “OG Kush” (OGKu) samples are purple, and “Durban Poison” (DuPo) samples are yellow.

Sample Selection for the Sensory Study

Genetic relatedness, PCoA clustering, genetic structure, and hierarchical clustering based on genetic distance clearly identified genetic outliers and consensus samples within strains, with the exception of "Sour Diesel" which was omitted from the sensory study. The genetic outliers "Durban Poison" 8GN, "OG Kush" 1SN, "Blue Dream" 3SN, and "Mob Boss" 5SN consistently showed differentiation from the other samples in the strain, but there was insufficient sample of "Durban Poison" 8GN to be included in the study. Three identical "Durban Poison" samples were included to examine scent variation among samples with identical genotypes. The remaining 12 samples selected for the sensory study had either identical or an extremely high degree of genetic similarity (Table 4.2). Samples from the sensory selection were included in a second dendrogram to confirm clustering and genetic outliers (Figure 4.7).

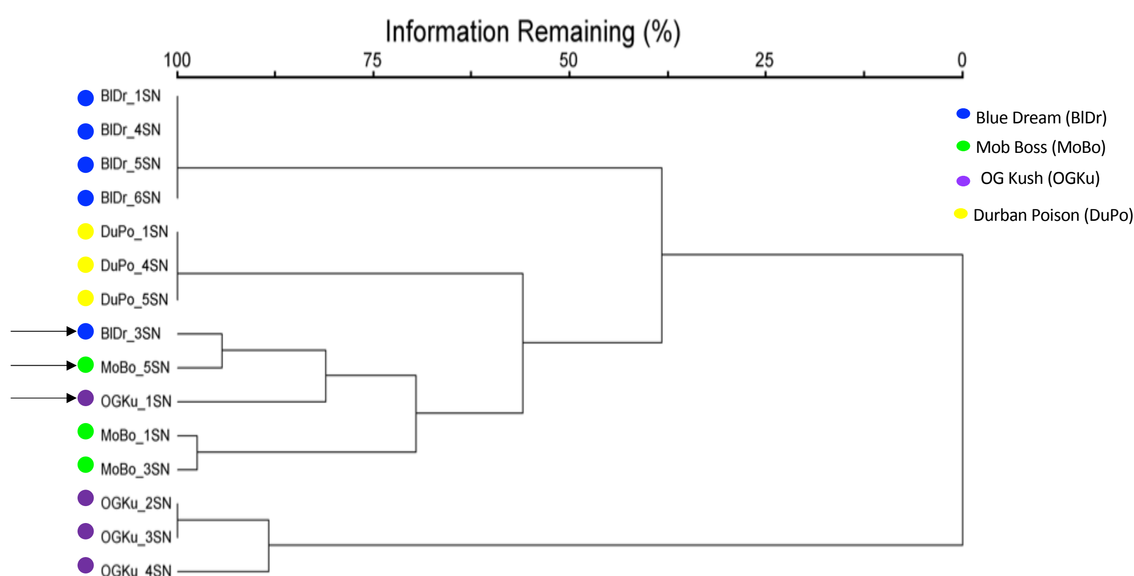


Figure 4.7. Hierarchical genetic cluster analysis of the cannabis samples. Arrows indicates genetic outliers. "Blue Dream" (BlDr) samples are red, "Mob Boss" (MoBo) samples are green, "OG Kush" (OGKu) samples are purple, and "Durban Poison" (DuPo) samples are yellow.

Sensory Results

Subject Demographics

Fifty-five people (33 men, 22 women; mean age 29.5 ± 7.8 years) were tested. Of these, all but eight had purchased *Cannabis* since January 1, 2014, and all but five subjects had smoked it. The high rates of purchase (85.5%) and use (90.9%) among study participants occurred despite efforts to recruit former and non-users as well. Seven subjects (12.7%) had taken part in previous cannabis sniff studies (Gilbert & DiVerdi 2018; Gilbert & DiVerdi, submitted).

Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) yielded a configuration consisting of two large clusters, designated as Cluster A' and Cluster B' (Figure 21). Previous sensory research examining scent profiles in *Cannabis* found two clusters and designated them as Cluster A and Cluster B (Gilbert and DiVerdi 2018). To avoid confusion between the previous work and the present study, we designated the two main clusters from this work as Cluster A' and Cluster B' (Figure 4.8). We feel that Gilbert and DiVerdi's (2018) Cluster A and Cluster B largely correspond to the current Cluster A' and Cluster B', but there are some minor discrepancies. Cluster A' contains "OG Kush", as it was previously assigned to Cluster A and described as citrus, lemon, sweet and pungent (Gilbert and DiVerdi 2018). Cluster A' contained all four samples of "OG Kush", which scored high on earthy, woody, and herbal descriptors, which is consistent with previous aromatic profiling and descriptions online (Gilbert and DiVerdi 2018; Leafly 2018b). Cluster B' contains "Durban Poison", as it was previously assigned to Cluster B, and described as citrus, lemon, sweet, pungent (Gilbert and DiVerdi 2018). However, in this analysis two

samples of "Durban Poison" were described as sweet, citrus, flowery, lemon, and the third was described as herbal, woody, flowery, and earthy, differing from Gilbert and DiVerdi (2018). This is the first study to provide olfactory analysis of the "Blue Dream" strain, and the results are ambiguous; three samples were assigned to Cluster B', while two samples, including the genetic outlier, were assigned to Cluster A'. "Mob Boss" was previously grouped with strains in Cluster A (earthy, woody, herbal) (Gilbert and DiVerdi 2018). In this study, only one sample of "Mob Boss" was characterized this way (MoBo_3SN), with the other two samples, including the genetic outlier, assigned to Cluster B'.

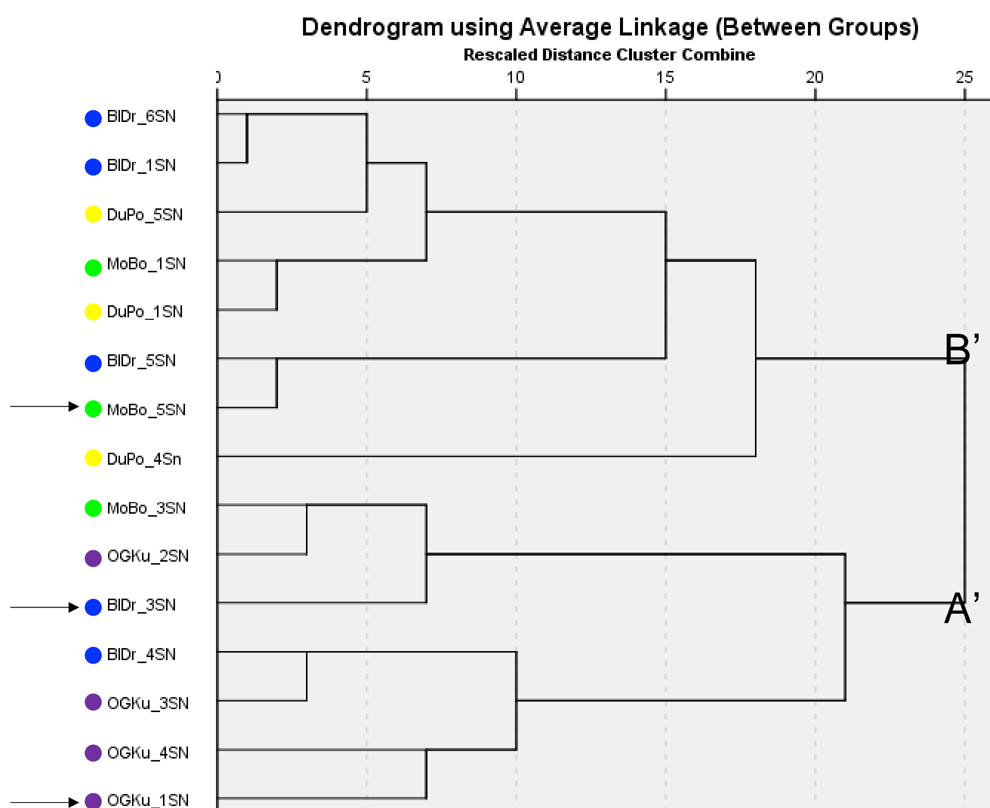


Figure 4.8. Hierarchical cluster analysis (HCA) for 15 samples included in the sensory study containing two large clusters, Cluster B' and Cluster A'. Strains are color coded and arrows indicate the genetic outlier for each if the strains. All "Durban Poison" samples were genetically identical. "Blue Dream" (BlDr) samples are blue, "Mob Boss" (MoBo) samples are green, "OG Kush" (OGKu) samples are purple, and "Durban Poison" (DuPo) samples are yellow.

Within Strain Descriptor Profiles

The aim of this study was to assess if the aroma profile of an anomalous genetic sample differed from those of the consensus samples. As a first effort, a Friedman's nonparametric repeated-measures ANOVA was performed on the summed frequencies for each odor descriptor across all samples within a given strain. All four strains returned significant chi-square values (Table 4.7). Thus, for example, aroma profiles of the five samples of "Blue Dream" differ significantly.

Table 4.7. Results of Friedman's test on the summed frequencies for each odor descriptor across all samples within a given strain for each *Cannabis* strain. The number of samples (N) in each strain, along with the Chi-square value, degrees of freedom (df) and asymptotic significance (Asymp. Sig).

	Blue Dream	**Mob Boss**	**OG Kush**	**Durban Poison**
N	5	3	4	3
Chi-Square	117.668	69.398	107.223	91.776
df	39	39	39	39
Asymp. Sig.	0.000	0.002	0.000	0.000

This analysis, however, does not address differences between specific samples within a strain, and in particular, if an anomalous sample differed from consensus samples. A Wilcoxon signed-rank test on all pairwise combinations of samples within each strain (analogous to post-hoc t-tests in ANOVA) was conducted. None of these pairwise comparisons yielded a statistically significant difference. We are not the first to note that this is a paradoxical result, given the significant results of the overall Friedman tests (Zimmerman and Zumbo 1993).

In order to assess the question of within strain aromatic differences with special interest to the genetic outlier, characterization of within-strain aroma differences was analyzed using the number of times particular odor descriptors were reported among the

participants. The five most frequently endorsed odor descriptors for each sample were identified and pooled within each strain. This resulted in 12 pooled descriptors for both “Blue Dream” and “Durban Poison”, nine for “Mob Boss”, and eight for “OG Kush”. Frequency counts for each descriptor were averaged across a strain’s consensus samples and compared to the counts for the genetic outlier sample (Figures 4.8 – 4.11).

The results for “Durban Poison” were interesting and somewhat unexpected. Although three samples were genetically identical, all three differed across the 12 pooled descriptors (Figure 4.9, Table 4.8). However, when examining the pooled five most frequently endorsed odor descriptors of the consensus samples of “Blue Dream”, “Mob Boss” and “OG Kush” (Figure 4.10, Table 4.8) there is also evidence of inconsistency in the aromas of the genetically cohesive samples of all strains. In order to examine the aroma consistency of “Durban Poison” we calculated the mean, standard deviation, and average standard deviation of the top descriptors for the consensus samples for each strain (Table 4.7). Since the strains have different aromas, we focused on the range of standard deviations and the average standard deviation: “Durban Poison” standard deviation range = 2.08-7.64, average standard deviation = 4.29; “OG Kush” standard deviation range = 1-7.57, average standard deviation = 4.37; “Blue Dream” standard deviation range = 2-7.18, average standard deviation = 5.12; and “Mob Boss” standard deviation range = 0.71-8.49, average standard deviation = 3.22.

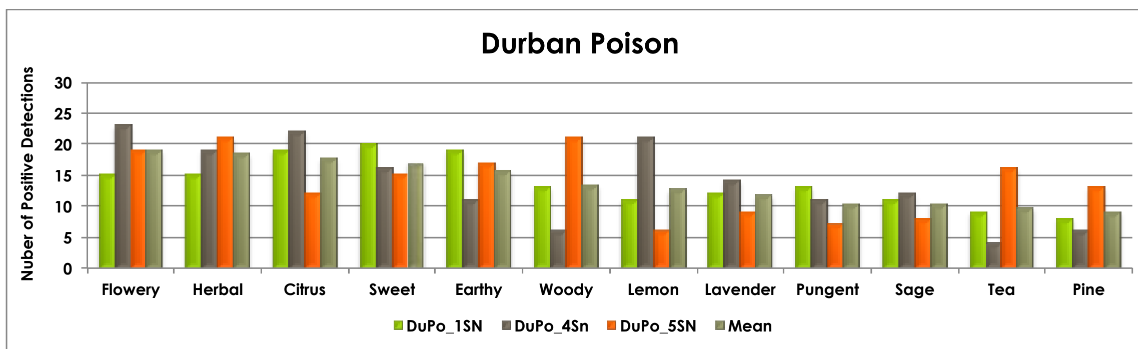
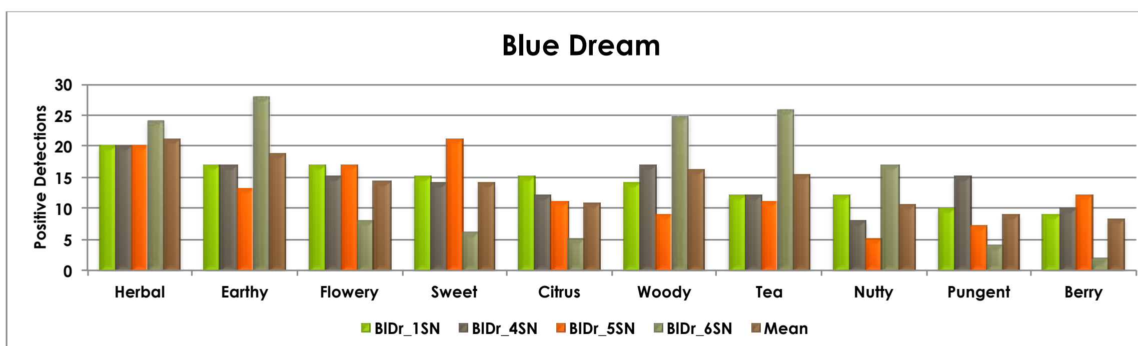
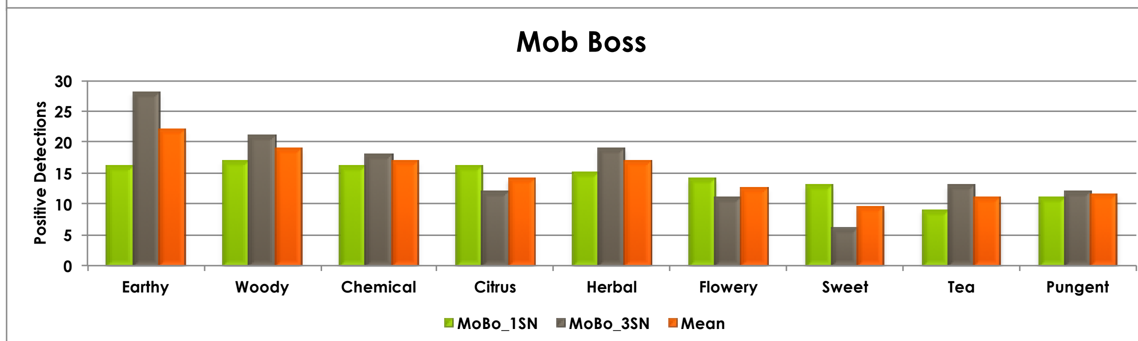


Figure 4.9. Detection frequency of top-rated odor descriptors for "Durban Poison".





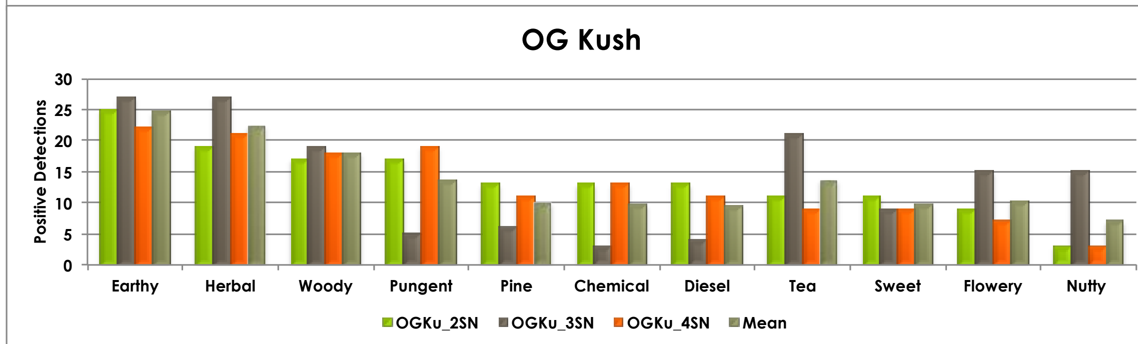


Figure 4.10. Detection frequency of top-rated descriptors for consensus samples of "Blue Dream", "Mob Boss" and "OG Kush" and the mean for each descriptor.

Table 4.8. Frequencies plus mean and standard deviation of the top five pooled odor descriptors for the genetic consensus samples of "Durban Poison", "OG Kush", "Blue Dream" and "Mob Boss".

	Flowery	Herbal	Citrus	Sweet	Earthy	Woody	Lemon	Lavender	Pungent	Sage	Tea	Pine
DuPo_1SN	15	15	19	20	19	13	11	12	13	11	9	8
DuPo_4SN	23	19	22	16	11	6	21	14	11	12	4	6
DuPo_5SN	19	21	12	15	17	21	6	9	7	8	16	13
Mean	19.00	18.33	17.67	17.00	15.67	13.33	12.67	11.67	10.33	10.33	9.67	9.00
SD	4.00	3.06	5.13	2.65	4.16	7.51	7.64	2.52	3.06	2.08	6.03	3.61

	Earthy	Herbal	Woody	Pungent	Pine	Chemical	Diesel	Tea	Sweet	Flowery	Nutty
OGKu_2SN	25	19	17	17	13	13	13	11	11	9	3
OGKu_3SN	27	27	19	5	6	3	4	21	9	15	15
OGKu_4SN	22	21	18	19	11	13	11	9	9	7	3
Mean	24.67	22.33	18.00	13.67	10.00	9.67	9.33	13.67	9.67	10.33	7.00
SD	2.52	4.16	1.00	7.57	3.61	5.77	4.73	6.43	1.15	4.16	6.93

	Herbal	Earthy	Flowery	Sweet	Citrus	Woody	Tea	Nutty	Pungent	Berry
BlDr_1SN	20	17	17	15	15	14	12	12	10	9
BlDr_4SN	20	17	15	14	12	17	12	8	15	10
BlDr_5SN	20	13	17	21	11	9	11	5	7	12
BlDr_6SN	24	28	8	6	5	25	26	17	4	2
Mean	21.00	18.75	14.25	14.00	10.75	16.25	15.25	10.50	9.00	8.25
SD	2.00	6.45	4.27	6.16	4.19	6.70	7.18	5.20	4.69	4.35

	Earthy	Woody	Chemical	Citrus	Herbal	Flowery	Sweet	Tea	Pungent
MoBo_1SN	16	17	16	16	15	14	13	9	11
MoBo_3SN	28	21	18	12	19	11	6	13	12
Mean	22.00	19.00	17.00	14.00	17.00	12.50	9.50	11.00	11.50
SD	8.49	2.83	1.41	2.83	2.83	2.12	4.95	2.83	0.71

The results for “OG Kush” (Figure 4.11) show that the genetic outlier sample was far more cheesy and less pungent and tea-like, than the mean of the consensus samples. “OG Kush” was characterized by Gilbert and DiVerdi (Gilbert and DiVerdi 2018) as having the earthy/woody/herbal aroma profile characteristic of Cluster A. All the “OG Kush” samples in the present study, including the genetic outlier, were rated highly on these three descriptors, confirming the earlier results. The fact that cheese was a relatively frequently endorsed descriptor for the anomalous genetic sample is noteworthy: cheese was not a highly ranked descriptor for any of the 11 strains tested previously (Gilbert and DiVerdi 2018). The blue cheese descriptor was also reported more than twice the frequency for the outlier than the consensus samples. Although blue cheese was not in the five most frequently endorsed odor descriptors, it lends more evidence for the unusual scent profile of the genetic outlier in the set. Thus, the “OG Kush” genetic outlier was distinctive from both the strain-typical profile, as well as the consensus samples.

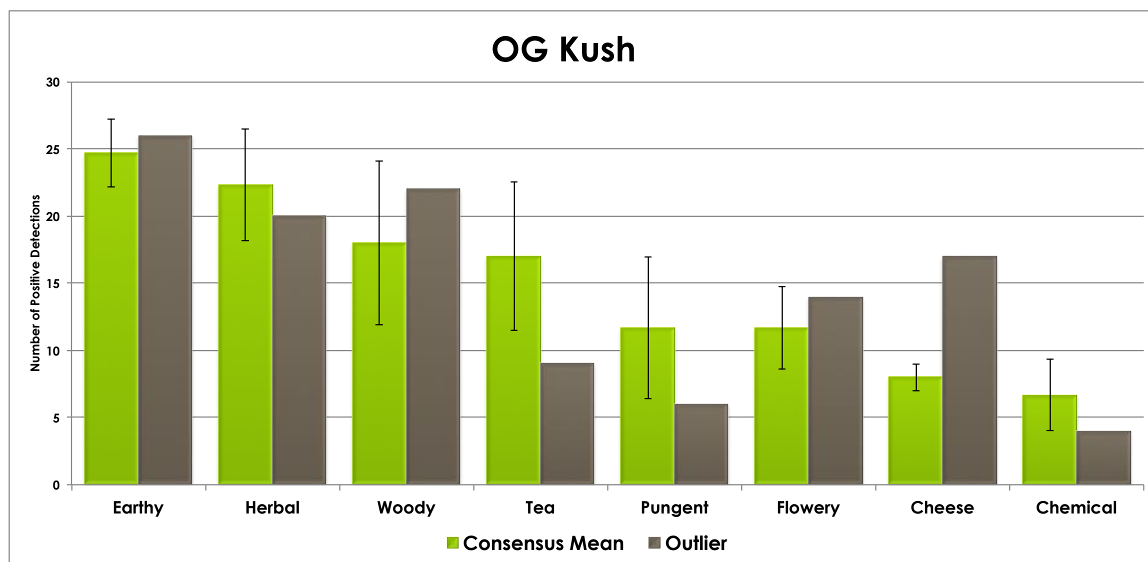


Figure 4.11. Detection frequency of top-rated odor descriptors for “OG Kush” samples; error bars indicate standard deviation from the mean of the consensus samples.

The results for “Blue Dream” (Figure 4.12) show that the genetic outlier was far more pungent, chemical, and skunk-like than the mean of the four consensus samples. HCA configuration (above) indicates that two “Blue Dream” samples align with Cluster A’, and three with Cluster B’. Despite this anomalous result, within-strain comparison shows the genetic outlier sample to have unique and marked differences from the other samples.

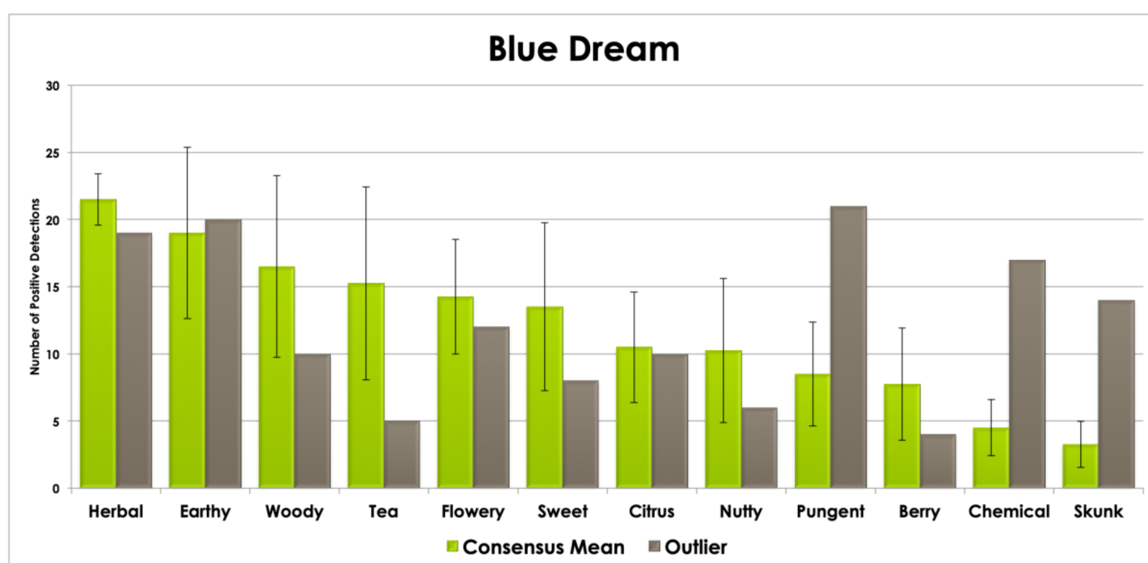


Figure 4.12. Detection frequency of top-rated odor descriptors for “Blue Dream” samples; error bars indicate standard deviation from the mean of the consensus samples.

The results for “Mob Boss” (Figure 4.13) show that the genetic outlier was strikingly more flowery, sweet, and berry-like, and less woody and chemical than the consensus samples. Gilbert & DiVerdi (Gilbert and DiVerdi 2018) found that “Mob Boss” has an earthy/woody/herbal aroma profile typical of Cluster A strains. The consensus genetic samples of “Mob Boss” tested here fit the Cluster A profile: they were described as earthy/woody/herbal/chemical and woody/earthy/citrus/chemical, respectively. In contrast, the description of the genetic outlier sample (MoBo_5SN) as

flowery/sweet/herbal/berry was at odds with both the consensus samples and the previously established Cluster A profile.

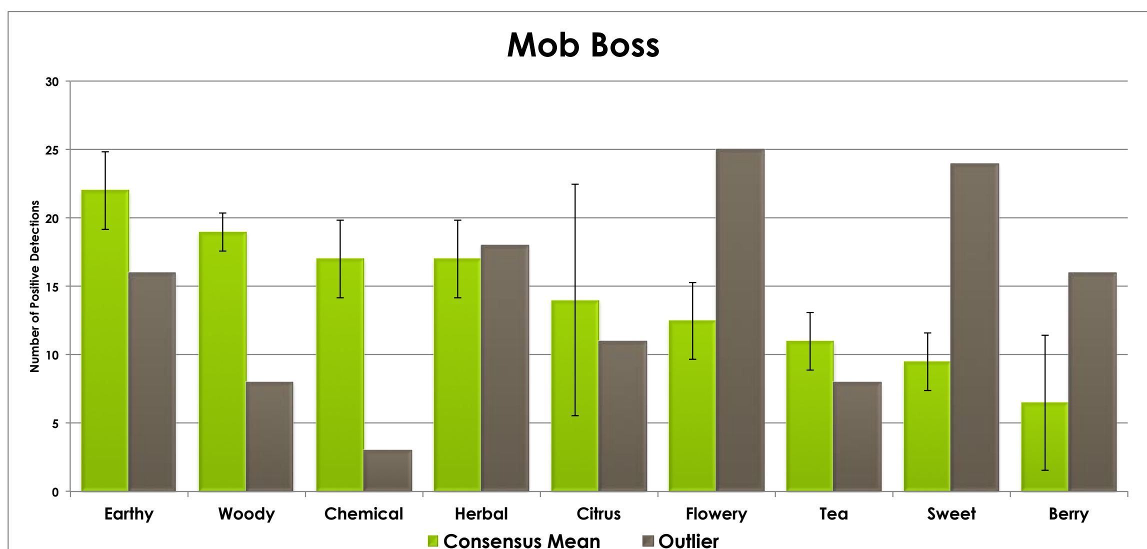


Figure 4.13. Detection frequency of top-rated odor descriptors for "Mob Boss" samples; error bars indicate standard deviation from the mean of the consensus samples.

Additional Within-Strain Scent Profiling Analyses

In order to lend a more holistic analysis of the aroma profiles, the 40 odor descriptors (Table 4.4) were broken down into smaller subsets which included categories of scents based on broad characters common to the scents included in each category. These categories were then used to determine if there were differences in overall characteristic aromas between the samples within strains that were identified as genetically cohesive and the sample identified as the genetic outlier (Table 4.3).

Within Strain Scent Profiling: Perceived Shared Character Categories

The Perceived Shared Character Category analyses was used to determine if genetics consensus and outlier samples differ when organized by categories including all scents. The "OG Kush" outlier was identified as pungent and earthy, but less sweet than

the consensus samples (Figure 4.14C). The genetic outlier of Blue Dream" was far more pungent than the consensus samples (Figure 4.14A). The genetic outlier of "Mob Boss" has unique and marked differences from the other samples in that it was identified as floral and sweet, and less pungent (Figure 4.14B). "Durban Poison" was not included in this analysis as all the samples were genetically identical.

Table 4.9. Shared Character Category Frequency of Detection. Scores for each scent category (Pungent, Sweet, Floral, Earthy, Spicy) for the genetic consensus samples and the genetic outlier.

		Pungent	Sweet	Floral	Earthy	Spicy
Blue Dream	Consensus Mean	0.074	0.084	0.081	0.139	0.141
	Outlier	0.239	0.074	0.085	0.135	0.142
Mob Boss	Consensus Mean	0.117	0.070	0.068	0.111	0.099
	Outlier	0.068	0.157	0.194	0.099	0.124
OG Kush	Consensus Mean	0.087	0.123	0.049	0.149	0.152
	Outlier	0.133	0.058	0.091	0.210	0.148

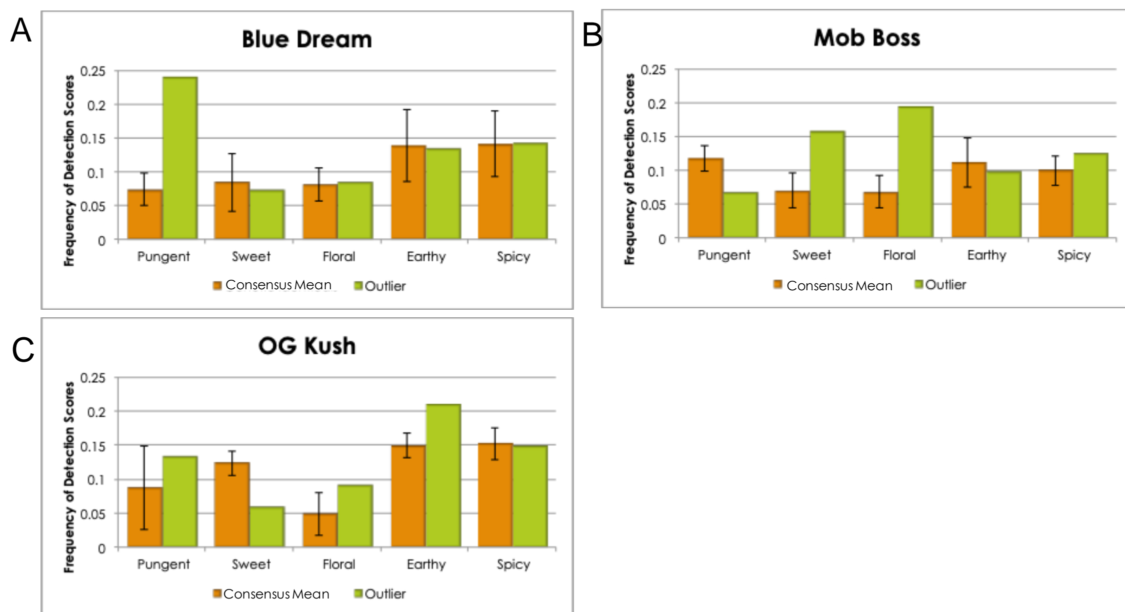


Figure 4.14. Shared Character Category Histograms. Frequency of Detection Scores of Pungent, Sweet, Floral, Earthy, Spicy detected among the samples by the 55 participants for "Blue Dream" (A), "Mob Boss" (B) and "OG Kush" (C).

Within Strain Scent Profiling:
***Cannabis* Lexicon Categories**

A second analysis looking at the impact of grouping scents used *Cannabis* Lexicon Categories. With these groupings, the "OG Kush" outlier identified far more similarly to the consensus samples but was less bitter (Figure 4.15C). The genetic outlier of Blue Dream" was far more bitter than the consensus samples (Figure 4.15A). The genetic outlier of "Mob Boss" was identified as sweet, with less spicy and bitter notes (Figure 4.15B). "Durban Poison" was not included in this analysis as all the samples were genetically identical.

Table 4.10. The Frequency of Detection Scores. Scores for each *Cannabis* Lexicon Category for the genetic consensus samples and the genetic outlier.

		Sweet	Sour	Spicy	Bitter
Blue Dream	Consensus Mean	0.105	0.158	0.158	0.130
	Outlier	0.063	0.139	0.139	0.195
Mob Boss	Consensus Mean	0.079	0.155	0.155	0.170
	Outlier	0.202	0.129	0.129	0.074
OG Kush	Consensus Mean	0.068	0.162	0.162	0.170
	Outlier	0.068	0.162	0.162	0.132

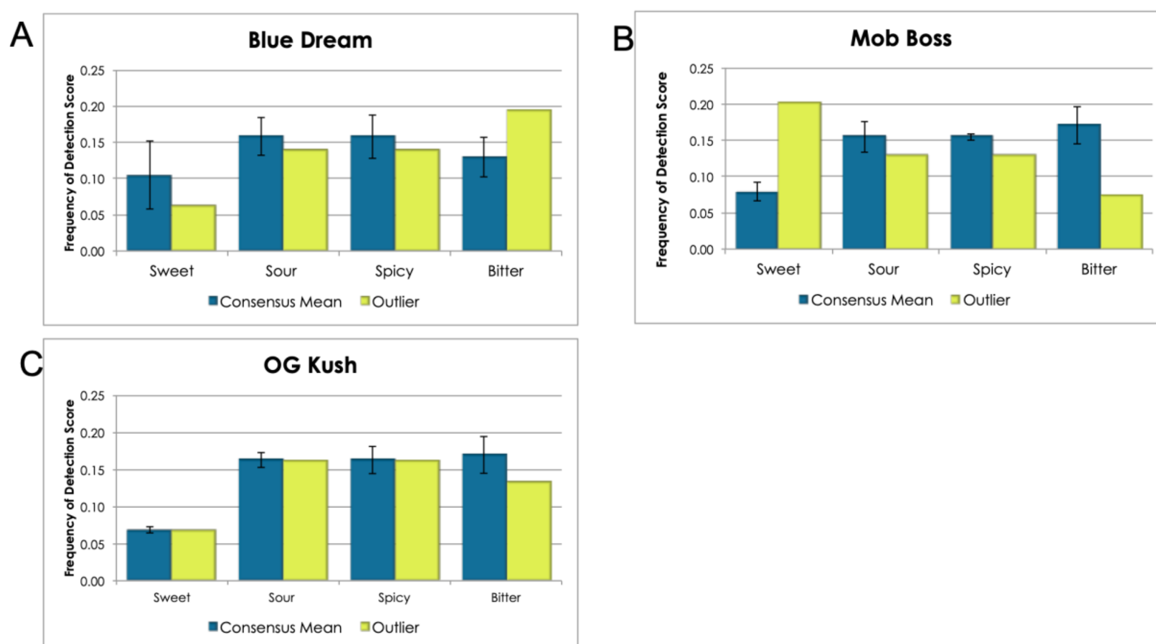


Figure 4.15. Histograms for Frequency of Detection Scores of each *Cannabis* Lexicon category (Sweet, Sour, Spicy and Bitter). Frequency of categorical scents detected among the samples by the 55 participants for "Blue Dream" (A), "Mob Boss" (B) and "OG Kush" (C).

Discussion

In the present study we purposefully targeted *Cannabis* strains with different scent profiles and sought to evaluate if genetic variation manifests as discrepant aromatic characterization. Using the combination of previously published genotyping methods and olfactory phenotyping techniques, a two-part study was designed to uncover the relationship between genotype and phenotype in *Cannabis*. First, samples were collected and genotyped to identify cohesive samples and genetic outliers in five strains. Then, a sensory study was conducted using non-expert participants to determine if participant descriptors aligned with the genetic results.

Genetic analyses were conducted on multiple samples of targeted strains to identify genetically cohesive and outlier samples. Previous research (Schwabe and McGlaughlin 2018) (Chapter II) found genetic variability in "Durban Poison", "OG

Kush", "Sour Diesel" and "Blue Dream", and although "Mob Boss" had not been genotyped, all strains analyzed showed variation. The genetic results were clear and robust with support for three genetic groups (Figures 4.4 and 4.5). Relatedness statistics and clustering analyses revealed four of five strains had a cohesive genetic signal in the majority of the samples, and one clear genetic outlier. The requirement for samples selection for the sensory portion were that the genetically cohesive and outlying samples were supported in each of the analyses conducted. Two strains, "OG Kush" and "Blue Dream", had consistent consensus genotypes with a clear outlier, fitting the requirements for inclusion in the sensory study. The outlier for "Mob Boss" was not as divergent but fulfilled the requirements for sensory inclusion. "Sour Diesel" had no consistent genetic outlier, and therefore was not included in the sensory study, even though this strain reportedly has a unique aromatic profile (Gilbert and DiVerdi 2018; Leafly 2018b). "Durban Poison" had a clear outlier, but there was not enough sample tissue to be included in the sensory portion. However, three identical "Durban Poison" samples were chosen to serve as a control to examine scent variation among samples with identical genotypes.

Fifteen samples from four strains were selected for inclusion in the sensory study: "Durban Poison" (3), "OG Kush" (4), "Blue Dream" (5), and "Mob Boss" (3) (Table 4.3). Clustering analysis of the sensory data was at odds with the clustering analysis of the genetic data (Figure 4.7 and 4.8). Three "Durban Poison" were assigned to Cluster A' indicating a degree of aroma similarity. Four "OG Kush" samples were assigned to Cluster B'. Genetic consensus samples of "Blue Dream" were assigned to both Cluster A' (1) and Cluster B' (3), indicating some dissimilarity between those genetic consensus

samples. One genetic consensus sample of "Mob Boss" was assigned each to Cluster A' and B', indicating some dissimilarity between those genetic consensus samples. The genetic outliers for "OG Kush", "Blue Dream", and "Mob Boss" clustered with at least one genetic consensus sample.

Analysis of variance (ANOVA) revealed samples within strains differed significantly from one another (Table 4.7). This is evident in "Durban Poison" where frequently endorsed odor descriptors show a high degree of variation, even though the genotypes are identical (Figure 4.9). "Blue Dream" and "Mob Boss" with identical genotypes, and "OG Kush" (4SN) with difference at one locus, all had a high degree of variation in the frequently endorsed odor descriptors (Figure 4.10). Due to this variation, a method to uncover within strain aromatic differences with special interest to the genetic outlier was needed. As this type of study combining genetic and sensory data had not previously been published, a novel approach was required. The first approach was to pool the five most frequently reported strain descriptors for the genetic consensus samples for each strain. The mean frequency for each descriptor for the genetic consensus samples was compared to the number of reports for the descriptor for the genetic outlier. This analysis revealed substantial differences in the aromatic profiles of the outliers in each strain (Figures 4.10 - 4.12). The second approach used categories of scents to calculate frequencies of detection in each category to compare aroma profiles of the genetic consensus samples to the genetic outlier. Two different scent category profiles were created, one from perceived shared characters (Shared Characteristics Category) and one from *Cannabis* flavor lexicons (*Cannabis* Lexicon Category). Both the profiles returned

at least one category in which the genetic outlier was considerably different than those of the genetic consensus samples (Figures 4.14 and 4.15).

"OG Kush" is reportedly a clone only strain (Chapter II) described online as woody, pine and earthy (Leafly 2018b) and is assigned to Gilbert and DiVerdi's (2018) earthy, woody and herbal Cluster A. Hierarchical Cluster Analysis of the scent data grouped all "OG Kush" samples, including the genetic outlier, in Cluster A' with the genetic outlier nested with consensus samples (Figure 4.8). Analyses using frequently endorsed odor descriptors (Figure 4.11) indicate the "OG Kush" outlier had less of a tea scent and a distinctive cheese aroma when compared to the consensus samples. The standard deviation from the mean of the consensus samples of the 11 pooled descriptors ranged from 1.0 (woody) to 7.57 (pungent) (Figure 4.10, Table 4.8). Analysis of frequencies in scent categories show the genetic outlier in "OG Kush" is less sweet and more earthy according to Shared Character Category (Figure 4.14C) and more bitter according to *Cannabis* Lexicon Category (Figure 4.15C) relative to consensus samples. The odor descriptors reported for all "OG Kush" samples align with consumer expectations as earthy, woody and herbal (Figure 4.11), however the outlier has some unique aromatic qualities that separate it from the consensus samples (Figure 4.11, 4.14C and 4.15C) indicating the genetic outlier could be identified using sensory perception methods.

"Blue Dream" is reportedly a clone-only strain (Chapter II) described as having blueberry, berry and sweet flavors (Leafly 2018b) but has not been previously described using validated sensory methods (Gilbert and DiVerdi 2018). Hierarchical Cluster Analysis of the scent data split the "Blue Dream" samples over the two major scent

clusters with two samples, including the genetic outlier, in a cluster A', and the remaining three samples in Cluster B' (Figure 4.8). HCA was unable to differentiate the "Blue Dream" outlier. Further analyses using frequently endorsed odor descriptors (Figure 4.12) indicate the "Blue Dream" outlier was less herbal and tea, but far more pungent, chemical and skunk aromas were detected compared to the consensus samples. The standard deviation from the mean of the consensus samples of the 10 pooled descriptors ranged from 2.0 (herbal) to 7.18 (tea) (Figure 4.10, Table 4.8). Analysis of frequencies in scent categories show that the genetic outlier in "Blue Dream" is overwhelmingly more pungent according to Shared Character Category (Figure 4.14A) and bitter according to *Cannabis* Lexicon Category (Figure 4.15A) than consensus samples. The odor descriptors reported for the outlier do not align with consumer expectations as a blueberry, berry and sweet strain, indicating the genetic outlier could be identified using sensory perception methods.

"Mob Boss" is described online as woody, pine and earthy (Leafly 2018b) and is assigned to Gilbert and DiVerdi's (2018) earthy, woody and herbal Cluster A. Hierarchical Cluster Analysis of the scent data split the "Mob Boss" samples over the two major scent clusters with two samples, including the genetic outlier, in Cluster B' and one consensus sample in Cluster A' (Figure 4.8). Analyses using frequently endorsed odor descriptors (Figure 4.13) indicate the "Mob Boss" outlier was less earthy, woody and chemical, but extensive flowery, sweet, and berry aromas were detected compared to the consensus samples. The standard deviation from the mean of the consensus samples of the 10 pooled descriptors ranged from 0.71 (pungent) to 8.49 (earthy) (Figure 4.10, Table 4.8). Analysis of frequencies in scent categories show the genetic outlier in "Mob Boss"

overwhelmingly more floral (Figure 4.14B) and sweet (Figure 4.14B and 4.15B) according to Shared Character Category, and less pungent (Figure 4.14B) and bitter (Figure 4.15B) according to *Cannabis* Lexicon Category than the consensus samples. The odor descriptors reported for the outlier do not align with consumer expectations as a woody, pine and earthy strain. Floral, sweet, and berry (Figures 4.9, 4.14B, and 4.15B), and are uncharacteristic of aromas associated with "Mob Boss", indicating the genetic outlier was identified using sensory perception methods.

"Durban Poison" is described online as earthy, pine and sweet (Leafly 2018b) which is quite different from Gilbert and DiVerdi's (2018) assignment to the citrus, lemon, sweet, and pungent Cluster B. Three genetically identical "Durban Poison" samples were selected for the sensory study (Table 4.3, Figures 4.2, 4.3, 4.5 - 4.7) to determine variation in the scent profile of samples from different origins with no genetic variation. The HCA assigned the three samples in to the Cluster B' (Figure 4.8) indicating a similar major aromatic profile among the three. The frequency of detection for 12 descriptors (Figure 4.13) indicated there was variation in scent detection by the participants. As these three samples were genetically identical, it would stand to reason that they should have similar aroma profiles. This expectation was not met in "Durban Poison" as all samples deviated from one another across the most frequently reported odor descriptors. Moreover, there was not a single sample that consistently deviated from the others. There was variation in scent detection by the participants in the genetic consensus samples in all four strains. (Figure 4.13 and 4.14; Table 4.8). Additionally, "Durban Poison" and "Blue Dream" had higher average standard deviations (4.37 and 5.12 respectively) than "Durban Poison" and "Mob Boss" (4.29 and 3.22 respectively).

From this analysis it is clear that genetic identity alone does not control aroma profiles and that differences were observed among consensus samples for all strains.

The utility of combining genetic and sensory methods to determine if genetic anomalies are detectable through sensory perception was confirmed for the samples used here. However, several challenges and limitations need to be addressed. It is unclear from the results of this study how much variation in aromatic profiles is standard within strains, as all the consensus sequences have scent profile variation (Figure 4.9, 4.10; Table 4.8). Participants for sensory studies such as this are limited by sensory overload to a maximum of 15 samples per sitting, resulting in a forced small samples size, but the number of participants in the study give strong support for the results herein. We are aware that sensory perception is personal and subjective, but the aim of the study was not to identify samples, but rather to determine if genetic outliers have a different aroma profile from genetically cohesive samples, which was achieved in the three strains with an outlier. We are also aware that an untrained consumer panel such as that used here tends to show greater variation in descriptor use than a highly trained expert panel. Thus, some participants may be very lax in reporting the odors they detect, while others may be more conservative, and of course there is everything in between. However, we feel that the number of participants in this study well represent consumer perceptions. The data were difficult to analyze because each 55 participants have the option to detect 40 aromas in 15 samples, and both presence and absence of scents are considered character states. Standard statistical analyses methods detected differences among all samples, and within strains, but given the variation seen among genetically cohesive samples of the same strain, standard statistical analyses were not appropriate. For this reason, the data analysis

required a novel approach. Grouping scents into categories is novel, and not necessarily a validated method for analyzing sensory data such as these, but this approach did reveal differences between genetically anomalous samples and those with cohesive genotypes.

Terpenes are aromatic phytochemicals produced in many plants that contribute to the multitude of aromas associated with leaves, flowers, and fruits, including those of *Cannabis*. Previous research has found terpenes vary in concentration over time (Aizpurua-Olaizola et al. 2016). As there are no standard growing conditions or harvest protocols in place for the *Cannabis* industry, scent variation in genetically identical samples could be due to differences in growing conditions (nutrients, light, etc.), harvest time and/or post-harvest flower processing (Cervantes 2006) among different grow facilities. In drug-type *Cannabis* strains, terpenes in flowers increase in concentration from day 122 until about day 165 of the growth cycle when levels began to decrease (Aizpurua-Olaizola et al. 2016). However, there is no standard time to harvest *Cannabis* flowers. Maturity varies among plants, and different strains reportedly mature at different times, ranging from 3-6 months (Leafly 2018b). Growers harvest flowers by examining the color of the stamens and the cloudiness of the trichomes under a magnifying glass. Growers aim to harvest at the height of THC production, which is when the stamens have dried and turned amber and trichomes have developed a spherical head and have clear or a creamy appearance (Figure 4.16) (Cervantes 2006). Although there have been studies analyzing cannabinoid and terpene production over time in plants grown under standardized conditions (Aizpurua-Olaizola et al. 2014), harvesting and processing information are not provided to consumers, and may well be beyond what consumers are willing to take into account when purchasing *Cannabis*. Our results suggest there can be

significant variation in terpenes, as seen through aroma descriptors, even when samples are genetically identical (Table 4.8), which could be surprising for those expecting certain characteristics in a strain. Although this study included a relatively small sampling of *Cannabis* strains, it demonstrates genetic variation is reflected in aromatic profile differences. However, in order to determine what is a reasonable amount of variation in scent profiles of *Cannabis*, more in-depth studies examining other factors known to influence terpene production are needed.

Figure 4.16. The glandular trichomes of the female flower of *Cannabis sativa* at 40X magnification.

Conclusion

The purpose of this study was to investigate if genetic inconsistencies are reflected in aroma profiles. Strains were selected based on availability, previous genotyping (Schwabe and McGlaughlin 2018) (Chapter II), and sensory research (Gilbert and DiVerdi 2018). We found that the tested strains provided additional evidence that samples of the same strain from different origins can have unexpected genetic and aroma variation, but consensus samples seem to align with previous sensory analysis profiles

(clusters A and B; Gilbert and DiVerdi 2018), as well as online descriptions (Gilbert and DiVerdi 2018; Leafly 2018b). Comparison of top-rated odor descriptors revealed that, compared to consensus samples, genetic outliers have distinctive aroma profiles. Broader scent categories, Shared Character Category and *Cannabis* Lexicon Category revealed genetic variants have different scent profiles than those of the consensus samples. In samples with identical genotypes there were notable differences in reported aroma descriptors, which could be attributed to differences in growing and curing processes among different grow facilities. We believe genetic variation in the samples included here were adequately reflected in differences in aromas, but aroma variation in samples with highly similar genotypes is substantial, which could be the result of different cultivation practices among different *Cannabis* grow facilities. These results show not only that genetic imposters within a strain can be detected and result in aromatic differences, but also that there is considerable variation in aromas among samples with identical genetic identity.

CHAPTER V

PHYTOCHEMICAL VARIATION

Contribution of Authors and Co-Authors

Manuscript in Chapter V

Author: Anna Schwabe

Contributions: Conceived the project, provided some funding, recruited participants, collected samples, conducted DNA extractions, designed and optimized microsatellite primers, compiled and analyzed data, and drafted manuscript content.

Co-Author: Josh Harrelson, Mile High Labs

Contributions: Conducted cannabinoid analysis and data collection, provided cannabinoid analysis methods.

Co-Author: Richard Crawford, Colorado Green Films Technology

Contributions: Conducted terpene analysis and data collection, provided terpene analysis methods.

Co-Author: Dr. Mitchell McGlaughlin

Contributions: Directed the project, provided some funding, contributed statistical analysis and manuscript revisions.

Abstract

This research investigates how genetic variation in neutral genetic markers is related to variation in cannabinoid and terpene content of *Cannabis*. The expanding *Cannabis* industry needs to provide consistent products to the ever-increasing customer base in both the recreational and medical marketplaces. Sources of variation are numerous and genetic variation has been found where there should be little to none, especially when striving for consistency. Moreover, there are currently no widely practiced standard growing or harvesting protocols to minimize variation among growing facilities. Variation in growing conditions, harvest time, curing and storage conditions can affect *Cannabis* chemotypes, but to what extent is largely a mystery. We conducted a small investigation with 15 samples from four strains. Using 10 microsatellite markers, we identified genetically cohesive samples of “Durban Poison” (n=3), “Blue Dream” (n=4), “Mob Boss” (n=2), and “OG Kush” (n=3). We also identified one genetic outlier for “Blue Dream”, “Mob Boss”, and “OG Kush”. We compared the chemotypes of the genetic outliers to the genetic consensus samples to investigate whether genotypic differences are reflected as chemotypic differences. A panel of nine cannabinoids and 21 terpenes were analyzed using HPLC-DAD and GC-MS, respectively. An additional three strains were added to the terpene analysis to examine terpene variation among strains. The results from this study show that cannabinoids and terpenes are highly variable among samples independent of their genetic similarity, as well as among strains. The relationship between genetic assignment and cannabinoid profile is less pronounced than the terpene profile, where genetic outliers differed substantially from the consensus samples in $>50\%$ of the 21 terpenes analyzed. Although the sampling in this study was

not extensive and include a relatively small number of cannabinoids and terpenes, the results clearly demonstrate both are quantitatively and qualitatively inconsistent among samples, and, therefore, using chemotype to identify *Cannabis* strains is not recommended.

Introduction

Overview

Human driven dispersal across the globe along with cultivation for thousands of years have resulted in many different varieties of *Cannabis,* which have been selected for a multitude of desirable characters, the most notorious of which are the phytocannabinoids. Although research on the chemical constituent differences among *Cannabis* strains is abundant (e.g., Pacifico et al. 2006), there are few studies examining the genetic contribution to chemotypic variation within strains (de Meijer and Hammond 2005; Lynch et al. 2016; Sawler et al. 2015).

Cannabis varietals are preferably propagated through cloning techniques to ensure minimal phenotypic variation among plants within strains. For some strains, stable seeds are available, but there is a lack of information for many of these seeds, and it is not clear how much variation there is within seed lots from many sources. Recent research has shown substantial genetic variation within strains from different sources (Schwabe and McGlaughlin 2018) as well as within seeds of the same varietal (Soler et al. 2017).

Medical and recreational drug type *Cannabis* breeders aim to produce a consistent product and maximize yield by using all female clones from a single mother plant. Given that desirable *Cannabis* strains are grown from clones or from a single parent seed, strains should be similar regardless of where the product was purchased. An organism’s

physical appearance (phenotype) is a product of both environmental factors and genetic makeup (genotype). Likewise, the chemical profile (chemotype) of *Cannabis* is determined by a combination of both genotype and environmental factors. While the chemical profile is inherited from the parents, the expression levels of the various chemicals within each plant can vary under different growing conditions (Cervantes 2006; Fischedick et. al 2010; Elzinga 2015; Jikomes and Zoorob 2018). To what extent the environment contributes to variation in the chemotype of *Cannabis* strains is largely unknown. Even when grown in controlled conditions, small but significant variation in mean THC/CBD ratios have previously been found among offspring of the same inbred line (de Meijer et al. 2002; de Meijer and Hammond 2005). As more people look to *Cannabis* for medical and recreational purposes, it is important to determine if there are inconsistencies observed in the chemotype within *Cannabis* strains.

Chemical Constituents of *Cannabis*

The number of isolated chemical constituents from *Cannabis* has increased from 423 in 1980 (Turner et al. 1980) to 490 in 2005 (ElSohly and Slade 2005), and there are currently more than 560 described chemical constituents (ElSohly et al. 2017). Chemical constituents found in *Cannabis* include approximately 120 phytocannabinoids, approximately 140 terpenes and an additional approximately 305 non-cannabinoids which include fatty acids, amino acids, carbohydrates and other chemical constituents (ElSohly et al. 2017). While terpenes are abundant in the plant kingdom, phytocannabinoids are rarely found outside the Cannabaceae, and are largely unique to *Cannabis sativa* (Small 2015b). Phytocannabinoids are a unique set of chemicals, which mimic compounds in the endocannabinoid system of many animals, excluding insects

(McPartland et al. 2001; Small 2015a). Glandular trichomes on the flowers of female plants are the main production site of phytocannabinoids and terpenes (Figure 5.1).

Figure 5.1. The glandular trichomes of the female flower of *Cannabis sativa* magnified 4X.

Cannabinoids. Anandamide (N-arachidonylethanolamine) is a neurotransmitter of the endocannabinoid system in humans. Anandamide binds to neuromodulatory CB_1 receptors in the central nervous system and the immunomodulatory CB_2 receptors in the peripheral nervous system (Small 2015a). The cannabinoids THC (Δ^9-tetrahydrocannabinol) and CBD (cannabidiol) mimic anandamide and 2-arachidonoylglycerol by binding to the CB_1 and CB_2 receptors (Mechoulam et al. 1995; Pertwee 2008). THC functions as an agonist by binding to and activating the receptor, while CBD is an agonist that binds to but does not activate the receptor. However, THC and CBD are not naturally produced in the *Cannabis* plant. Rather, *Cannabis* produces

THCA (Δ^9-tetrahydrocannabinolic acid) and CBDA (cannabidiolic acid), which are non-bioactive precursors to THC and CBD. Both THCA and CBDA can be converted to the bioactive forms through decarboxylation methods, which is often achieved through heating the flower material. Consumers are generally most interested in the effects of the active cannabinoid forms, and therefore *Cannabis* products that are not prepared for smoking or vaporizing have been processed prior to sale. Smoking and vaporizing are accomplished by applying a flame or heat to the *Cannabis* product and inhaling the smoke or vapors. Cannabinoids are fat soluble and infusing fat or oil allows the compounds to be more biologically available and absorbed more readily. 'Edibles' can be prepared using butter, oils, or sprays that have been infused with THC and/or CBD by heating flower material in oil or butter slowly over time. The four most widely recognized cannabinoids (THCA, CBDA, THC and CBD) are often reported separately or as a total calculated amount (see **Methods** Cannabinoids) on the labels of retail products as a percent of dry weight. Other cannabinoids and compounds, however, are produced in varying levels in *Cannabis,* and are only sometimes included on retail labels.

Medical and retail marijuana laws in Colorado as of January 2018 are mandated by the Colorado Department of Revenue Marijuana Enforcement Division. Testing protocols are published in the Colorado Department of Public Health and Environment's reference library. Retail Marijuana Cultivation Licensees are required to test two harvest batches for microbial contaminants every 30 days and potency once per quarter (Colorado Department of Revenue 2017). Cannabinoids are required to be listed as a percentage which represents an average of the results from all batch test samples (Colorado Department of Revenue 2017). Colorado requires analysis of five compounds

(THC, THCA, CBD, CBDA and cannabinol (CBN)) (Colorado Department of Revenue 2017).

Chemical analysis in *Cannabis* is generally conducted using gas chromatography (GC), high performance liquid chromatography (HPLC), mass spectroscopy (MS), or some variation of these assays such as HPLC- Diode Array Detector (HPLC-DAD) or High-Speed Liquid Chromatography (HSLC) (DeBacker et al. 2009, Hazecamp & Fischedick 2012, de Cássia Mariotti et al. 2015, Aizpurua-Olaizola et al. 2014, Chan 2014, Elzinga et al. 2015, Gul et al. 2015). Cannabinoids commonly analyzed include THCA, THC, CBDA, CBD, CBN, cannabigerolic acid (CBGA), cannabigerol (CBG), tetrahydrocannabivarin (THCV), and Δ^9-*trans*- cannabichromene (CBC), (Gul et al. 2015). Gas chromatography is less expensive than HPLC, but GC uses high temperatures and therefore is not an accurate way to measure levels of THCA and CBDA since those compounds are converted to THC and CBD when heated. Chemical analysis for commercial *Cannabis* in Colorado can be conducted by any licensed testing facility that adheres to state-mandated protocols, but variation in sample storage, testing equipment, and technicians may introduce variation.

Terpenes. Terpenes are a large group of chemicals that contribute to characteristics found in many herbal plants and essential oils. Terpenes are also partially responsible for some of the pharmacological properties of *Cannabis.* The terms terpene and terpenoid are often used interchangeably, although terpenes are basic hydrocarbons, while terpenoids contain additional functional groups. There are more than 120 identified terpenes found in *Cannabis* that are categorized as primary or secondary terpenes based on abundance. It is thought that terpenes work in concert with cannabinoids, and the

various levels and combinations of each in different varieties are responsible for the suite of pharmacological benefits reported for medicinal marijuana. This synergistic mechanism of action has been termed "the entourage effect" (Ben-Shabat et al. 1998). Terpenes are classified according to the number of repeating units of 5-carbon building blocks (isoprene units). The molecular structure of monoterpenes has 10 carbons, sesquiterpenes have 15 carbons, and triterpenes have 30 carbons. Yield and distribution of phytochemicals are the result of genetics, environmental conditions, and plant maturity (Meier and Mediavilla 1998). Monoterpenes dominate the volatile terpene profile of *Cannabis,* but there are a few common, sesquiterpenes. Research on *Cannabis* phytochemicals has traditionally focused on terpene and cannabinoid levels, because it is these phytochemicals that give the psychoactive effects, therapeutic benefits, and unique aromas associated with *Cannabis.*

Monoterpenes. Monoterpenes are the most common terpenes found in *Cannabis* (ElSohly 2007). α-Pinene (Figure 5.2A) is the most abundant terpene in the plant kingdom and is found in conifers, pines, sage, parsley, dill and basil (PCRlabs 2019). The therapeutics of α-pinene are anti-inflammatory, anti-osteoarthritic, and anti-nociception properties (PCRlabs 2019). β-Myrcene (Figure 5.2B) is found in lemongrass, basil, bay leaves, thyme, parsley, hops and tropical fruits (PCRlabs 2019). The therapeutics of β-myrcene are anti-inflammatory, analgesic, muscle relaxation, and sedative/hypnotic properties (PCRlabs 2019). D-limonene (limonene) (Figure 5.2C) is the second most abundant terpene and is found in the rinds of citrus fruits (PCRlabs 2019). The therapeutics of limonene are anti-inflammatory, analgesic, anti-depressant, muscle relaxation and sedative properties (PCRlabs 2019). Linalool (Figure 5.2D) is found in

flowers and spices such as lavender, rosewood, birch, and coriander (PCRlabs 2019). The therapeutics of linalool are anti-inflammatory, anesthetic, analgesic, anti-convulsant, anti-anxiety, and sedative properties (PCRlabs 2019).

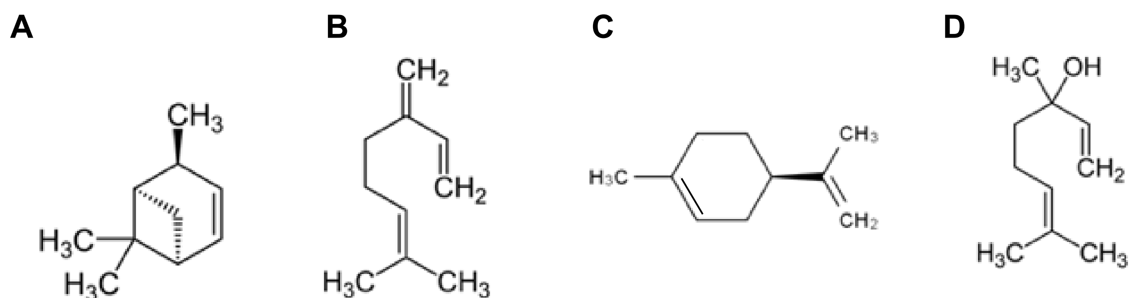


Figure 5.2. The chemical structures of four monoterpenes. (A) α-pinene), (B) β-myrcene, (C) D-limonene, and (D) linalool.

Sesquiterpenes. Sesquiterpenes are also found in *Cannabis* but are less prevalent (ElSohly 2007). β-Caryophyllene (Figure 5.3A) is found in many spices and plants such as clove, cinnamon, black pepper, hops, oregano and basil (PCRlabs 2019). Therapeutics of β-caryophyllene are anti-inflammatory and analgesic effects. α-Caryophyllene (humulene) (Figure 5.3B) is found in spices and herbs such as clove, basil, hops, sage, spearmint, ginseng, as well as some fruits and vegetables (PCRlabs 2019). Therapeutics of α-caryophyllene are anti-inflammatory and analgesic effects (PCRlabs 2019).

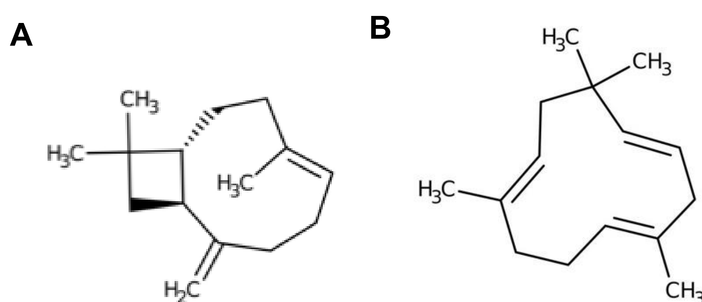


Figure 5.3. The chemical structures of two sesquiterpenes. (A) β-caryophyllene, and (B) α-caryophyllene.

Summary

This study aimed to investigate if genetic variation in neutral genetic markers is paralleled with variation in nine cannabinoids and 21 terpenes. The genotypes from four strains totaling 15 samples were previously assessed for genetic variation (Chapter IV). Samples from an additional three strains were added to the terpene analysis to examine terpene variation among strains. Cannabinoid profiles were analyzed using HPLC-DAD, and terpenes were analyzed using GC. To determine phytochemical variation in retail *Cannabis* strains, this study examined variation (1) among samples that are genetically identical (2) within strains, and (3) among different strains. Variation within strains was compared to genetic variation to assess whether genetic variation is associated with cannabinoid and terpene variation.

Methods

The genetic assessment (see Chapter IV) generated a set of 15 *Cannabis* samples selected from four strains: "Blue Dream" (5), "OG Kush" (4), "Mob Boss" (3), and "Durban Poison" (3). All three "Durban Poison" samples had identical genotypes, while one sample each of "Blue Dream" (3SN), "OG Kush" (1SN), and "Mob Boss" (5SN) was identified as having a unique genotype from the others, which were genetically cohesive within the strain. Samples were purchased from 11 dispensaries licensed by the state of Colorado for retail recreational sales located in Denver, Fort Collins, and Garden City. The coded, dried cannabis flower samples (1 g) from the sensory study (Table 11; Chapter IV) were sent to Mile High Labs (Loveland, Colorado) for blind analysis.

Cannabinoids

Potency and related cannabinoid analysis of raw material were reported for nine cannabinoids (THC, THCA, CBD, CBDA, CBG, CBGA, THCV, CBN, and CBC). Mile High Labs analyzed the samples following standard protocols with a gradient HPLC system using a reverse-phase column and guard column with a C18 stationary phase. A 1260 Infinity II HPLC with DAD detection at 240 nm, (bandwidth 4 nm) and reference 360 nm (bandwidth 100 nm) was used per Colorado state compliance requirements (Table 5.1 - 5.3).

Table 5.1. HPLC Method Parameters

<table>
<tr><th colspan="5">HPLC Information</th></tr>
<tr><td>Column Type:</td><td colspan="4">Restek Raptor ARC-18, 2.7 mm, 150 x 4.6 mm, PN 9314A65</td></tr>
<tr><td>Guard Cartridge:</td><td colspan="4">Restek Raptor ARC-18, 2.7 mm, 5 x 4.6 mm, PN 9314A0250</td></tr>
<tr><td>Column Temperature:</td><td colspan="4">35 °C</td></tr>
<tr><td>Sample Tray Temperature:</td><td colspan="4">Ambient</td></tr>
<tr><td>Flow Rate:</td><td colspan="4">1.5 mL/min</td></tr>
<tr><td>Stop Time:</td><td colspan="4">12 min</td></tr>
<tr><td>Post Time:</td><td colspan="4">3 min</td></tr>
<tr><td>Method Type:</td><td colspan="4">Gradient</td></tr>
<tr><td>MP A:</td><td colspan="4">0.015% formic acid in water</td></tr>
<tr><td>MP B:</td><td colspan="4">0.010% formic acid in acetonitrile</td></tr>
<tr><th colspan="3">Gradient</th><th colspan="2">DAD Detector Settings</th></tr>
<tr><td>Time</td><td>MP A (%)</td><td>MP B (%)</td><td>Detection:</td><td>240m nm</td></tr>
<tr><td>0:00</td><td>30</td><td>70</td><td>Detection BW:</td><td>4 nm</td></tr>
<tr><td>2:00</td><td>30</td><td>70</td><td>Reference:</td><td>360 nm</td></tr>
<tr><td>8:00</td><td>5</td><td>95</td><td>Reference BW:</td><td>100 nm</td></tr>
<tr><td>12:00</td><td>5</td><td>95</td><td>Collect UV Spectra:</td><td>190 -400 nm</td></tr>
<tr><td colspan="3" rowspan="2"></td><td>Peakwidth:</td><td>> 0.05 min (2.5 Hz)</td></tr>
<tr><td>Slit Width:</td><td>4 nm</td></tr>
<tr><th colspan="5">Injector Settings</th></tr>
<tr><td>Injection Volume</td><td colspan="4">10 µL</td></tr>
<tr><td>Needle Wash</td><td colspan="4">Methanol</td></tr>
</table>

Table 5.2. Injection Sequence

Description	# of Injections	Parameter
Reagent Blank	NLT 1	System equilibration
Reagent Blank	1	Contamination check and non-interference
RL Standard	1	Reporting Limit Recovery
Resolution Standard	1	Resolution
Standard A	6	• CBD / ISTD peak area ratio • Mean CBD USP Tailing Factor • Mean CBD USP Theoretical Plates
Standard B	1	CBD Percent Recovery
Samples	1	Sample analysis
Standard A*	1	Check Standard Recovery

*Inject standard A as a drift check at least every six samples and at the end of each sequence

Table 5.3. Peak Relative Retention Time (RRT) and Relative Response Factor (RRF) Values

Cannabinoids	Approx. retention time	Approx. RRT	RRF
tetrahydrocannabinolic acid (THCA)	7.28	1.69	0.66
tetrahydrocannabinol (Δ9-THC)	6.38	1.48	0.89
cannabidiolic acid (CBDA)	3.67	0.85	0.71
cannabidiol (CBD)	4.3	1	1
cannabigerolic acid (CBGA)	3.96	0.92	1
cannabigerol (CBG)	4.15	0.97	1.02
tetrahydrocannabivarin (THCV)	4.46	1.04	1
cannabinol (CBN)	5.64	1.31	0.33
Δ9-trans- cannabichromene (CBC)	7.07	1.64	0.32

The analytical column used was a Restek Raptor ARC-18, 2.7 µm, 150 x 4.6 mm, PN 9314A65. The guard column was a Restek Raptor ARC-18, 2.7 µm, 5 x 4.6 mm, PN 9314A0250 and Restek EXP Direct Connect Holder PN 25808. HPLC-grade acetonitrile (CAS# 75-050-8), methanol (CAS# 76-56-1) and water (CAS# 7732-18-5) were used, as well as GC grade di-n-octyl phthalate (CAS# 117-84-0), reagent grade formic acid (CAS# 64-18-6), Cerilliant 1 mg/mL Cannabigerol Ampoule PN C-141 (dissolved in

methanol), Cerilliant 1 mg/mL Cannabidiol Ampoule PN C-045 (dissolved in methanol), and Cerilliant © Certified Cannabinoid Standards for ID (Redrock, Texas): C-045, C-046, C-144, T-005, C-140, C-141, T-093, C-143, C154, C171, C-150, C-152, T-032, C-153, T-094, and C-142. A 10ul aliquot of composite cannabinoid sample was injected using Acq. Method TM-001 – 1260.amx and Processing method *TM-002UV.pmx with no manual modification (Figure 5.4, Table 5.4)

Figure 5.4. Composite HPLC-DAD chromatogram.

Table 5.4. Retention Times, Peak Height, Peak Area and Response Factor for the composite HPLC-DAD Analysis.

Signal: DAD1A,Sig=240,4 Ref=360,100

Name	Tailing	Plates	RT (Min)	Peak Height (mAU)	Area (mAU*sec)	Res (USP)
	0.90	18093.00	1.56	0.68	1.17	
	1.00	15160.00	1.62	1.03	2.00	1.10
	1.00	10827.00	1.73	0.80	2.24	1.90
	1.40	4192.00	1.96	1.24	5.21	2.40
	1.00	14444.00	2.08	1.69	4.34	1.20
	0.80	11883.00	2.29	4.34	14.73	2.80
	1.20	17393.00	2.42	105.22	293.92	1.60
CBDV	1.00	16971.00	2.68	74.34	230.00	3.40
	1.10	21209.00	3.10	0.68	2.20	5.00
	1.40	17392.00	3.27	1.00	4.13	1.80
CBDA	1.10	23925.00	3.58	79.60	278.90	3.20
	1.00	34299.00	3.76	1.74	5.17	2.10
CBGA	1.10	29880.00	3.88	86.21	291.14	1.40
CBG	1.00	30199.00	4.04	64.32	224.49	1.80
CBD	1.00	33788.00	4.18	62.61	212.93	1.60
THCV	1.00	39558.00	4.34	79.61	260.05	1.70
	2.00	63140.00	5.00	0.38	1.42	8.00
	1.10	68243.00	5.32	0.55	1.67	4.00
CBN	1.10	74373.00	5.51	242.42	737.08	2.30
	1.10	90899.00	6.04	0.39	1.13	6.60
	0.80	92846.00	6.25	95.07	264.29	2.60
THC	1.80	38836.00	6.32	155.45	749.53	0.60
	1.00	121577.00	6.80	125.48	367.14	4.70
CBC	1.00	136042.00	6.94	245.48	701.31	1.90
THCA	1.20	122548.00	7.18	87.26	273.92	3.00
	1.50	142309.00	7.56	292.93	902.20	4.70
	1.30	114404.00	7.67	73.85	255.63	1.40
	1.50	27536.00	7.95	1.14	10.58	1.90
	0.90	149960.00	8.39	0.31	1.33	3.20

Mobile phase A (0.015% formic acid in water) was prepared by combining 150 μL of formic acid with 1 L of water. Mobile phase B (0.010% formic acid in acetonitrile) was prepared by mixing 100 μL of formic acid with 1 L of acetonitrile. Stock Standard

solutions were stored at -80°C for < 88 days, and the Internal Standard (ISTD), A/B Standard solutions and RL solutions were stored at -80° C for < 7 days. The Sample ISTD was prepared by diluting 200.00 ± 20 mg di-n-octyl phthalate in methanol to a total volume of 100.0-mL. The Standard ISTD was prepared by adding methanol to 5.0 mL of the Sample ISTD to a final volume of 20 mL. Standard A was prepared with 100.00 ± 10 mg of Cannabidiol Isolate Standard to a sufficient volume of methanol in a 100-mL volumetric flask to just below the QS line and sonicated for 2 minutes. The solution was then allowed to equilibrate to room temperature and diluted to volume with methanol. In order to create the Standard A Preparation with a nominal concentration of 50 μg/mL cannabidiol, 5.0 mL of the sonicated Cannabidiol Isolate Standard in methanol was diluted with 5.0 mL and brought to a final volume of 100.0 mL with methanol. Standard B was prepared with 100.00 ± 10 mg of Cannabidiol Isolate Standard to a sufficient volume of methanol in a 100-mL volumetric flask to just below the QS line and sonicated for 2 minutes. The solution was then allowed to equilibrate to room temperature and diluted to volume with methanol. In order to create the Standard B Preparation with a nominal concentration of 50 μg/mL cannabidiol, 5.0 mL of the sonicated Cannabidiol Isolate Standard in methanol was diluted with 5.0 mL and brought to a final volume of 100.0 mL with methanol. Standard A/B are prepared to ensure standards were made correctly. If Standard B recovery fell within a 4% range of Standard A then all quantitation was determined to be accurate using Standard A during sample processing. 5.0 mL of Stock Standard A was diluted to a volume of 100.0 mL with methanol to prepare the RL Stock Standard I Preparation with a nominal concentration of 50 μg/mL cannabidiol. A 5.0 mL aliquot of Stock RL Standard was diluted to a volume of 100.0

mL with methanol to prepare the RL Stock Standard II Preparation with a nominal concentration of 2.5 μg/mL cannabidiol. A 10.0 mL aliquot of RL Stock Standard II and 5.0 mL of the Standard ISTD was diluted to a volume of 100.0 mL with methanol to prepare the RL Standard Preparation with a nominal concentration of 0.25 μg/mL cannabidiol. The Resolution and RC Identification Standard was prepared by adding ~1.0 mL of each Cerilliant © Certified Cannabinoid Standards (Redrock, Texas) to a 20.0 mL volumetric flask and diluted to volume with methanol. A 10 uL aliquot of RL standard was injected and Acq. Method TM-001 – 1260.amx and Processing method *TM-002UV.pmx were used with no manual modification (Figure 5.5, Table 5.5).

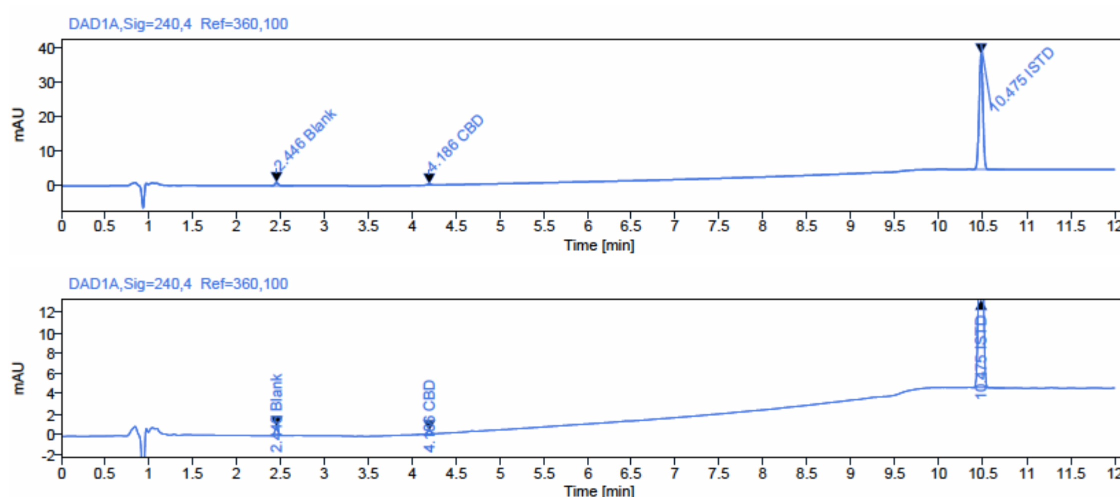


Figure 5.5. RL standard HPLC-DAD chromatogram.

Table 5.5. Retention Times, Peak Height, Peak Area and Response Factor for the RL standard HPLC-DAD analysis.

Signal: DAD1A,Sig=240,4 Ref=360,100

Name	Tailing	Plates	RT (Min)	Peak Height (mAU)	Area (mAU*sec)	Res (USP)
Blank	1.10	18713.00	2.45	0.87	2.47	
CBD	1.20	34891.00	4.19	0.30	1.02	21.60
ISTD	0.90	271187.00	10.48	33.46	101.04	74.00

Cannabis samples were prepared by weighing 200.00 ± 10 mg of dried flower material and transferred to a 100 mL volumetric flask, and 5.0 mL of the Sample ISTD was added. To prepare a Potency and Related Cannabinoid Stock Sample Preparation with a nominal concentration of 2000 µg/mL, a sufficient amount of methanol was added to bring the volume to just below the QS line and the solution was sonicated at 40° C for 10 minutes. The solution was mixed thoroughly then brought to room temperature. The Stock Sample Solution was then filtered into a scintillation vial using a Pall Acrodisc CR13 0.2-µm PTFE syringe filter. The first ~1 mL of filtrate was discarded followed by collection of > 5.5 mL of filtrate. The final Potency and Related Cannabinoid Sample Preparation with a nominal concentration of 500 µg/mL was prepared by dilution 5.0 mL of Stock Sample Solution to 20.0 mL with methanol. A 10 uL aliquot of sample extraction (For example 752, OG Kush 2 Figure 5.6, Table 5.6) was injected using Acq. Method TM-001 – 1260.amx, Processing method *TM-002UV.pmx with no manual modification.

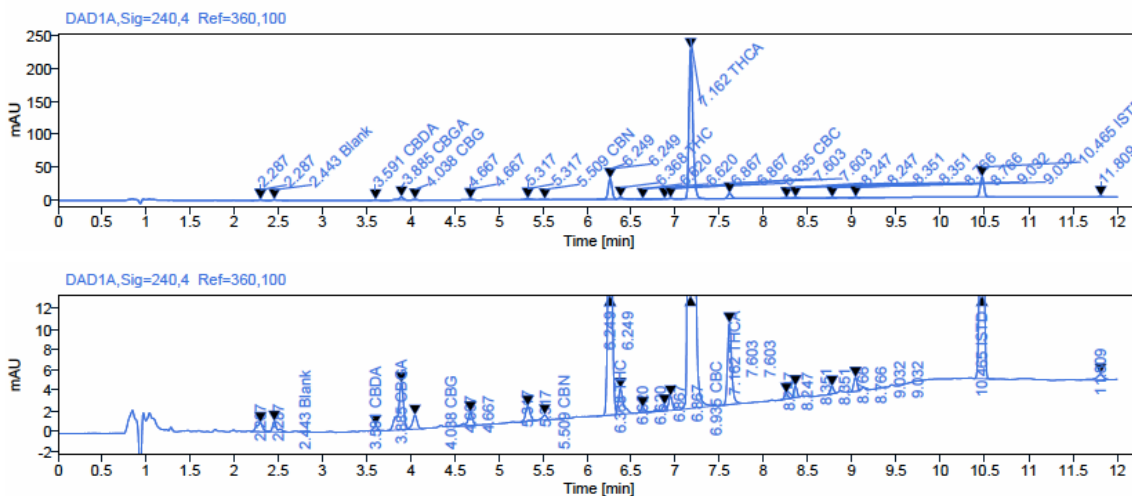


Figure 5.6. Sample HPLC-DAD chromatogram for sample 752 (OG Kush 2).

Table 5.6. Retention Times, Peak Height, Peak Area and Response Factor for the sample 752 (OG Kush 2) HPLC-DAD analysis.

Signal: DAD1A,Sig=240,4 Ref=360,100

Name	Tailing	Plates	RT (Min)	Peak Height (mAU)	Area (mAU*sec)	Res (USP)
	1.10	5418.00	2.29	0.88	4.32	
Blank	1.60	17959.00	2.44	1.04	3.81	1.60
CBDA	1.50	27959.00	3.59	0.50	1.71	14.50
CBGA	0.80	28583.00	3.89	4.61	20.09	3.30
CBG	1.00	28834.00	4.04	1.35	5.02	1.60
	0.80	45876.00	4.67	1.30	5.05	6.90
	1.50	51644.00	5.32	1.48	5.25	7.20
CBN	1.00	57983.00	5.51	0.48	1.70	2.10
	1.10	99474.00	6.25	31.11	93.18	8.70
THC	1.80	77805.00	6.37	2.52	9.92	1.40
	1.40	98167.00	6.62	0.50	1.62	2.80
	0.60	103677.00	6.87	0.58	2.26	2.90
CBC	1.20	132966.00	6.94	1.39	3.99	0.80
THCA	1.30	124679.00	7.16	227.63	713.80	2.90
	1.30	131948.00	7.60	8.00	26.04	5.30
	1.10	200307.00	8.25	0.54	1.46	8.20
	1.30	225435.00	8.35	1.13	2.93	1.40
	1.10	245819.00	8.77	0.73	1.90	5.90
	1.10	223202.00	9.03	1.36	3.93	3.60
ISTD	0.90	271602.00	10.47	30.46	91.75	18.30
	0.80	127297.00	11.81	0.54	3.12	12.60

The Reagent Blank Solution was prepared by following the steps for sample preparation but excluding the addition of the dried flower material (Figure 5.7, Table 5.7). Each sample was extracted in triplicate to ensure consistency and ensure quantitation was accurate.

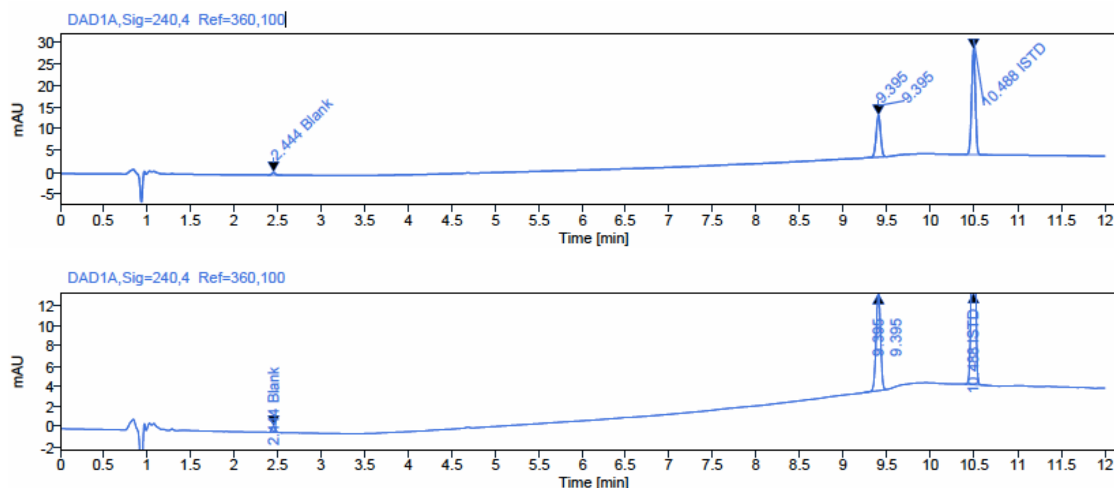


Figure 5.7. HPLC-DAD chromatogram for the blank.

Table 5.7. Retention Times, Peak Height, Peak Area and Response Factor for the blank HPLC-DAD analysis.

Signal :DAD1A,Sig=240,4 Ref=360,100

Name	Tailing	Plates	RT (Min)	Peak Height (mAU)	Area (mAU*sec)	Res (USP)
Blank	1.00	19083.00	2.44	0.63	1.67	
	1.00	154781.00	9.40	9.66	34.88	83.60
ISTD	1.00	270998.00	10.49	24.36	73.85	12.40

Many of the major cannabinoids (THC, CBD, CBG, CBC, but not CBN) have the molecular formula $C_{21}H_{30}O_2$. However, the acidic forms (THCA, CBDA, CBGA and CBCA) are the naturally occurring form manufactured by the trichomes in the flowers. The acidic forms have an additional CO_2 and chemical formula $C_{22}H_{30}O_4$. In order to calculate the THC_{TOTAL}, a conversion factor must be used to account for the CO_2 molecule removed during decarboxylation. The molecular weight of the acidic cannabinoids is 358.48 and the neutral cannabinoids is 314.47, therefore 12.28% THCA is lost in the form of $CO_2(g)$ during the decarboxylation process. The conversion factor to account for this loss is 0.877 (314.47/358.48). This formula can be applied to calculate THC_{TOTAL},

CBD_{TOTAL}, CBG_{TOTAL} and CBC_{TOTAL}. The conversion factor can be used to calculate the conversion of acidic cannabinoid forms when they lose a CO_2 molecule (decarboxylation) and are converted to the neutral bioactive form (Equation 5.2 & 5.3).

THC_{TOTAL} (%) = % THC + (% THCA x 0.877) Equation 5.1.

CBD_{TOTAL} (%) = % CBD + (% CBDA x 0.877) Equation 5.2.

Total Cannabinoids (%) = Σ individual related cannabinoids ≤ 0.05% Equation 5.3.

Dispensaries may report %THC and %THCA as separate values or $\%THC_{TOTAL}$. The sample packaging used in this study did not disclose analysis methods; although we assume if %THC and %THCA were reported, HPLC-DAD analysis was conducted. Since both HPLC-DAD and GC are acceptable analyses to use to determine potency in retail cannabis, the $\%THC_{TOTAL}$ can be calculated from HPLC-DAD analysis.

The percent dry weight of nine cannabinoids in the fifteen samples were compiled into a table using Microsoft Excel and bar plots for the cannabinoid profiles were graphed in Excel. For each sample, bar plots were generated for (1) the percent dry weight for the full cannabinoid panel, (2) the proportions of cannabinoids within the total cannabinoid fraction, (3) the percent dry weight for THCA and THC, and (4) the percent dry weight for CBDA, CBG, CBGA, THCV, CBN, and CBC. None of the samples possessed measurable amounts of CBD; therefore CBD was not included in any of the subsequent analyses. Histograms of percentages by dry weight of the average level of THCA in the consensus sequences and the genetic outlier were generated in Microsoft Excel. Histograms of average level of the remaining cannabinoids in the consensus sequences

were graphed with standard deviation error bars against the cannabinoid levels of the genetic outlier. Microsoft Excel was also used to generate a scatter plot to determine the relationship between the two cannabinoids (THCA and CBDA) representing the majority of the variation in the data.

A hierarchical cluster analysis dendrogram for eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, THCA) in 15 samples using Ward's method and Euclidean distance parameters was conducted in PC-ORD (McCune and Mefford 1999). However, because the investigation was examining variation within strains, cannabinoid dendrograms for each strain were also generated to assess the relationship between the genetic outlier and the consensus samples. Since THCA had overwhelmingly higher levels than the other cannabinoids, a second clustering analysis was conducted with THCA and THC removed from the data set. Dendrograms were generated for the full samples set and for each strain. PC-ORD was also used to generate a Principal Components Analysis (PCA) for eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, THCA) using Ward's method and Euclidean distance parameters. Since THCA had overwhelmingly higher levels than the other cannabinoids, a second PCA was conducted with THCA and THC removed from the data set.

Terpenes

The terpene profiles for 21 *Cannabis* samples from seven strains were determined using a GC-MS. Mile High Labs analyzed the samples following standard protocols with reference standards to calibrate the GC-MS system. Samples were prepared by weight in ACS-grade methanol. Preparation concentrations were from approximately 10-15% by weight. All weights were obtained to within 0.1 mg. Dry sample material was crushed to

a fine solid, suitable for weighing and transfer. Tare weights were obtained for the crimp-sealed vials, and then the sample was added and weighed followed by addition of methanol with subsequent weighing. Samples were then placed into a 50-55°C sonicating bath for 15 minutes followed by vortexing while still warm and allowed to settle at room temperature. Once cooled, the liquid was decanted via pipette and syringe filtered through a 0.45 µm filter into GC vials for analysis. An Agilent 6890 GC with a 5973 inert MS detector in EI mode was used for analysis. The calibration curves for components were created using the 21-component terpene standard from RESTEK (data not provided). A five- point calibration was used for standards. The column used was a RESTEK Rxi-624 and helium was used as the carrier gas. A computing integrator recorded the chromatograph and mass spectrometer analysis, and relative retention times of terpene peaks from the reference standards were used to identify and determine presence and average abundance of terpenes in samples of seven strains. Two replicates for each sample were analyzed. The peaks were then calculated as normalized area percentages of each of the terpenes present in the samples. Terpene distribution and relative proportions for 21 terpenes was assessed. The terpene levels were used to assess the relationship between genetic consensus samples and a genetic outlier within a strain, as well as terpene variation among different strains.

PCA eigenvalues generated in PC-ORD for 21 terpenes using Ward's method and Euclidean distance parameters were used to plot the PCA in RStudio with the ggplot package (R Studio Team 2015). Histograms of peak area percentages of each terpene present in the strain samples were generated in Microsoft Excel. The average terpene levels of the genetic consensus samples were graphed with standard deviation error bars

against the terpene levels of the genetic outlier. The five highest terpene of the genetic consensus samples were pooled for “Blue Dream” (12), “Mob Boss” (12), and “OG Kush” (10), and compared to terpene levels of the genetic outlier. Twenty-one terpenes levels for seven strains and the average level each terpene was reported when multiple samples were included. Finally, the terpenes with the highest average standard deviation were graphed for the seven strains in order to assess terpene variation among strains.

Results

Cannabinoids

All the samples used in this investigation had high THC concentrations (7.97 – 22.70 % dry weight) (Table 5.8). Five samples had a detectable level of CBDA (0.04294 – 0.06520 % dry weight), while CBD was not detected in any sample and was not included in the results (Table 5.8).

Table 5.8. Percentages of eight cannabinoid levels in four strains by HPLC-DAD as percent dry weight for eight cannabinoids, and calculated THC_{TOTAL} (Equation 5.1). Identification abbreviations are listed for each sample and the genetic outliers are indicated with an asterisk and highlighted in gray.

Strain	Identification	THCA	THC	CBDA	CBGA	CBG	THCV	CBN	CBC
Durban Poison	1SN	12.73	0.42	0.00	0.32	0.00	0.00	0.04	0.09
	4SN	12.28	0.04	0.00	0.76	0.09	0.00	0.00	0.02
	5SN	17.52	0.27	0.05	0.79	0.07	0.05	0.00	0.00
mean ± SD	Consensus	14.17 ± 2.90	0.24 ± 0.19	0.02 ± 0.03	0.62 ± 0.26	0.05 ± 0.05	0.02 ± 0.03	0.01 ± 0.02	0.04 ± 0.05
Blue Dream	1SN	14.48	0.10	0.00	0.12	0.00	0.00	0.00	0.00
	4SN	8.85	0.21	0.00	0.14	0.00	0.00	0.00	0.00
	5SN	12.72	0.20	0.00	0.11	0.00	0.00	0.00	0.00
	6SN	17.11	0.25	0.04	0.13	0.00	0.00	0.00	0.00
mean ± SD	Consensus	13.29 ± 3.47	0.19 ± 0.06	0.01 ± 0.02	0.12 ± 0.01	0.00	0.00	0.00	0.00
	3SN*	15.07	0.15	0.00	0.11	0.00	0.00	0.00	0.02
Mob Boss	1SN	11.88	0.24	0.00	0.96	0.16	0.05	0.04	0.04
	3SN	15.05	0.27	0.00	1.41	0.08	0.00	0.00	0.00
mean ± SD	Consensus	13.47 ± 2.25	0.26 ± 0.02	0.00	1.19 ± 0.32	0.12 ± 0.05	0.03 ± 0.04	0.02 ± 0.03	0.02 ± 0.03
	5SN*	11.54	0.31	0.00	0.21	0.12	0.00	0.05	0.04
OG Kush	2SN	25.34	0.47	0.07	1.08	0.28	0.00	0.03	0.07
	3SN	20.52	0.07	0.04	0.74	0.11	0.00	0.00	0.00
	4SN	15.24	0.23	0.05	0.21	0.08	0.00	0.04	0.02
mean ± SD	Consensus	20.37 ± 5.05	0.26 ± 0.20	0.05 ± 0.01	0.68 ± 0.44	0.15 ± 0.11	0.00	0.02 ± 0.02	0.03 ± 0.04
	1SN*	12.59	0.18	0.00	0.39	0.07	0.00	0.02	0.00

To compare cannabinoid composition and quantity among samples within strains, histograms of eight cannabinoids were generated to show variation (Figure 5.8). The "Durban Poison" samples were genetically identical, and yet the cannabinoid levels varied among all samples. When the total cannabinoid fraction (% by dry weight) with relative abundance of the chemotype by proportion was graphed, THCA levels appear to be similar, but levels of THC, CBGA and CBC seem to be the highest contributors to the variation (Figure 5.9). Samples 1SN and 4SN had similar levels of THC (Table 5.6, Figure 5.10A) and 5SN had higher THC. Examination of the other cannabinoids showed 1SN and 4SN differ substantially in both composition and relative levels of CBGA, CBG, CBC, and CBN, while 4SN and 5SN had similar levels of CBGA and CBG but differed in CBC, CBDA and THCV (Figure 5.11A). PCA clustering analysis using data from eight cannabinoids produced two large clusters, with assignment of 1SN and 4SN to one cluster and 5SN to the other (Figure 5.14), indicating 5SN is more distinct in cannabinoids than1SN and 4SN are from each other. This is supported in the clustering analysis containing only "Durban Poison" samples where 1SN and 4SN clustered together (Figure 5.15A). There is a possibility that the less prevalent minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV) may be a better reflection of the genetic relationships among the samples in drug-type strains. Therefore, additional PCAs of the minor cannabinoids were generated (Figure 5.13 and 5.17A). Although the "Durban Poison" samples were still assigned to different clusters (Figure 5.16), and "Durban Poison" 4SN and 5SN clustered together while 1SN was the chemical deviant (Figures 5.16 and 5.17A).

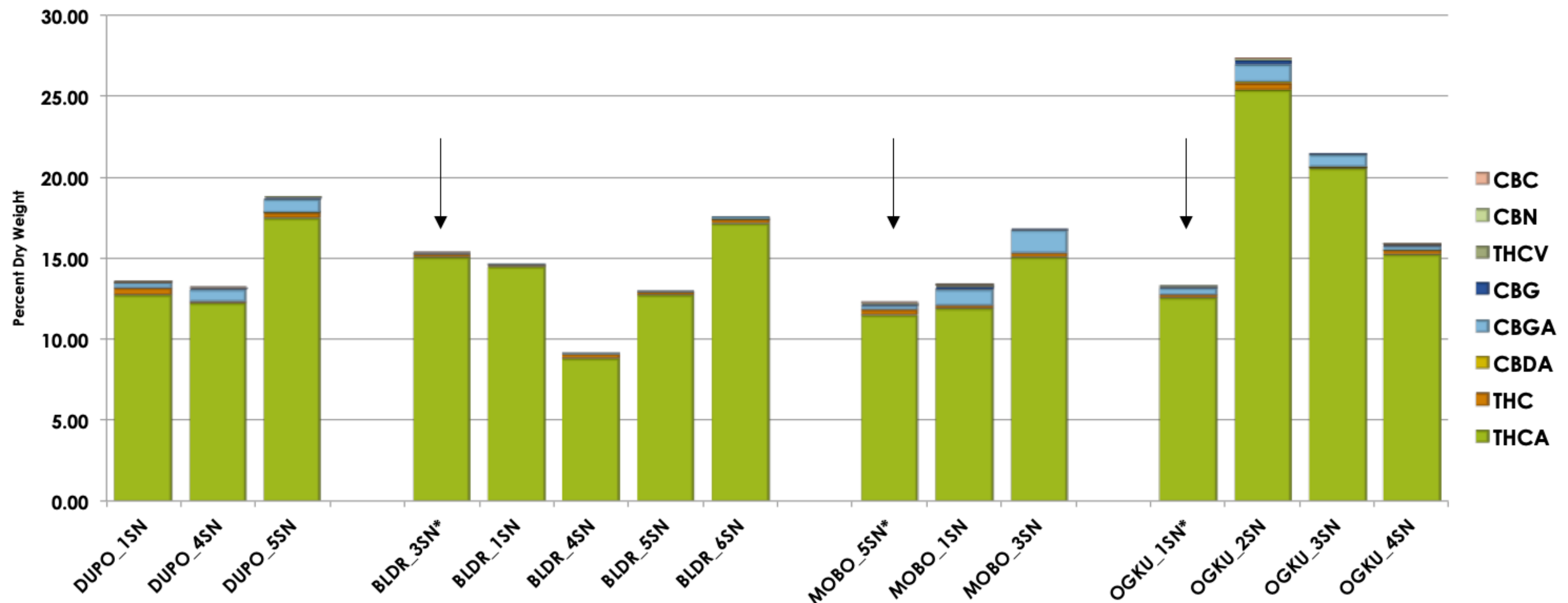


Figure 5.8. Percent dry weight (mg) of eight cannabinoids measured by HPLC-DAD in 15 samples of four commercially available strains. Arrows and asterisks indicate the genetic outlier.

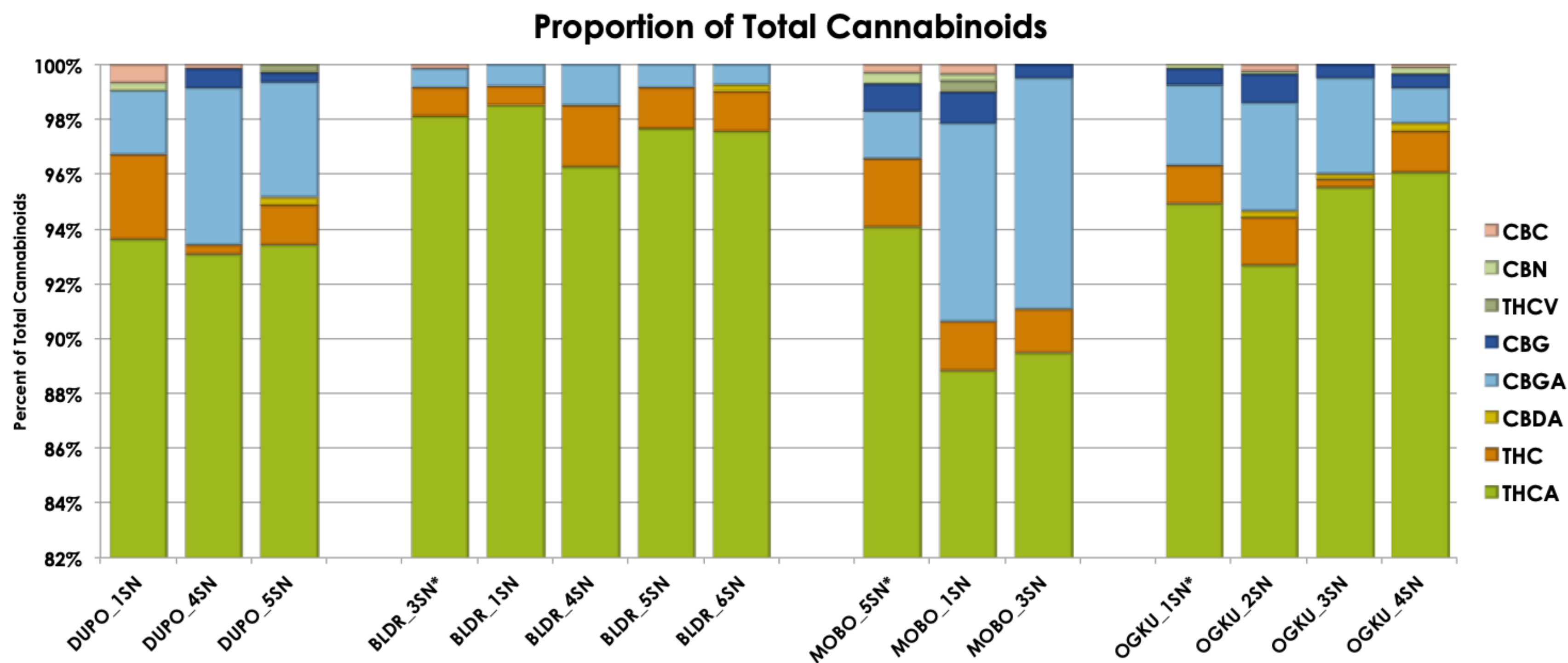


Figure 5.9. Chemotype by proportions of detected cannabinoids within the total cannabinoid fraction (% by dry weight) measured by HPLC-DAD in 15 samples of four commercially available strains. Asterisks indicate genetic outlier.

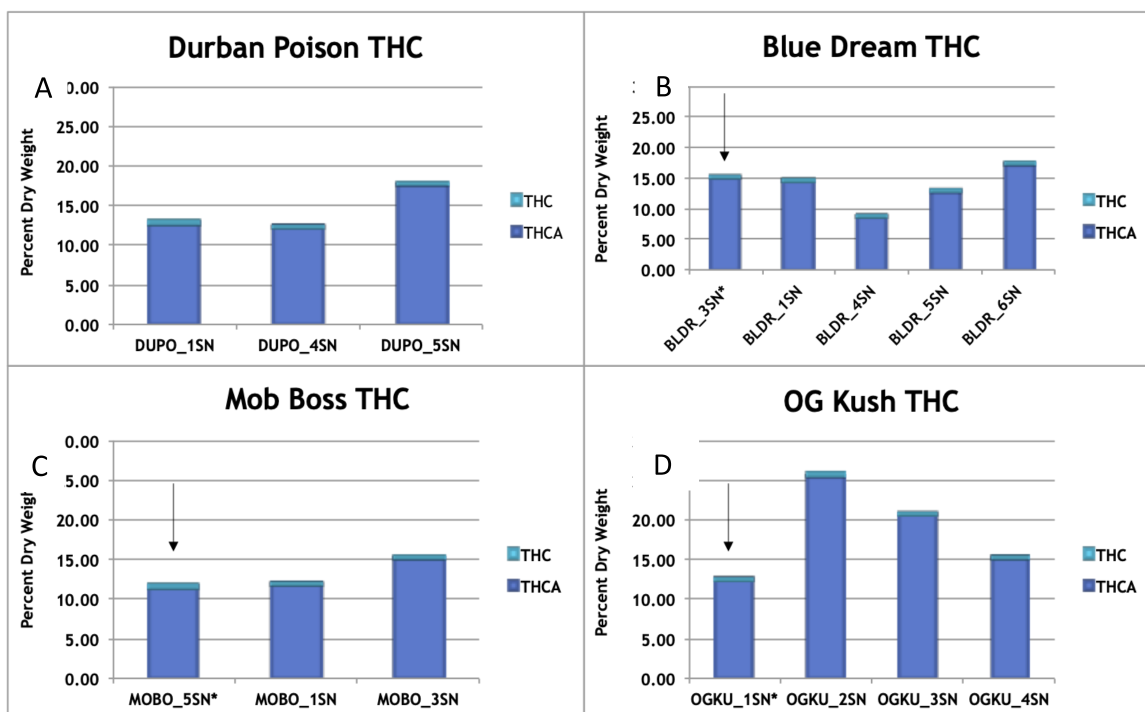


Figure 5.10. Percent dry weight (mg) of THC and THCA (six cannabinoids excluded). Arrows and asterisks indicate the genetic outlier.

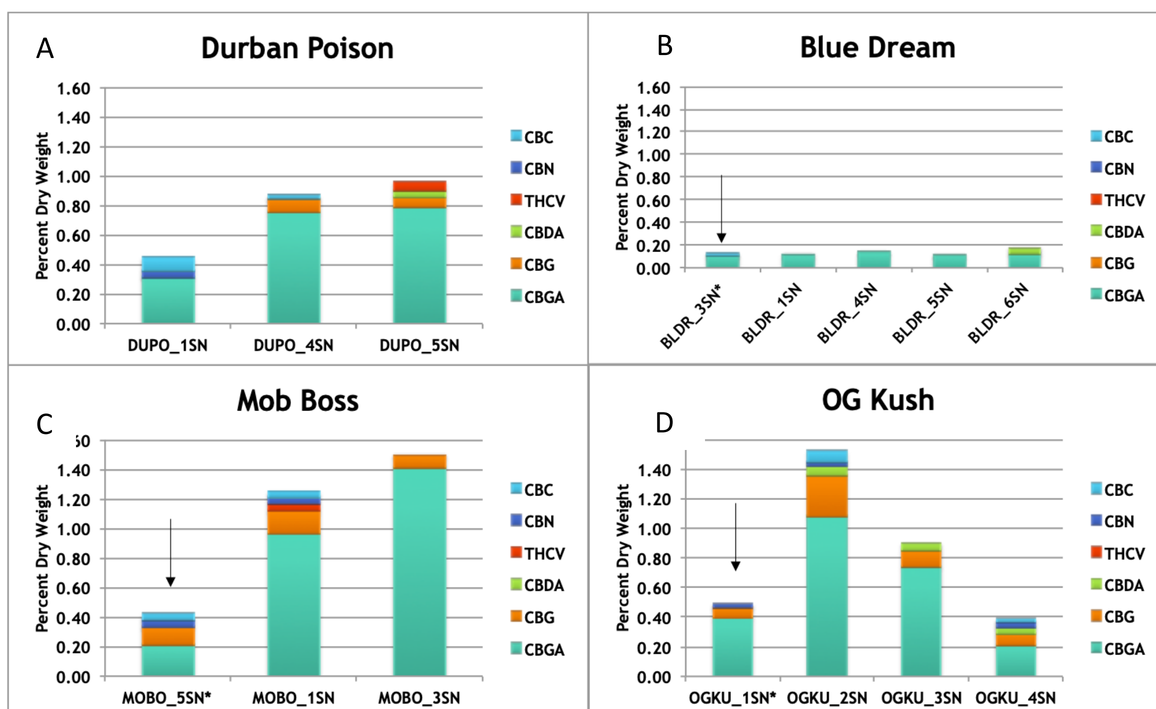


Figure 5.11. Percent dry weight (mg) of six minor cannabinoids (THC and THCA excluded). Arrows and asterisks indicate the genetic outlier. Predominant minor cannabinoids are CBGA and CBG.

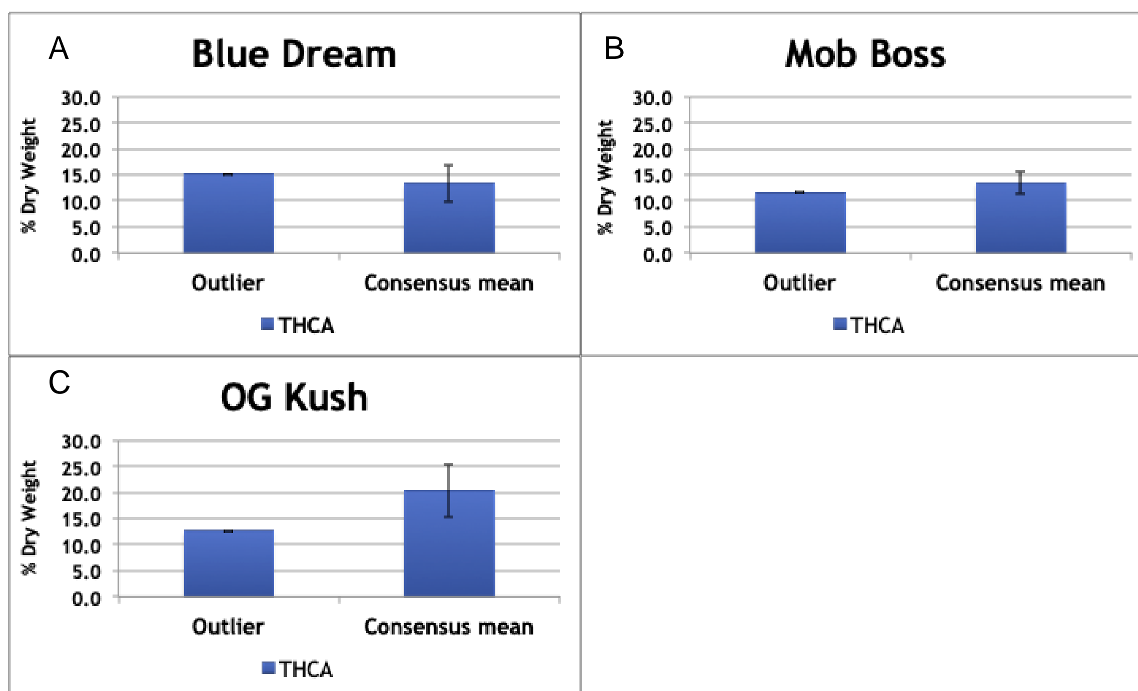


Figure 5.12. Average levels of THCA (total % dry weight) in the genetic consensus samples compared to the genetic outlier. Error bars represent the standard deviation.

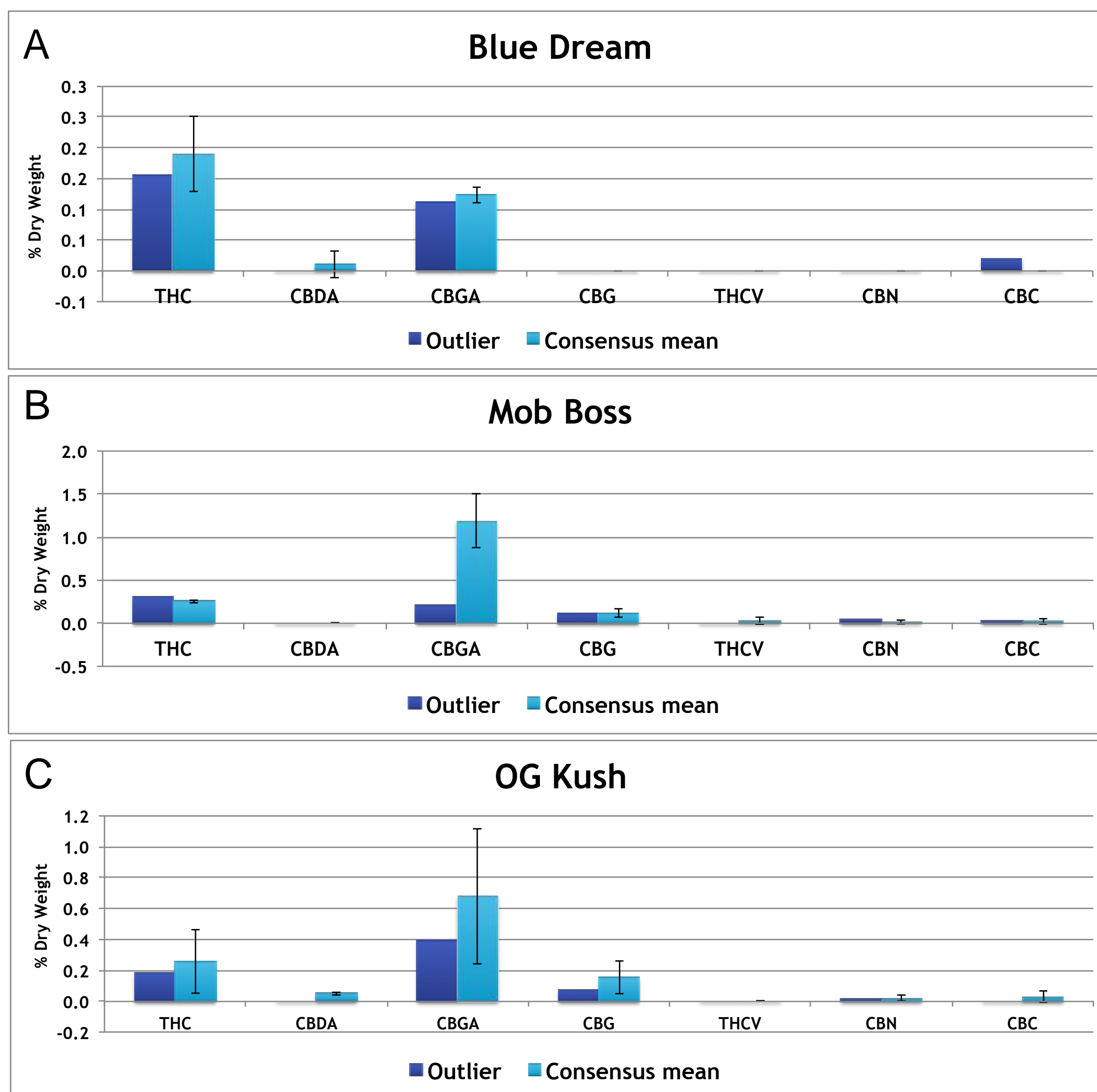


Figure 5.13. Average levels of the minor cannabinoids (total % dry weight) in the genetic consensus samples compared to the levels of the genetic outlier. Error bars represent the standard deviation.

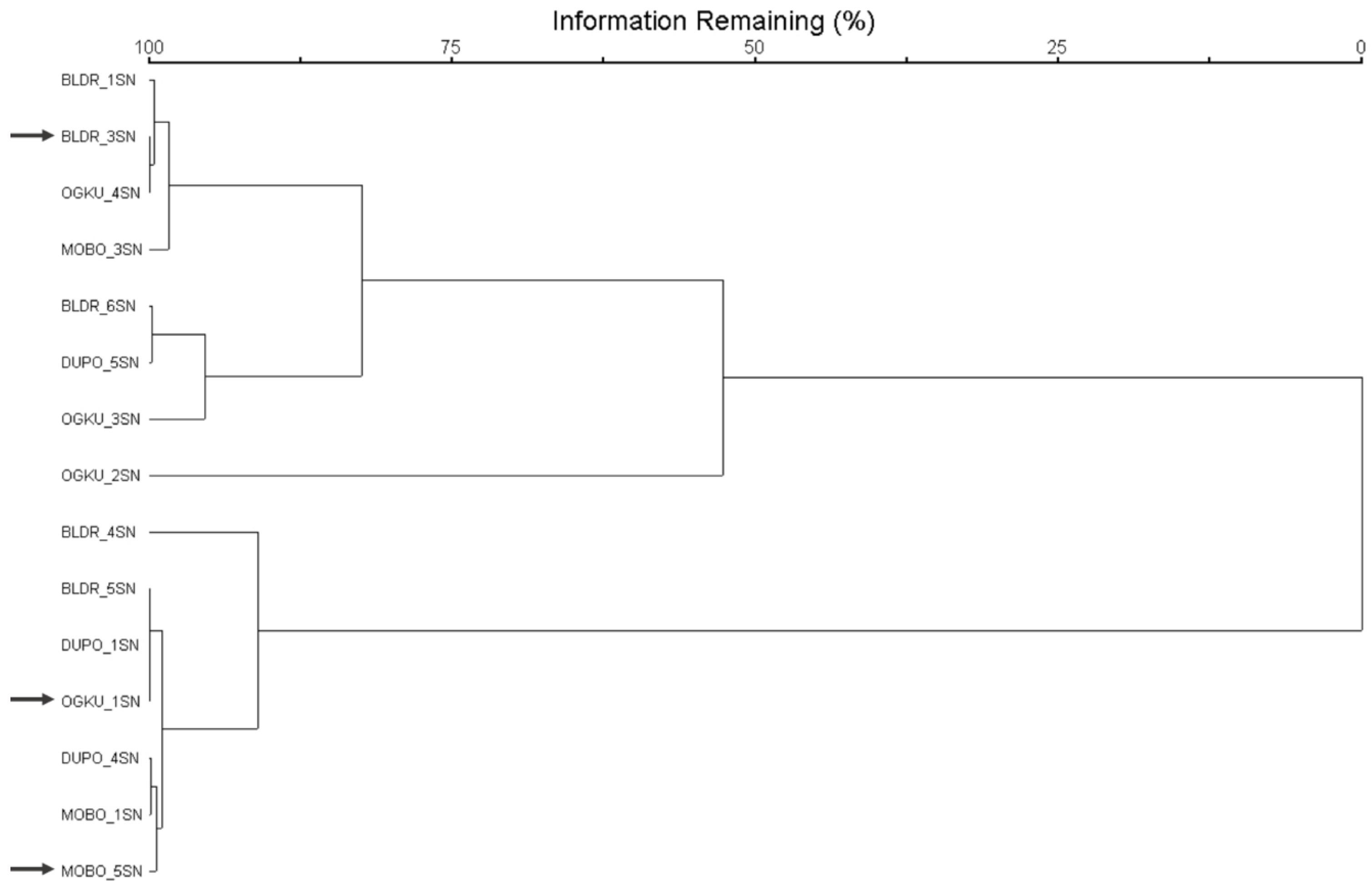


Figure 5.14. Hierarchical clustering dendrogram for 15 samples of four strains based on levels of eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, and THCA). The genetic outliers are indicated with arrows.

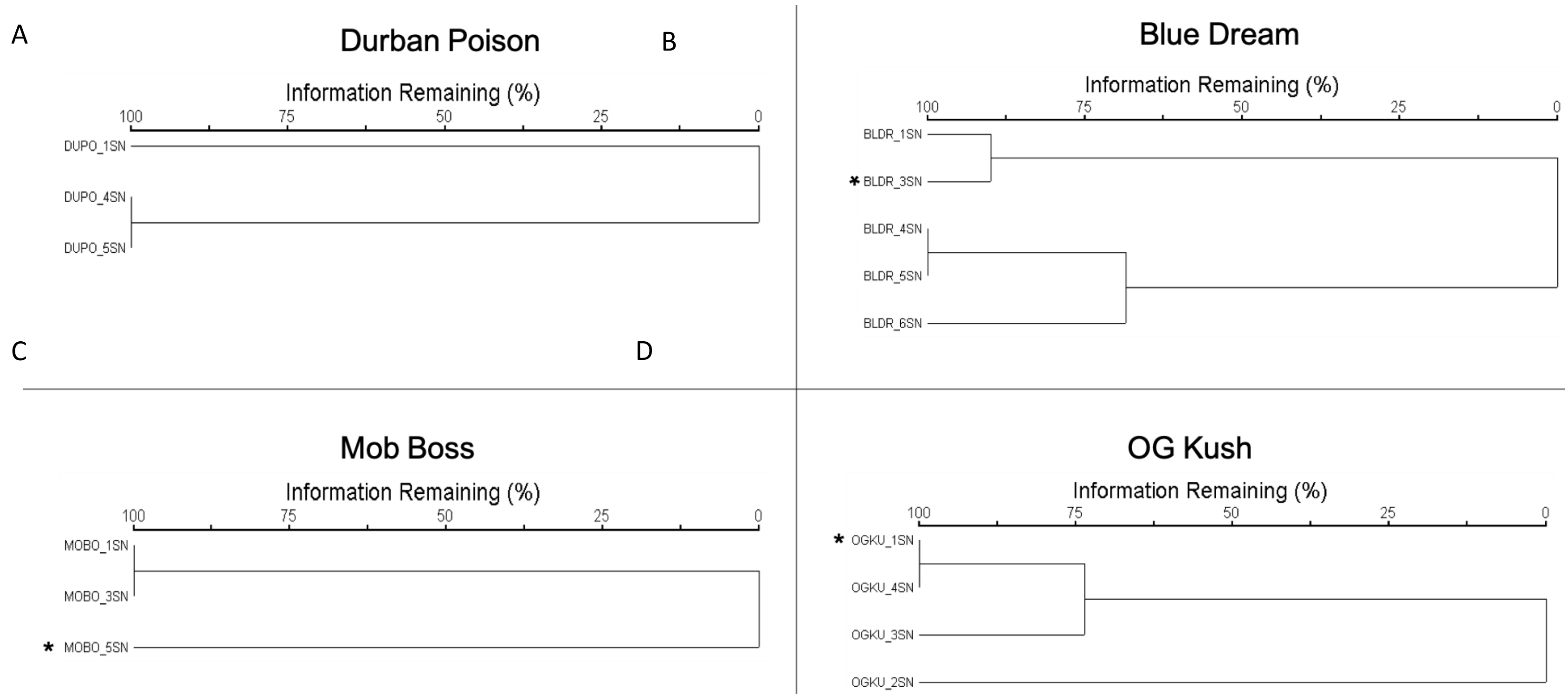


Figure 5.15. Hierarchical clustering dendrogram for 15 samples within (A) "Durban Poison", (B) "Blue Dream", (C) "Mob Boss", and (D) "OG Kush" based on levels of eight cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV, THC, and THCA).

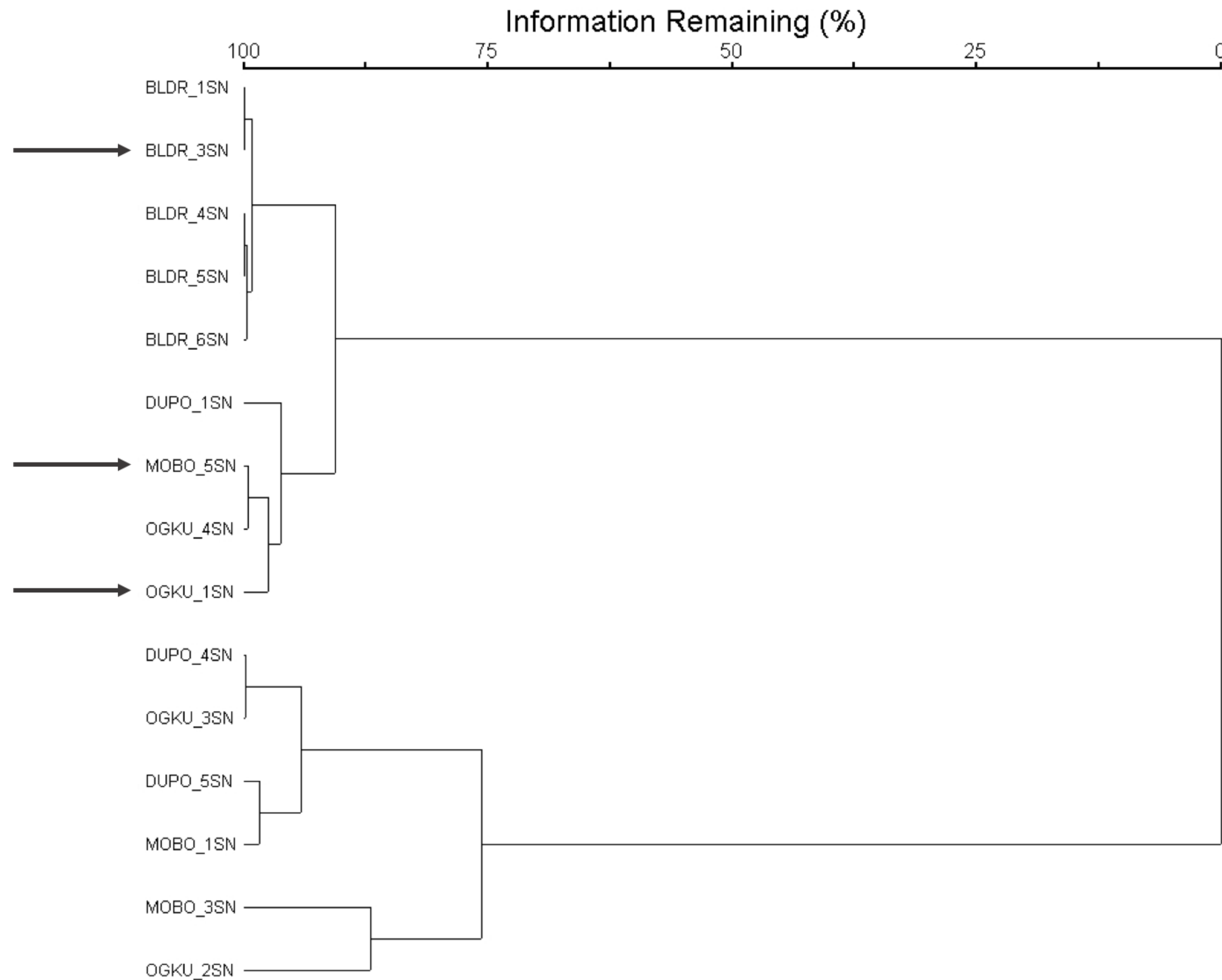


Figure 5.16. Hierarchical clustering dendrogram for 15 samples of four strains based on levels of six minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV).

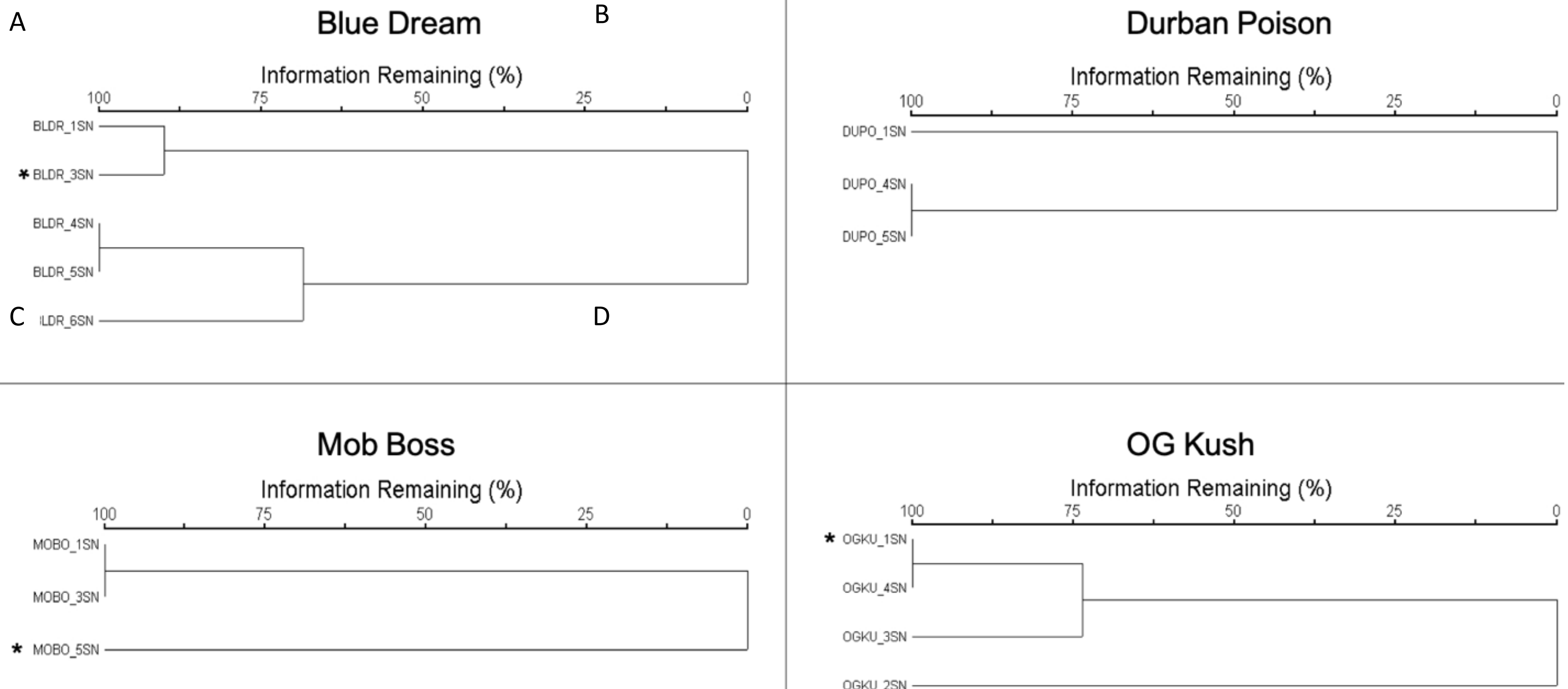


Figure 5.17. Hierarchical clustering dendrogram for 15 samples of (A) "Durban Poison", (B) "Blue Dream", (C) "Mob Boss", and (D) "OG Kush" based on levels of six minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV).

The genetic consensus “Blue Dream” samples varied from one another, and the genetic outlier did not appear to be anomalous compared to the consensus samples. When the total cannabinoid fraction (% by dry weight) with relative abundance of the chemotype by proportion was graphed, levels of THCA, THC and CBGA seem to be the highest contributors to the variation, although the levels of each are within 3% of the relative abundance of THCA, 4SN seems to be the most different (Figure 5.9). Among “Blue Dream”, most samples are similar in THCA (Table 5.4) although 4SN is noticeably lower (Figure 5.10B). Examination of the minor cannabinoids in “Blue Dream” are low, but have comparable levels of CBGA, and differing in only CBC and CBDA in 3SN and 6SN, respectively (Figure 5.11B). No other cannabinoids were detected in the “Blue Dream” samples. Hierarchical clustering analysis using data from eight cannabinoids produced two large clusters, with assignment of 1SN, 3SN* and 6SN to one cluster, while 4SN and 5SN were assigned to the other (Figure 5.14). Clustering analysis containing only “Blue Dream” samples clustered 1SN and 3SN together, while 4SN, 5SN and 6SN are assigned to the other cluster (Figure 5.15B). Hierarchical clustering analysis of the minor cannabinoids assigned all five “Blue Dream” samples to one cluster indicating none of the samples substantially differed in the less prominent cannabinoids (Figure 5.16). However, when “Blue Dream” samples were examined together, 1SN and 3SN* formed a cluster, while the remaining samples were assigned to a second cluster (Figure 5.17B).

The genetic consensus “Mob Boss” samples varied from one another, although the genetic outlier did not appear to be anomalous compared to the consensus samples and had a similar level of THCA to 1SN (Figure 5.8). When the total cannabinoid

fraction (% by dry weight) with relative abundance of the chemotype by proportion was graphed, levels of THCA and CBGA seem to be the highest contributors to the variation, although levels of THCA, THC and CBGA in the genetic outlier is markedly different (Figure 5.9). Among “Mob Boss”, the outlier shares a similar THC profile with 1SN (Figure 5.10C). Examination of the minor cannabinoids reveals the “Mob Boss” samples are vastly different in CBGA content, and differ in CBN, CBC and THCV (Figure 5.11C). Only CBG and CBGA were detected in 3SN, while CBN and CBC were detected in 5SN* and 1SN, and THCV in 1SN. Hierarchical clustering analysis using data from eight cannabinoids produced two large clusters, with assignment of the genetic outlier and one consensus sample to one cluster, while the other consensus sample was assigned to the other cluster (Figure 5.14). Clustering analysis containing only “Mob Boss” samples clustered 1SN and 3SN together, while the outlier was assigned to the other cluster (Figure 5.15C). Hierarchical clustering analysis of the minor cannabinoids assigned all genetic consensus “Mob Boss” samples to one cluster and the genetic outlier to the other, indicating the genetic outlier differed in the less prominent cannabinoids (Figure 5.16). Clustering analysis of “Mob Boss” samples supported the difference in less prominent cannabinoids in the genetic outlier (Figure 5.17C).

The genetic consensus “OG Kush” samples varied, and the genetic outlier appeared to be most similar to 4SN in cannabinoid levels (Figure 5.8). When the total cannabinoid fraction (% by dry weight) with relative abundance of the chemotype by proportion was graphed, levels of THCA, THC and CBGA seem to be the highest contributors to the variation, although levels of THCA are similar in 1SN*, 3SN and 4SN (Figure 5.9). THC in 3SN differs from the other samples, and 1SN* is the only sample

with no CBDA detected. All samples contained CBG and CBGA, although 4SN contained far less CBGA. CBC and CBN were detected at low levels in 2SN and 4SN. Among “OG Kush”, the outlier shares a similar THC profile with 4SN (Figure 5.10D). Examination of the minor cannabinoids reveals the “OG Kush” samples are vastly different in CBGA content, and CBG is higher in 2SN (Figure 5.11D). Only CBG, CBGA, and CBN were detected in 1SN*, while CBD was detected in the three consensus samples. 2SN and 4SN both contained CBN and CBC. Hierarchical clustering analysis using data from eight cannabinoids produced two large clusters, with assignment of the genetic outlier to one cluster, while the consensus sample were assigned to the other cluster (Figure 5.14). Clustering analysis containing only “OG Kush” samples placed the genetic outlier with two consensus samples and assigned 2SN to the other cluster (Figure 5.15D). Hierarchical clustering analysis of the minor cannabinoids assigned 4SN, and 3SN and 2SN were assigned to the other cluster (Figure 5.16). Clustering analysis of “OG Kush” assigned 1SN*, 4SN and 3SN to one cluster and 2SN to the other (Figure 5.17D).

Principal Components Analysis generated plots for eight cannabinoids which represented 99.9% of the variation in the data and is explained almost entirely by THCA concentrations. The PCA scaled by THCA (Axis 1; 99.002 % of variation; $r = 1.00$, $tau = 0.981$) (Figure 5.18) confirms that THCA is the overwhelming source of variation. A PCA was conducted on the minor cannabinoids (CBDA, CBG, CBGA, CBC, CBN, THCV) to determine if any minor cannabinoids are driving variation (Figure 5.19). The PCA scaled by CBGA (Axis 1, 99.2% of variation; $r = -1.00$, $tau = -1.00$) confirms that CBGA is the overwhelming source of variation.

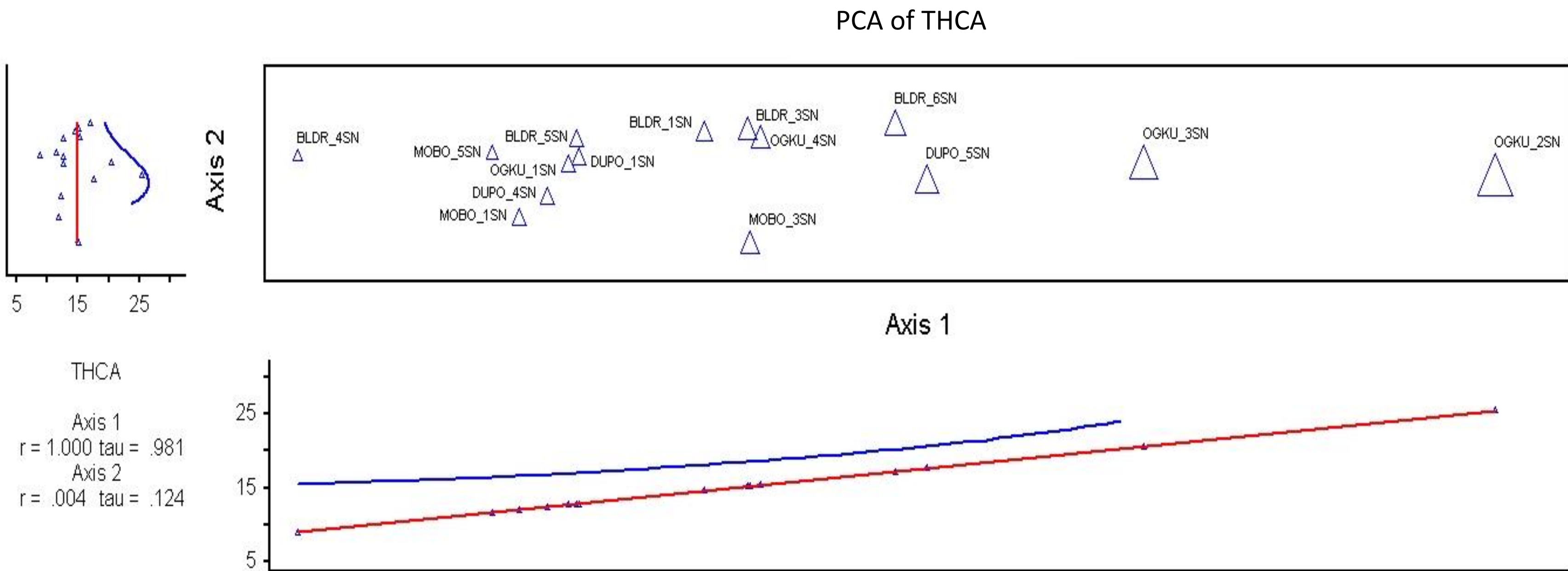


Figure 5.18. Principal Components Analysis of scaled by THCA on Axis 1 (99.002% of variation) and remaining variation on Axes 2 and 3 (0.900% and 0.079% of variation).

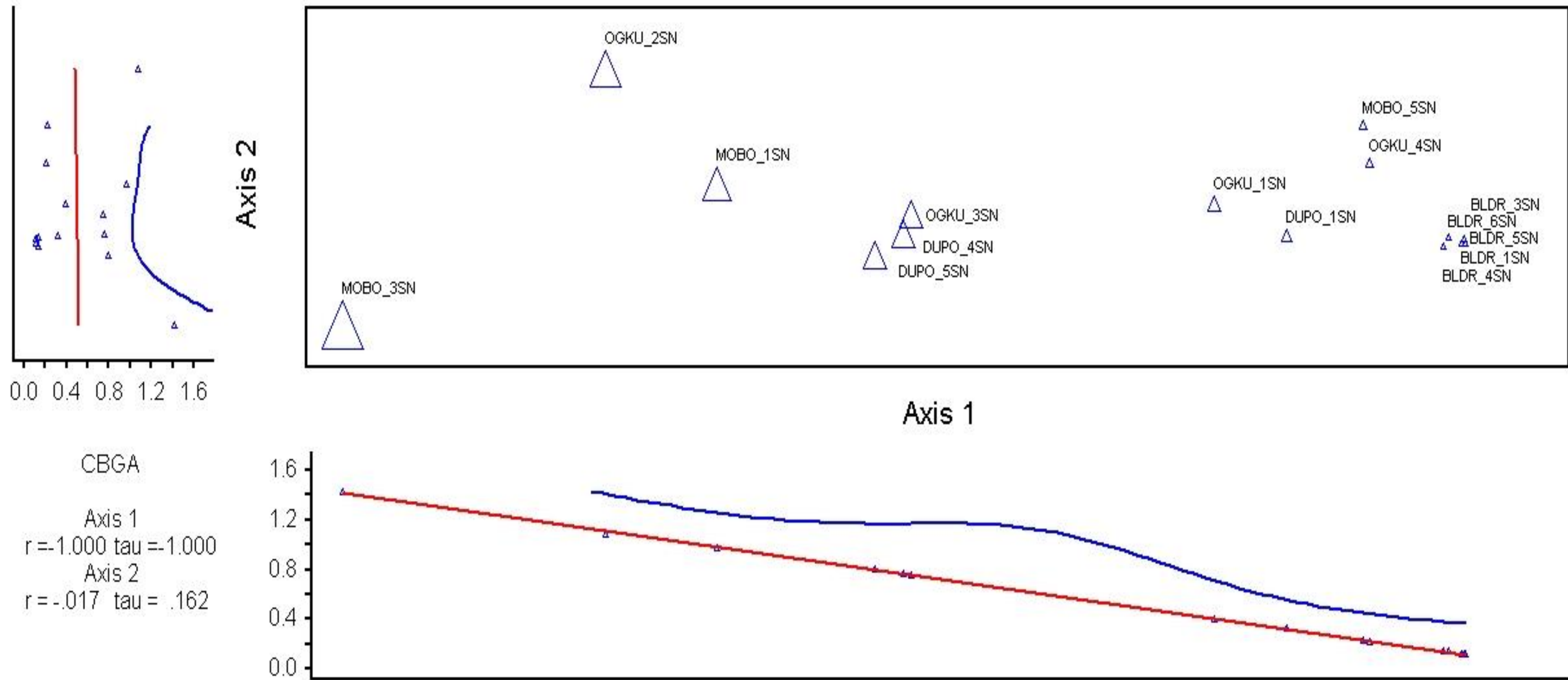


Figure 5.19. Principal Components Analysis of scaled by CBGA on Axis 1 (99.925% of variation) and the remaining variation on Axis 2 (0.075% of variation).

Given that THCA and CBGA were found to be driving variation, a linear regression analysis of THCA and CBGA was conducted to determine if there is a relationship between these two cannabinoids (Figure 5.20). There is a weak positive correlation ($R^2 = 0.1749$) between THCA and CBGA, and samples within strains do not cluster, although all 'Blue Dream' samples were low in CBGA.

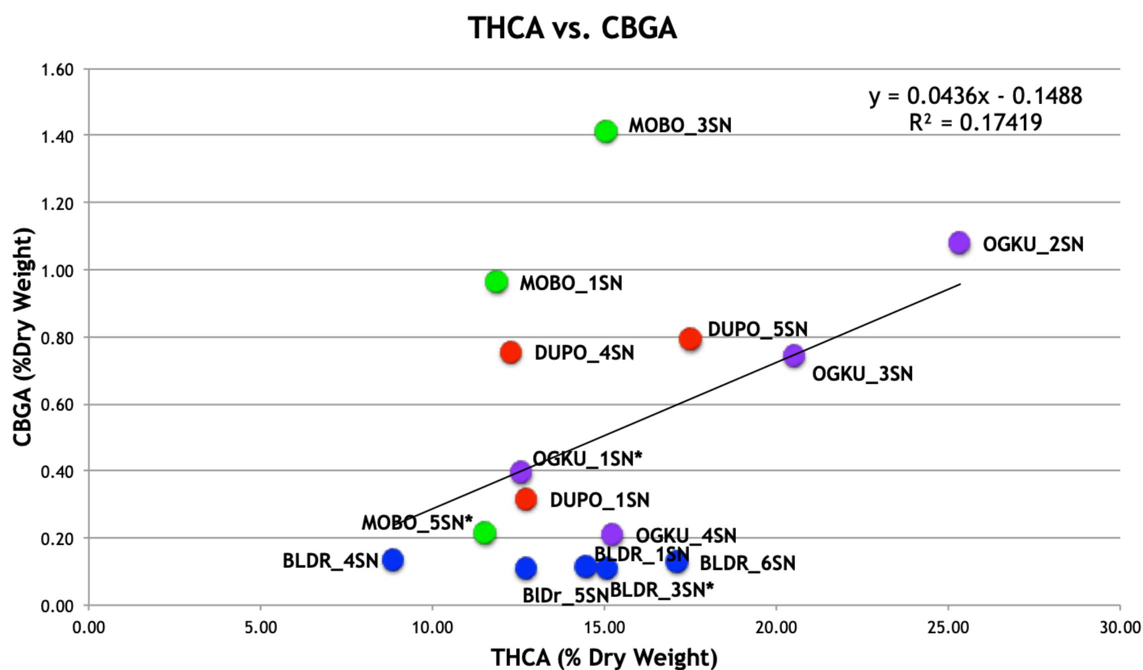


Figure 5.20. Linear regression analysis of THCA level (% dry weight) against CBGA (% dry weight). Samples are color coded and labeled with the abbreviated identification (Table 5.4).

Terpenes

A panel of 21 terpenes were analyzed in seven strains of commercially-available *Cannabis* (Table 5.9). Nineteen cannabis samples drawn from seven strains were analyzed: "Durban Poison" (3), "Blue Dream" (5), "Mob Boss" (3), "OG Kush" (4), "White Widow" (1), "White Urkle" (2), and "Tangerine Haze" (1). The analyses included samples of the same strain acquired from different dispensaries, clones from the same

dispensary (“White Urkle”), and a single-representatives of two strains (“White Widow” and “Tangerine Haze.

Table 5.9. Average terpene percent distribution by mass (mg) calculated as normalized area percentages of each of the terpenes present in the samples. The general molecular class is denoted as monoterpene (m) or sesquiterpene (s).

		α-pinene (m)	camphene (m)	myrcene (m)	β-pinene (m)	3-carene (m)	α-terpinene (m)	limonene (m)
Blue Dream	1SN	11.29 (0.25)	3.66 (0.17)	9.03 (0.20)	6.66 (0.33)	3.24 (NA)	3.46 (NA)	1.80 (2.54)
	4SN	11.87 (0.26)	4.64 (0.13)	6.93 (0.15)	6.13 (0.23)	4.60 (NA)	4.53 (0.13)	5.57 (NA)
	5SN	11.14 (0.14)	3.34 (0.39)	7.82 (0.09)	5.68 (0.26)	3.14 (0.25)	2.96 (NA)	3.19 (NA)
	6SN	12.65 (0.50)	3.58 (0.08)	8.65 (0.58)	6.36 (0.22)	3.51 (0.01)	3.47 (0.06)	3.29 (NA)
	Consensus	11.71 (0.69)	3.81 (0.58)	8.11 (0.93)	6.21 (0.41)	3.67 (0.64)	3.61 (0.66)	3.46 (1.56)
	3SN*	7.67 (0.22)	3.02 (0.11)	6.26 (0.20)	4.54 (0.04)	2.98 (0.11)	2.81 (0.11)	4.54 (1.09)
Mob Boss	1SN	7.10 (0.34)	3.06 (0.46)	3.64 (0.44)	3.50 (0.43)	2.87 (0.32)	2.89 (0.31)	2.01 (2.84)
	3SN	12.86 (0.61)	2.81 (0.37)	3.72 (0.39)	3.33 (0.47)	2.84 (NA)	2.14 (NA)	2.90 (NA)
	Consensus	9.98 (4.08)	2.93 (0.17)	3.68 (0.06)	3.42 (0.12)	2.85 (0.02)	4.09 (0.53)	2.46 (0.63)
	5SN*	7.95 (0.18)	4.50 (0.45)	7.15 (0.27)	5.26 (0.77)	0.00	4.09 (0.33)	7.42 (0.30)
OG Kush	2SN	2.13 (0.03)	1.26 (0.02)	3.21 (0.07)	2.04 (0.02)	0.93 (0.05)	0.49 (0.69)	12.79 (0.15)
	3SN	2.19 (0.28)	1.53 (0.24)	7.57 (0.34)	4.30 (0.01)	0.53 (0.75)	0.57 (0.80)	3.34 (1.34)
	4SN	3.40 (0.15)	2.63 (NA)	9.89 (0.52)	6.35 (0.07)	2.63 (0.19)	2.56 (0.19)	0.00
	Consensus	2.57 (0.72)	1.81 (0.72)	6.98 (3.39)	4.23 (2.15)	1.36 (1.11)	1.21 (1.18)	5.44 (6.74)
	1SN*	3.05 (0.19)	2.70 (0.31)	8.32 (0.50)	5.63 (0.02)	2.89 (NA)	2.42 (NA)	2.74 (0.53)
Durban Poison	1SN	4.64 (0.31)	3.26 (0.09)	3.94 (0.22)	3.67 (0.10)	2.96 (0.21)	2.84 (NA)	3.67 (NA)
	4SN	5.28 (0.18)	3.47 (0.11)	3.43 (0.00)	4.68 (0.15)	4.01 (0.12)	3.81 (0.15)	0.00
	5SN	4.61 (0.26)	2.62 (0.12)	4.78 (0.21)	3.71 (0.25)	3.52 (0.06)	3.34 (0.09)	3.46 (158)
	Consensus	4.85 (0.38)	3.12 (0.44)	4.05 (0.68)	4.02 (0.57)	3.50 (0.52)	3.33 (0.49)	2.38 (2.06)
White Widow	WW1	5.02 (0.33)	4.33 (0.39)	9.32 (0.81)	7.68 (0.98)	0.00	4.18 (NA)	0.00
White Urkle	WE	11.42 (1.89)	1.89 (0.44)	5.78 (2.22)	3.20 (0.45)	1.68 (0.32)	1.75 (0.40)	1.15 (1.62)
Tang. Haze	TANG1	3.51	2.14	2.6	2.41	2.11	2.09	0.00

Table 5.9 *continued*

		p-cymene (m)	ocimene (m)	g-terpinene (m)	terpinolene (m)	linalool (m)	isopulegol (m)	geraniol (m)
Blue Dream	1SN	0.00	3.46 (0.28)	2.94 (NA)	3.26 (0.19)	6.50 (0.10)	2.51 (0.03)	2.57 (0.16)
	4SN	0.00	4.93 (0.02)	3.92 (NA)	0.00	4.86 (0.19)	3.23 (NA)	3.60 (0.21)
	5SN	0.00	3.05 (0.28)	3.04 (NA)	4.18 (1.33)	3.45 (2.04)	2.20 (NA)	2.75 (NA)
	6SN	0.00	3.47 (0.03)	3.02 (NA)	3.30 (0.06)	6.94 (0.12)	2.46 (0.15)	2.54 (NA)
	Consensus	0.00	3.59 (0.57)	3.23 (0.46)	2.68 (1.84)	5.44 (1.60)	2.60 (0.44)	2.87 (0.50)
	3SN*	0.00	3.47 (0.02)	2.58 (0.17)	4.91 (0.06)	5.02 (0.10)	1.97 (0.01)	2.25 (0.00)
Mob Boss	1SN	0.00	3.09 (0.40)	1.15 (1.62)	2.74 (0.21)	4.11 (0.26)	2.74 (0.20)	2.52 (0.25)
	3SN	0.00	3.02 (0.34)	2.18 (0.47)	2.64 (0.36)	3.71 (0.30)	1.88 (0.05)	1.90 (0.35)
	Consensus	0.00	3.06 (0.05)	1.67 (0.73)	2.69 (0.07)	3.91 (0.28)	2.17 (0.42)	2.21 (0.44)
	5SN*	0.00	3.82 (0.30)	3.54 (0.20)	3.72 (0.06)	5.45 (0.29)	2.92 (0.17)	3.45 (0.67)
OG Kush	2SN	0.00	0.94 (0.10)	0.80 (0.04)	1.16 (0.09)	9.73 (0.34)	1.13 (0.02)	0.91 (0.01)
	3SN	0.00	1.20 (0.09)	0.57 (0.81)	1.38 (0.14)	10.18 (0.45)	1.15 (0.07)	1.13 (0.17)
	4SN	0.00	2.81 (0.11)	2.33 (NA)	2.69 (0.20)	6.57 (0.15)	1.84 (0.19)	2.13 (0.03)
	Consensus	0.00	1.65 (1.01)	1.23 (0.96)	1.75 (0.83)	8.83 (1.97)	1.37 (0.41)	1.39 (0.65)
	1SN*	0.00	3.05 (0.11)	2.48 (NA)	2.78 (0.19)	6.19 (0.04)	1.79 (0.22)	2.09 (0.06)
Durban Poison	1SN	0.00	3.11 (0.34)	2.59 (NA)	2.81 (0.11)	6.08 (0.09)	2.33 (0.34)	2.42 (0.13)
	4SN	4.39 (0.18)	6.95 (0.19)	3.66 (0.21)	12.37 (0.19)	3.95 (0.06)	2.38 (0.06)	2.94 (0.37)
	5SN	1.53 (2.17)	8.08 (0.16)	3.04 (0.03)	16.89 (0.97)	2.65 (0.15)	1.92 (0.01)	2.49 (0.02)
	Consensus	1.97 (2.23)	6.05 (2.60)	3.09 (0.54)	10.69 (7.19)	4.23 (1.73)	2.21 (0.25)	2.62 (0.28)
White Widow	WW1	6.12 (0.83)	3.59 (0.33)	3.47 (0.41)	1.60 (2.26)	4.48 (0.52)	2.77 (0.24)	2.94 (0.38)
White Urkle	WE	1.24 (1.75)	3.23 (1.02)	1.44 (0.24)	0.69 (0.97)	1.51 (0.34)	1.19 (0.22)	1.30 (0.34)
Tang. Haze	TANG1	3.26	2.28	1.81	1.89	4.83	1.81	1.55

Table 5.9 *continued*

		β-caryophyllene (s)	α-caryophyllene (s)	nerolidol-1 (s)	α-gurjunene (s)	nerolidol-2 (s)	guaiol (s)	α-bisabolol (s)
Blue Dream	1SN	16.03 (0.99)	7.08 (0.33)	4.21 (0.30)	4.82 (0.46)	4.87 (0.36)	2.50 (0.09)	4.85 (0.17)
	4SN	16.67 (2.07)	7.62 (0.33)	4.10 (NA)	5.18 (0.06)	4.60 (0.32)	3.10 (0.01)	4.45 (0.23)
	5SN	19.07 (NA)	6.67 (2.32)	5.84 (NA)	6.39 (1.62)	4.20 (2.78)	4.57 (2.98)	5.01 (NA)
	6SN	14.98 (0.75)	6.80 (0.38)	2.50 (1.63)	4.74 (0.02)	5.02 (0.19)	2.68 (0.14)	4.48 (0.68)
	Consensus	16.69 (1.73)	7.04 (0.42)	4.16 (1.36)	5.28 (0.76)	4.67 (0.36)	3.21 (0.94)	4.70 (0.28)
	3SN*	14.00 (0.18)	5.91 (0.00)	3.46 (0.03)	3.90 (0.07)	5.75 (0.07)	8.73 (0.11)	7.47 (0.01)
Mob Boss	1SN	15.01 (0.04)	7.12 (0.08	6.12 (0.03)	6.28 (0.33)	13.35 (0.18)	2.30 (0.23)	8.86 (0.37)
	3SN	18.90 (0.05)	8.46 (0.29)	3.94 (3.39)	6.15 (0.32)	12.71 (0.08)	1.83 (0.40)	6.02 (0.25)
	Consensus	16.95 (2.75	7.79 (0.95)	5.03 (1.54)	6.21 (0.10)	13.03 (0.45)	2.07 (0.33)	7.35 (1.89)
	5SN*	10.01 (0.97)	6.18 (0.08)	1.31 (0.09)	6.00 (0.09)	9.87 (0.68)	3.04 (0.37)	4.33 (0.05)
OG Kush	2SN	15.81 (0.03)	7.83 (0.04)	2.43 (0.12)	10.54 (0.10)	24.10 (0.10)	0.67 (0.07)	0.92 (0.37)
	3SN	27.77 (0.09)	10.04 (0.14)	2.86 (2.55)	5.40 (0.09)	10.80 (0.22)	7.49 (0.12)	0.00
	4SN	22.63 (0.27)	7.57 (0.09)	0.95 (0.08)	5.04 (0.15)	7.94 (0.16)	6.88 (0.09)	5.54 (0.16)
	Consensus	22.07 (6.00)	8.51 (1.32)	2.08 (1.00)	6.99 (3.08)	14.28 (8.62)	5.01 (3.77)	2.15 (2.97)
	1SN*	23.13 (1.42)	8.05 (0.39)	1.00 (0.01)	4.95 (0.09)	7.88 (0.65)	7.22 (0.56)	5.54 (0.18)
Durban Poison	1SN	13.13 (0.32)	5.63 (0.05)	4.60 (NA)	5.18 (0.03)	10.77 (0.54)	10.44 (0.21)	12.19 (0.38)
	4SN	11.53 (0.54)	4.39 (1.43)	5.11 (2.63)	4.73 (0.07)	7.49 (0.62)	5.42 (0.09)	0.00
	5SN	9.21 (1.18)	4.26 (0.41)	8.36 (1.56)	3.63 (0.34)	5.84 (1.02)	4.97 (0.47)	1.08 (1.53)
	Consensus	11.29 (1.97)	4.76 (0.75)	6.02 (2.04)	4.51 (0.80)	8.04 (2.51)	6.94 (3.03)	4.43 (6.75)
White Widow	WW1	16.84 (NA)	6.64 (2.68)	10.01 (1.66)	4.80 (0.57)	7.03 (1.09)	6.61 (1.12)	0.00
White Urkle	WE	34.12 (0.98)	699 (6.10)	6.29 (0.96)	4.20 (4.20)	4.72 (5.04)	1.26 (0.37)	4.95 (4.07)
Tang. Haze	TANG1	28.31	2.49	5.14	4.35	9.21	6.78	11.43

Histograms of 21 terpenes were generated to compare terpene composition among samples within strains (Figure 5.21). The level of terpenes varied within strains as well as across the four strains. “Durban Poison” samples were genetically identical, and yet the terpene levels varied among the samples (Figure 5.21A). Although many terpenes were similar among the samples, limonene, p-cymene, ocimene, terpinolene, nerolidol-1, nerolidol-2, guialol and α-bisabolol had at least one sample that was markedly different. In the PCA of terpenes “Durban Poison” showed a wide distribution across Coordinate 1 (Figure 5.23).

Four “Blue Dream” samples were genetically identical (1SN, 4SN, 5SN and 6SN) and 3SN* was a genetic outlier. Although many of the terpenes had similar levels, α-pinene, terpinolene, nerolidol-1, guialol, and α-bisabolol had at least one sample that was markedly different (Figure 5.21B). Seventeen of 21 terpene levels in the genetic outlier fell outside the standard deviation of the genetic consensus (Figure 5.22A). Of the five highest levels of terpenes pooled within each strain, eight of 12 terpenes in the genetic outlier fell outside the standard deviation of the genetic consensus average (Figure 5.23A). In the PCA of terpenes “Blue Dream” samples cluster, indicating shared terpene characteristics (Figure 5.24).

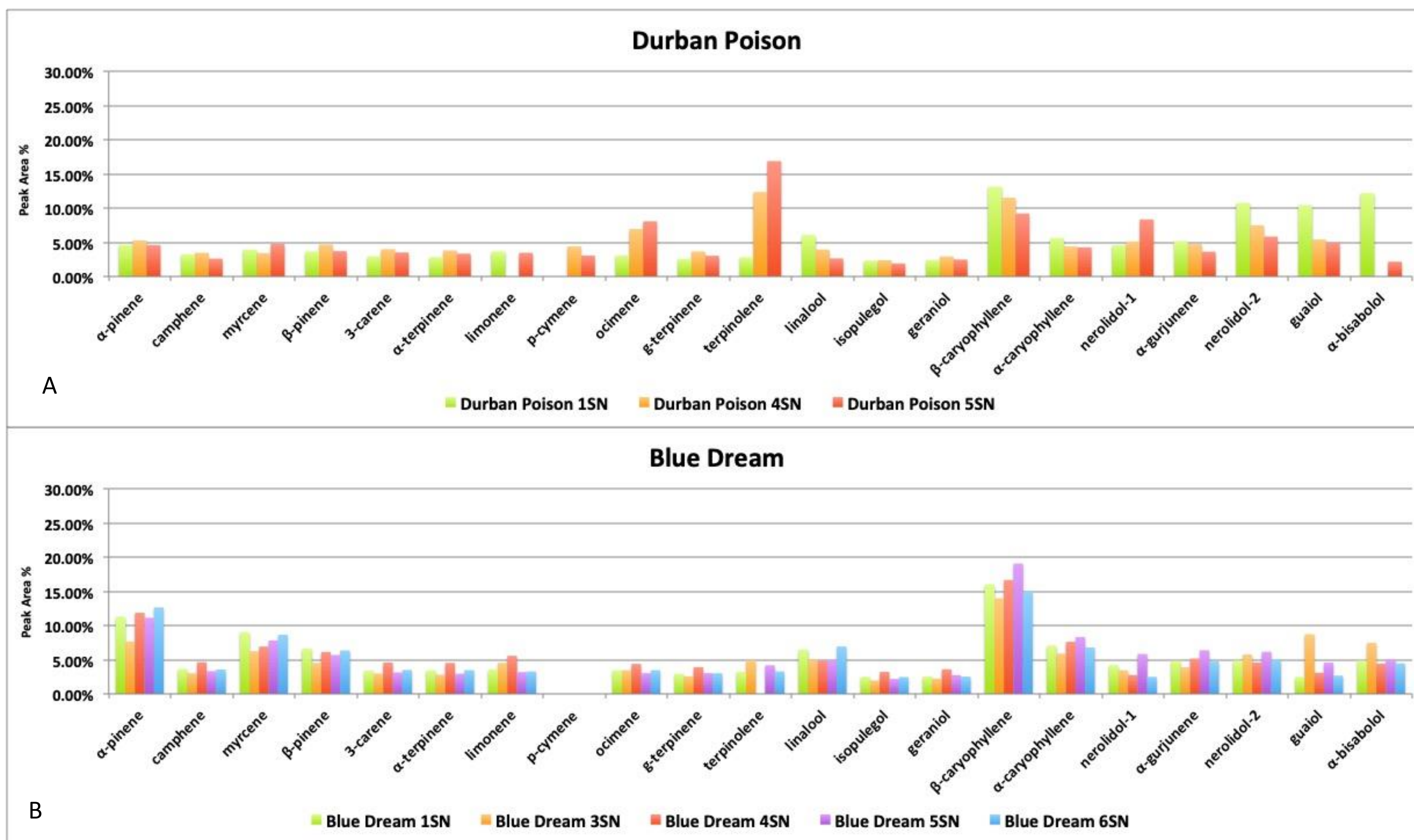


Figure 5.21. cont. next page

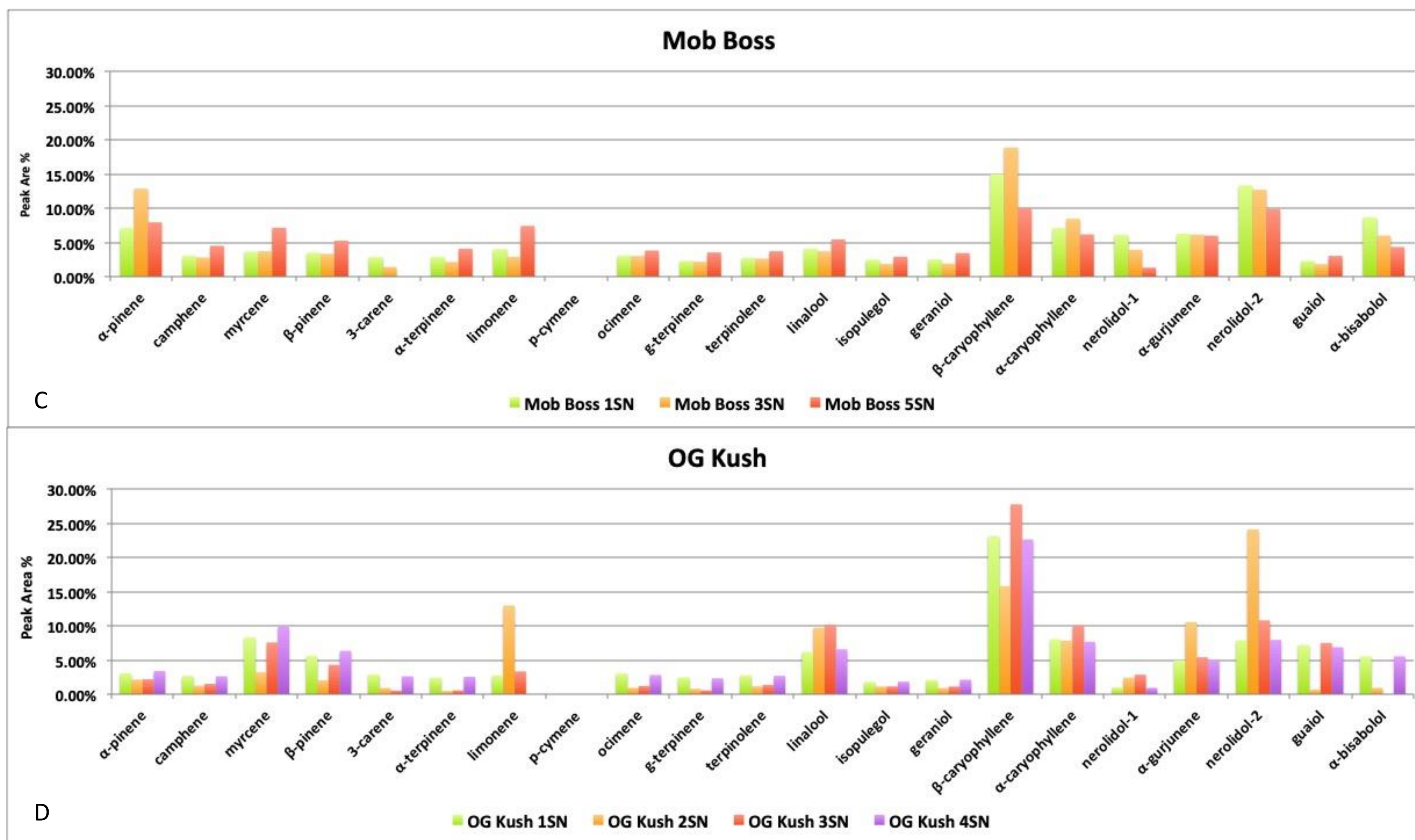


Figure 5.21. Terpene variation within each accession of four strains (% distribution) for (A) "Durban Poison", (B) "Blue Dream", (C) "Mob Boss", and (D) "OG Kush". The genetic outlier is indicated with an asterisk.

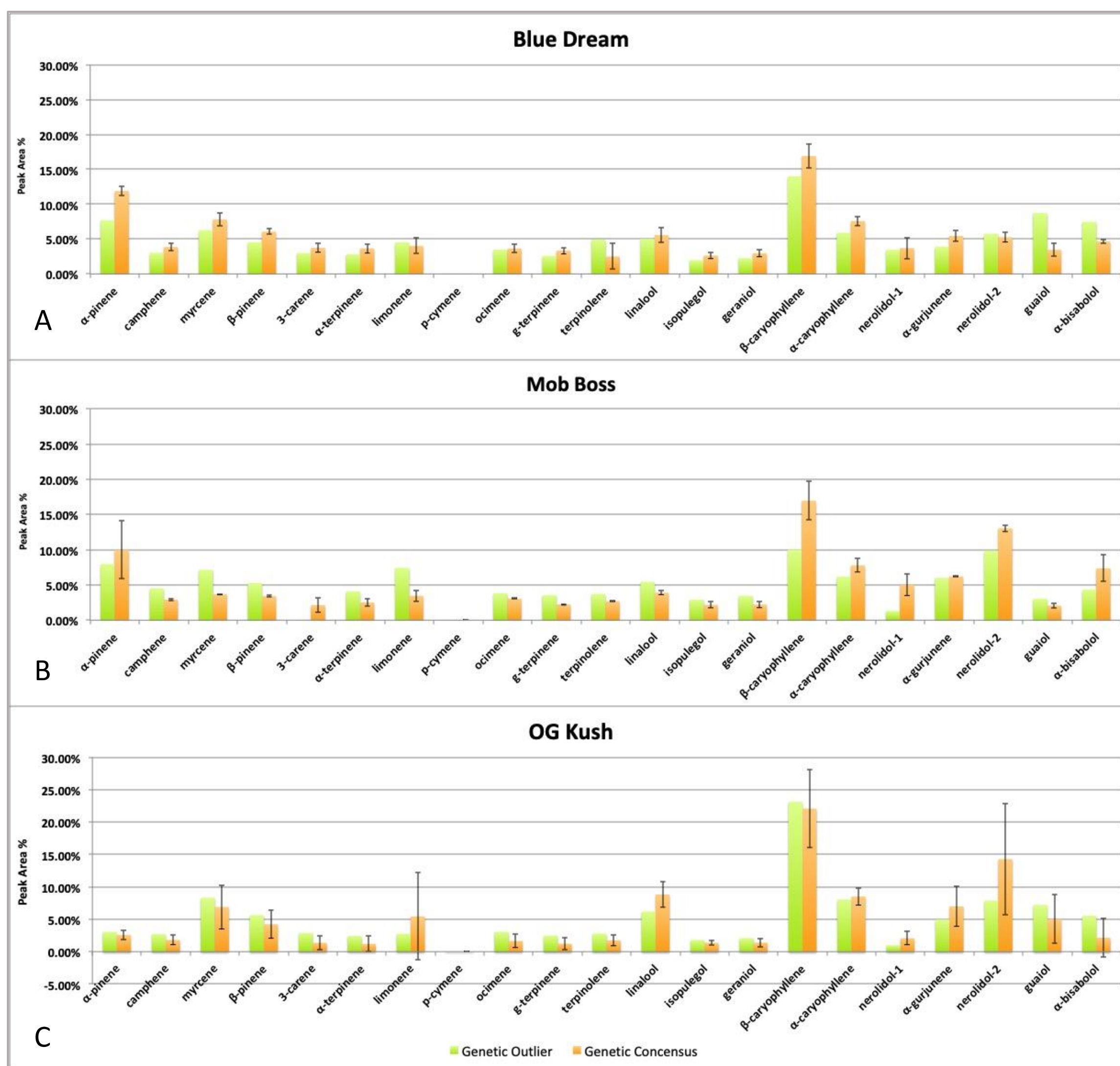


Figure 5.22. Twenty-one terpenes (A) "Blue Dream", (B) "Mob Boss", and (C) "OG Kush", with the average for the genetic consensus samples and the genetic outlier. Error bars represent the standard deviation.

Two “Mob Boss” samples were genetically identical (1SN and 3SN) and 5SN* was a genetic outlier. Although terpenes such as ocimene, terpinolene, isopulegol, and α-gurjunene had similar levels, almost all terpenes had at least one sample that was markedly different, and in many cases the sample with the different level was the genetic outlier 5SN* (Figure 5.21C). Nineteen of 21 terpene levels in the genetic outlier fell outside the standard deviation of the genetic consensus (Figure 5.22B). Of the five highest levels of terpenes pooled within each strain, 10 of 12 terpene levels in the genetic outlier fell outside the standard deviation of the genetic consensus (Figure 5.23B). In the PCA of terpenes “Mob Boss” samples cluster, indicating shared terpene characteristics (Figure 5.24).

Three “OG Kush” samples were genetically similar (2SN, 3SN and 4SN) and 1SN* was a genetic outlier. Although terpenes such as isopulegol, and geraniol had similar levels, almost all terpenes had at least one sample that was markedly different, although there does not seem to be a sample that is consistently different across terpenes (Figure 5.21D). However, samples 1SN* and 4SN have highly similar levels in almost every terpene. This was supported in the average terpene level of the genetic consensus samples compared to the genetic outlier, where only thirteen of 21 terpenes fell outside the standard deviation of the genetic consensus (Figure 5.22C). Of the five highest levels of terpenes pooled within each strain, and only one of 10 terpene levels in the genetic outlier fell outside the standard deviation of the genetic consensus average (Figure 5.23C). The PCA of “OG Kush” samples do not form a tight cluster, indicating minimal shared terpene characteristics (Figure 5.24).

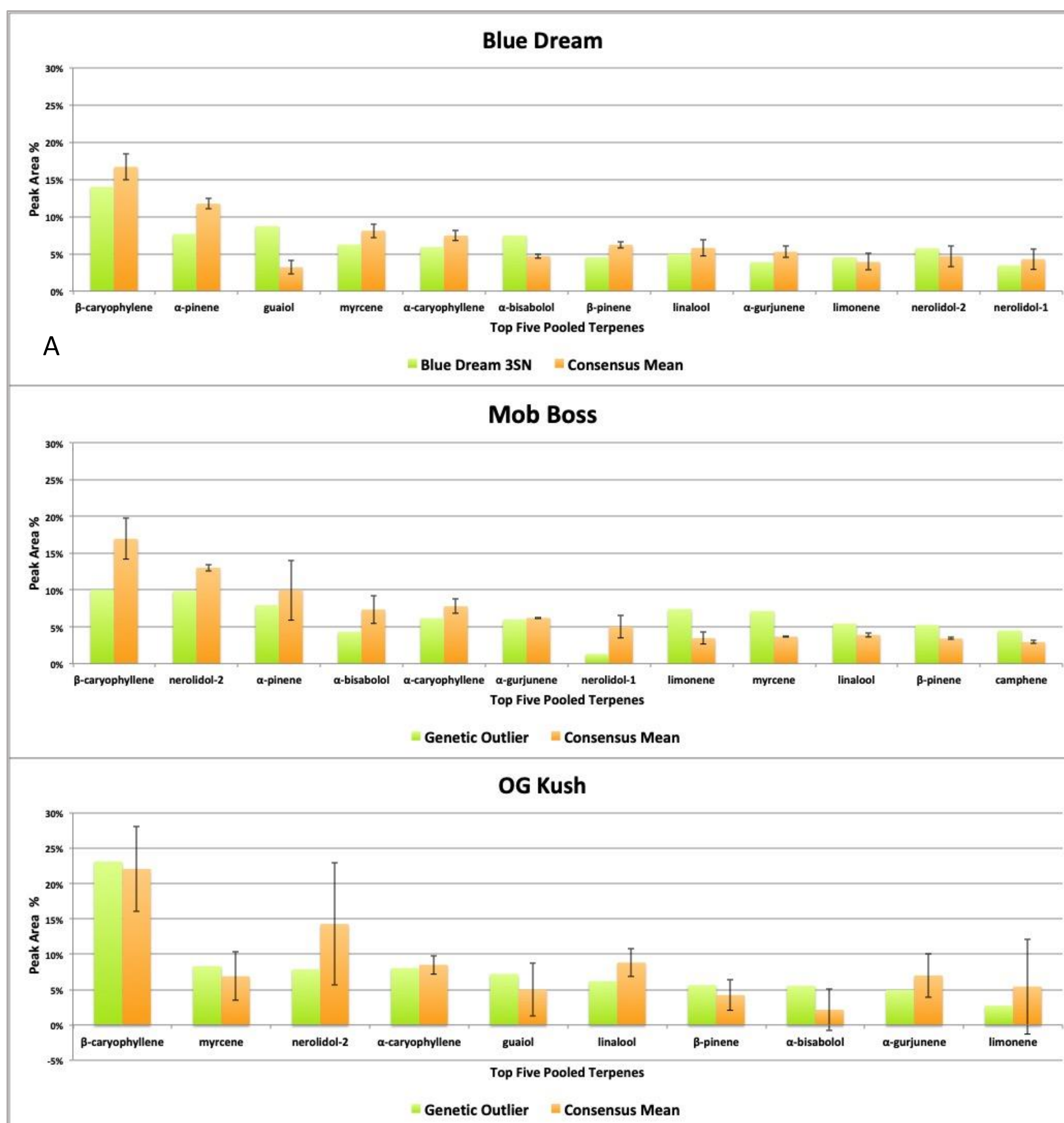


Figure 5.23. Top five terpenes with the highest levels pooled for (A) "Blue Dream", (B) "Mob Boss", and (C) "OG Kush", with the average for the genetic consensus samples and the genetic outlier. Error bars represent the standard deviation.

Histograms of the average terpene levels for seven strains was generated to examine terpene variation of different strains to explore the possibility that strains may have unique terpene composition (Figure 5.24). From this analysis, it is clear that strains have varying terpene profiles. For example, "White Widow" had no detectable levels of

α-bisabolol and "Durban Poison" has a much higher level of terpinolene than the other strains.

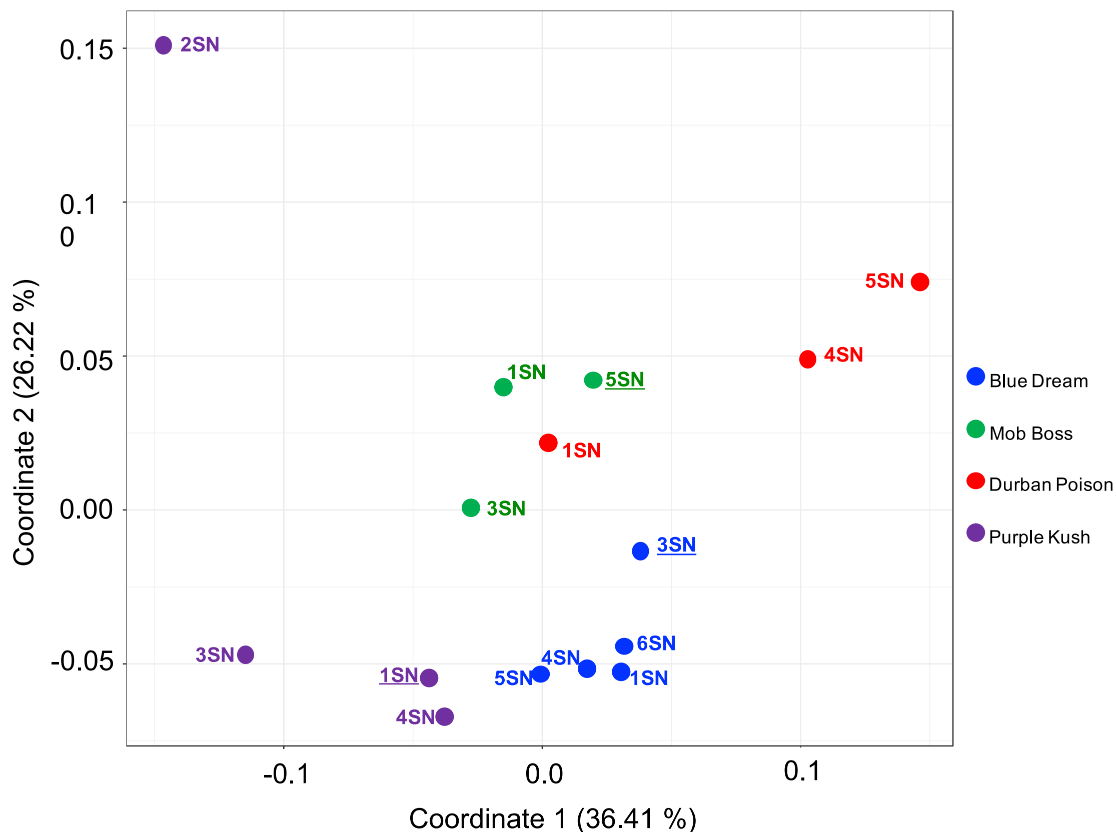


Figure 5.24. PCA clustering analysis. Genetic outlier labels are underlined.

In order to determine if specific terpenes contribute to variation among strains, standard deviations across all samples were calculated and terpenes with the highest deviation were graphed by strain (Figure 5.25). "OG Kush" has relatively high levels of β-caryophyllene and neridiol-2 compared to α-pinene, α-bisabolol and terpinolene.

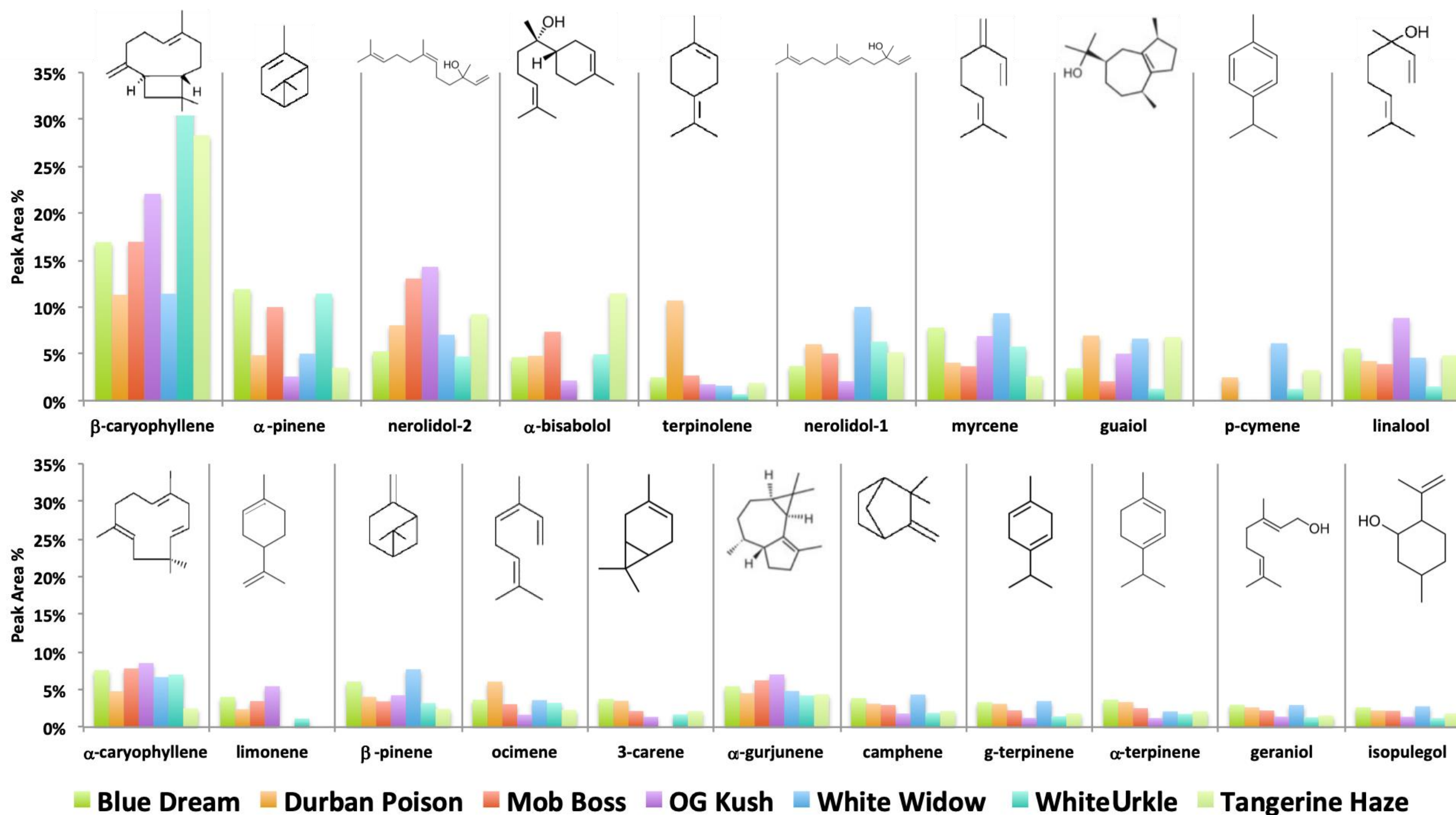


Figure 5.25. Average terpene levels in seven strains (% distribution by terpene). "Blue Dream", "Mob Boss" and "OG Kush" averages only include the genetic consensus samples. The molecular structure for each terpene is displayed above the graph.

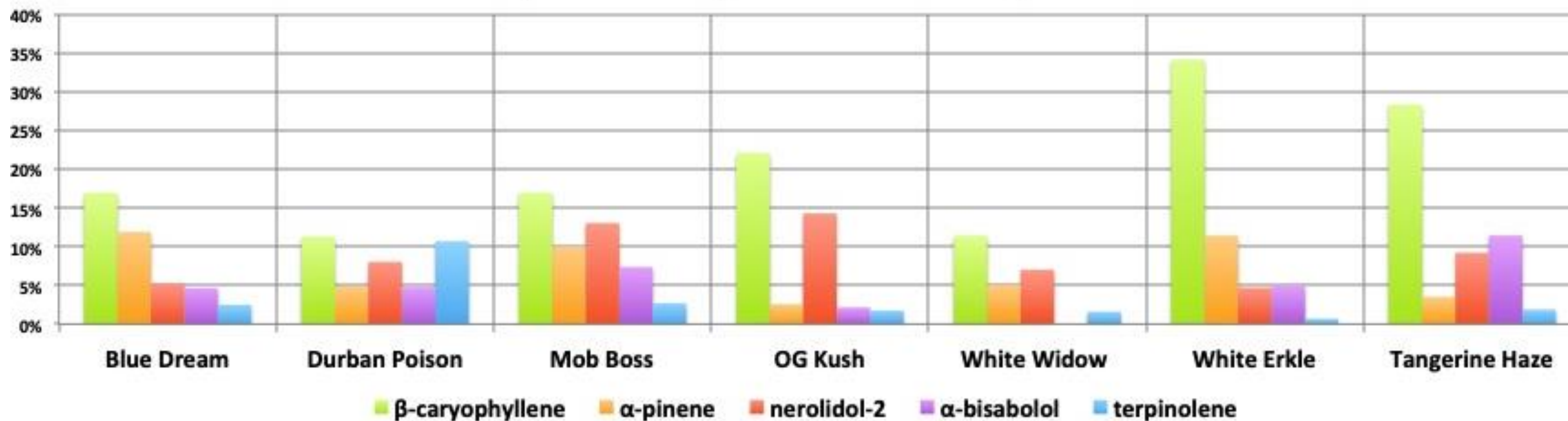


Figure 5.26 Terpene profiles for seven strains. These five terpenes had the five highest average % distributions across samples and are arranged by highest average terpene level. "Blue Dream", "Mob Boss" and "OG Kush" averages only include the genetic consensus samples.

Discussion

Cannabis diversity is reflected in a wide range of phenotypes. Over the last several decades of prohibition, underground breeders have created new strains with variable and unique characters that are desirable to consumers. Now that *Cannabis* is a growing legal market in the majority of the United States, variation is problematic as medical and recreational consumers and industry look for consistency in products. Consumers accept slight variation in plants of the same variety, but there is also an expected level of consistency when looking for a particular product. A 'Granny Smith' apple is not an acceptable substitute for a 'Honey Crisp' apple, while there may be some variability within either type sourced from different producers or regions. Variation in any plant variety (including *Cannabis*) can come from several sources, such as growth conditions (soil, light, temperature, water, air flow, etc.), harvest time (early, mid or late), storage conditions, and shelf life (time until deterioration), but no matter the environmental variation affecting the quality of the harvest, the genetic integrity is maintained. Moreover, cannabinoid concentrations can differ significantly not only at different times during flower maturation, but also in flowers from different locations (high or low) on the same plant (Aizpurua-Olaizola et al. 2016; Richins et al. 2018). Previous work has uncovered genetic variation within *Cannabis* strains (Chapter II), which we consider in some cases to be an initial source of variation; if there is genetic variation within a strain, the potential to grow a consistent product among facilities is more difficult.

Public interest in *Cannabis* is increasing, and breeders continue to produce strains with unique chemical compositions. Interest in medical applications of *Cannabis* is

expanding, and clinicians are seeking information on the benefits of particular phytochemical constituents, or combinations thereof, that contribute to alleviating certain symptoms of a variety of medical conditions. Evidence suggests *Cannabis* phytochemicals work synergistically (Ben-Shabat et al. 1998), and the unique combination of cannabinoids and terpenes in different strains is a centerpiece of interest in the industry. Recently, the potential of using a chemical fingerprint to identify specific strains and provide consistent products to consumers has been suggested (Hazekamp and Fischedick 2012; Hillig and Mahlberg 2004; Jikomes and Zoorob 2018).

The purpose of this study was to determine if genetic differences are mirrored as differences in chemical profiles within the same strain. Four strains, three of which included samples that were genotyped as identical (“Durban Poison”, “Blue Dream” and “Mob Boss”), and one which had samples differing at only one locus (“OG Kush”), were analyzed. Three of the strains included a previously determined genetic outlier (“Blue Dream”, “Mob Boss” and “OG Kush” (Chapter IV). Cannabinoid and terpene levels of genetic consensus samples were compared to levels in the genetic outlier to determine if phytochemicals vary (1) among samples that are genetically identical (2) within strains, and (3) among different strains. *Cannabis* research has shown some strains have limited variability and may be distinguishable by chemotype (Hazekamp and Fischedick 2012; Jikomes and Zoorob 2018). However, data are limited describing chemical composition and quality within strains resulting from variation in the genotype and/or varying environmental conditions (Richins et al. 2018).

Cannabinoids produced in the trichomes of the female flower are phytochemicals that elicit physiological effects and are known to vary in potency due to environmental

and genetic differences. Our cannabinoid analyses demonstrate that not only is there variation in levels of phytochemicals, but also in the relative ratios of cannabinoids (THC, CBD, and CBG) among plants with identical genotypes (Figure 5.9). THCA/THC varied among samples within the four strains (Figure 5.14), and the levels in the genetic outlier fell within one standard deviation for “Blue Dream” and “Mob Boss”, but not for “OG Kush”. Given that THCA is one of the main phenotypic traits breeders are selecting, it is possible that the minor cannabinoids paint a more accurate picture of how genetic differences are reflected in chemotype. The minor cannabinoids varied among samples within the four strains (Figure 5.13). Genetic outliers varied in cannabinoid levels in “Blue Dream” (CBC) “Mob Boss” (THCV and CBN), and “OG Kush” (CBDA). Clustering analysis scaled to THCA and CBGA indicated these two cannabinoids are driving variation in all samples (Figures 5.18 and 5.19).

Terpenes produced in the trichomes of the female flower are phytochemicals that contribute to aromas and flavors in *Cannabis* and are known to vary in potency due to environmental and genetic differences. Our terpene analyses demonstrate there is variation among plants with identical genotypes (5.21). Terpenes varied among samples within the four strains (Figure 5.21), and the levels in the genetic outlier fell within one standard deviation for four of 21 terpenes for “Blue Dream”, two of 21 terpenes for “Mob Boss”, and eight of 21 terpenes for “OG Kush” (Figure 5.18). Therefore the terpenes measured in the genetic outlier of “Blue Dream” differed by 80%, “Mob Boss” differed by 90%, “OG Kush” differed by 61%. Although variation in terpenes is expected due to differences in cultivation and storage practices, these results suggest that the genetic outlier is not only genetically different, but also chemotypically different. This analysis

gives more evidence to mislabeling and also indicate the need for further examination of genetic differences in strains and how they may be driving variability in chemotypes.

The variation assessment of cannabinoids among strains show THCA, THC, and CBGA are the largest constituents in the profile with variable levels within strains (Figure 5.10 and 5.11). In the light of recent work highlighting genetic variation within strains (Schwabe and McGlaughlin 2018), it is possible that genetic variability would result in deviations from chemotype expectations. Fischedick et al. (2010) collected samples of "White Widow" and "Amnesia" from 10 coffee shops in the Netherlands and found chemical deviants in both strains, which coincidently aligned with the chemotype of the alternate strain, possibly indicating a mix-up at the coffee shop or the supplier. However, without examining the genetic identity of the samples, it is unclear if the chemical deviation from the other samples in the set were due to genetic differences, or differences in flower maturity, or post-harvest processing and storage. As there are no standard growing or harvesting protocols, variation in abundance of phytochemicals is expected among facilities. Previous work has shown chemical composition and THC_{TOTAL} can be highly variable within strains and has suggested that strain names are an unreliable indicator of potency, and therefore strain names should be eliminated and *Cannabis* types should be based on chemotype (Hazekamp and Fischedick 2012; Elzinga et. al 2015). However, strain names are not potency specific, but rather a name given to a cultivar with a unique profile of hundreds of compounds. Although, accurate reporting of a wider scope of phytochemicals in products would be beneficial, the unique cultivars and varieties should have names as they allow consumers and growers to communicate and identify characteristics associated with a certain cultivar. Consistency in cultivars is

achievable through genotyping, stabilizing genetic lines, and standardizing protocols in the *Cannabis* industry. Without standard conditions, and a confirmed genotype, variation is not only expected but will continue to proliferate.

Chemotypes of "Blue Dream" and "OG Kush" were previously analyzed in a large study which included 35 strains with > seven samples in each (N = 494) (Elzinga et al. 2015). It was found that most replicates within strains did not cluster and showed highly variable chemotypes, but some strains, including "Blue Dream", formed clear clusters indicating a distinct chemical profile. In the current study, we also found that "Blue Dream" forms a distinct cluster, even with the genetic outlier (Figures 5.19 and 5.24). Fischedick et al. (2010) found that *Cannabis* clones of several strains could be distinguished from one another based on their cannabinoid and terpene content. *Cannabis* varieties are generally divided into three chemotype groups based on THC:CBD ratios (Hazekamp and Fischedick 2012; Hillig and Mahlberg 2004; Jikomes and Zoorob 2018). Most drug-type, and all the strains included in this study, are included in the chemotype I group, which is categorized as high THC and low CBD (Jikomes and Zoorob 2018). Given that there has been a relationship found between THC/CBD in different chemotypes, we thought perhaps there might also be a relationship between CBGA and THCA. Linear regression on the samples here indicate there is a weak relationship (R^2 = 0.174) between CBGA and THCA (Figure 5.20), but because there is so much variation, more sampling is needed to examine this further. Elzinga et al. (2015) proposed that strains with distinct chemical profiles might be more easily identified by their smell due to relatively high concentrations of specific terpenes. However, our analysis indicates "Blue Dream" does not have relatively high levels of any specific measured terpenes. The

highest terpene found in “Blue Dream” samples was β-caryophyllene, but levels were much higher in “OG Kush”, “White Urkle” and “Tangerine Haze” (Figure 5.25 & 5.26). The most abundant terpene in the seven strains we examined was β-caryophyllene, (Figure 5.25) although other terpene studies have found myrcene to be the most abundant terpene in *Cannabis* (Casano et al. 2011). It is not clear why there would be such a clear descrepancy between these two studies. Strain names were not provided in the Casano et al. (2011) study, but they categorized their samples as ‘mostly indica’ or ‘mostly sativa’, and the ‘mostly sativa’ had much lower levels of myrcene than the ‘mostly indica’ samples. All of the strains in our data set are Hybrid types with the exception of “Durban Poison”, which is a Sativa type strain.

Since this was an investigation of sources of variation within *Cannabis*, we thought it would be interesting to compare THC levels reported on the dispensary packaging to the total THC equivalent levels measured at Mile High Labs. Jikomes and Zoorob (2018) analyzed Washington state’s seed-to-sale tracabililty data set and found principle cannabinoid variation among state-certified testing laboratories. The 15 samples we used for this study had striking discrepancies in the reported levels of major cannabinoids (THCA and THC) between dispensary packaging labels and levels measured by Mile High Labs (Table 5.10). Every sample tested had lower THC_{TOTAL} content than was reported on packaging labels, with reported levels of THC_{TOTAL} by dispensaries being 23.5% – 61.48%. higher than what was measure at Mile High Labs. This discrepancy could be the result of several variables such as lab testing protocols, storage conditions, or age, but there is not enough information provided by distributors to determine the source of the discrepancy. However, this should be considered a major

issue in the industry, since the high potency strains that have been receiving attention in that last few years may not actually have the THC levels reported by the testing lab and claimed on the packaging. For example, "Mob Boss" 5SN had one of the highest reported THC levels (25.2 – 28.9 %), but tested at 10.42% at Mile High Labs, an average difference of 61.48%. The highest total THC equivalent measured at Mile High Labs was "OG Kush" 2SN, and although the packaging report was over-represented, this sample had the least discrepancy of only 23.5% more total THC equivalent than was determined by Mile High Labs.

Table 5.10. THC_{TOTAL} determined by mile high labs compared to THC reported on retail packaging. Fifteen samples from four strains with the THC_{TOTAL} (total % dry weight) measured by Mile High Labs (MHL). The THC levels reported by the dispensary, the discrepancy difference in average reported THC and the THC (total) measured by MHL, and the % of THC that was over-reported by the dispensary label vs. what MHL measured.

Strain	Sample ID	MHL THC (TOTAL)	Reported Range	Reported Average	THC Discrepancy	% THC overreporte d
Durban Poison	DuPo 1SN	11.58	NA	17.40	5.82	33.45
	DuPo 4SN	10.81	NA	20.14	9.33	46.33
	DuPo 5SN	15.63	NA	21.50	5.87	27.30
Blue Dream	BlDr 1SN	12.80	NA	17.33	4.53	26.14
	BlDr 3SN	13.37	NA	17.87	4.50	25.18
	BlDr 4SN	7.970	14.41-25.18	19.80	11.82	59.74
	BlDr 5SN	11.35	NA	16.64	5.29	31.79
	BlDr 6SN	15.26	NA	NA	NA	NA
Mob Boss	MoBo1SN	10.66	19.00-31.00	25.00*	14.34	57.36
	MoBo 3SN	13.47	22.12-24.87	23.50*	10.02	42.67
	MoBo 5SN	10.42	25.20-28.90	27.05*	16.63	61.48
OG Kush	OGKu 1SN❖	11.23	15.20-26.14	20.67*	9.44	45.67
	OGKu 2SN	22.70	28.07-31.28	29.68*	6.98	23.50
	OGKu 3SN	18.06	NA	24.10	6.04	25.06
	OGKu 4SN❖	13.60	NA	25.93	12.33	47.55

❖Asterisk indicates samples identified as the same grow facility and batch
*Average calculated based on the THC% range reported on the packaging

Although these results show variation among genetically identical samples, challenges and limitations need to be addressed. The samples were purchased three months prior to chemotype testing and had been moved in and out of -2 °C storage during the sensory study (Chapter IV). During this process, care was taken to limit the degradation of phytochemicals by storing at low temperatures (-2 °C) in amber glass jars with screw-top lids. Moreover, CBN, the degradation product of THCA, was not present in levels indicative of incorrect storage. It is possible, however, that phytochemical levels were affected. There may also be differences in lab testing protocols. Although there are standard protocols for testing in Colorado, labs are not restricted to either HPLC or GC, and it is possible these two tests may produce different results. Dispensaries and consumers often choose products with higher THC content, and for this reason, some labs may have protocols maximizing THC measures. Mile High Labs is traditionally a facility that caters to the Hemp industry and therefore not looking to maximize the measurement of THCA. Mile High Labs has a duty to accurately measure THCA and CBDA content following the same protocols as the DEA and USDA, as Hemp is defined by $THC_{TOTAL} < 0.3\%$, and crops that measure over that threshold are destroyed. However, because Mile High Labs is a Hemp testing facility, the protocol may not be calibrated to measure high amounts of THC, which we recognize as a limitation and therefore further investigation of the discrepancy is needed.

This study demonstrated significant variation in cannabinoids and terpenes even when samples are genetically identical, which could be annoying for recreational consumers expecting certain effects but is more problematic for those consuming to alleviate medical symptoms. This study included a relatively small sampling of *Cannabis*

strains, but nevertheless demonstrated genotype is not necessarily reflected as a predictable chemotype. In order to determine what is a reasonable amount of variation in chemotypes, more in-depth studies examining other environmental factors known to influence cannabinoid and terpene production such as growing conditions, harvest time, curing procedure and storage conditions, are needed.

Conclusion

The purpose of this study was to determine if genetic differences are mirrored as differences in chemical profiles of cannabinoids and terpenes within the same *Cannabis* strain. Based on these results, cannabinoids and terpene levels vary (1) among samples that are genetically identical, (2) within strains, and (3) among strains. The samples used in this study had variable chemotypes, but the variation did not seem to be linked to genetic identity. Given that chemotype is a phenotypic expression, and phenotype is the result of genotype and environment, the variation observed in cannabinoids and terpenes are most certainly the result of a combination of several variables. The *Cannabis* industry should aim to produce consistent products, especially for medicinal patients. There have been suggestions that focusing on chemotyping rather than using strain names for consistency would be a viable option. Minimizing phenotypic variation, including chemotype, would require standard growing conditions, harvesting protocols, curing techniques and storage conditions. However, the *Cannabis* industry currently has no such accepted standards for growing facilities and/or dispensaries, and therefore phytochemical variation within strains is inevitable, even if plants are genetically identical. Chemotyping for identity would be relatively easy if it were able to accurately identify and distinguish strains from one another. These results suggest that this is not the

case. Before chemotyping for identity can be utilized, it is crucial that industry standards are set and adhered to by every grower and producer. Until that happens, the only way to accurately identify *Cannabis* strains is through genotyping.

CHAPTER VI

CONCLUSION

Conclusion

Cannabis sativa is a highly versatile plant with variable phenotypes in fiber production, flower production, and phytochemical production (Clarke and Merlin 2013, 2016; Small 2015a; Small 2016). Artificial selection for desirable phenotypic characters throughout the history of the human relationship with *Cannabis* has led to genetic divergence between hemp and drug-types (Dufresnes et al. 2017; Henry 2015; Houston et al. 2017; Lynch et al. 2016; Sawler et al. 2015; Soler et al. 2017). *Cannabis* has largely been prohibited in the U.S. since the 1930's, but recent legalization in many states following growing public acceptance has led to a wave of new legislation allowing medical and adult recreational marijuana as well as hemp production. The *Cannabis* industry is growing at an unprecedented rate, but due to a tumultuous history, basic scientific knowledge and research on the plant is lacking. Not only that, but protections afforded to other plant varietals are not applicable to *Cannabis* and therefore stable genetic lines are relatively rare. Without taking the time and resources to stabilize the genetics through inbreeding, cloning is the preferred method for many growers to reproduce plants with desirable phenotypes. However, variation among plants that should be essentially identical has been identified as a problem (Chapter 2; Schwabe and McGlaughlin 2018) for the industry as they seek to provide customers with the quality and consistency they deserve. Variation within strains is of particular concern as

practitioners worldwide turn to *Cannabis* to treat a growing number of medical conditions.

The genetic, olfactory, and phytochemical research presented in this dissertation provides considerable evidence of both genotypic and phenotypic variation in *Cannabis sativa* (Chapters II-V). This investigation found variation is not limited to one source, but rather several possible sources. Potential origins for the genetic variation observed are numerous, but within strain variation was greater than expected (Chapter II), which is of particular concern for medical marijuana patients. Additionally, genotypic variation was not necessarily reflected in phenotypic variation, and genotypic cohesiveness among samples did not result in a consistent phenotype (Chapters IV and V). Although these studies were conducted using relatively small sample sizes, the results clearly demonstrate the need for regulatory systems.

Cannabis can be subdivided into two main categories: Hemp (<0.3% THC) and drug-types (>0.3% THC). The drug-type categories are further sub-divided into the commonly referenced Sativa, Indica, and Hybrid types, as well as the High CBD type which has lower THC levels but has >0.3% THC and is therefore assigned to the drug-type category. The commonly referenced Sativa, Indica and Hybrid types used to describe differences in psychoactive effects were not clearly resolved using genetic analysis (Chapter II and III), and this is probably due to extensive hybridizing of strains to create novel combinations with characteristics of both Sativa and Indica types. The genetic investigation had representative samples ranging from 100% Sativa to 100% Indica (Chapter II and III). However, the phytochemical portion of the study only included hybrids (“Blue Dream” 50:50, “Mob Boss” 50:50, “OG Kush” 55:45, “White

Widow" 60:40, "Tangerine Haze" 60:40, and "White Urkle" 50:50) and "Durban Poison", which is 100% Sativa. Therefore, investigating category assignment based on scent profile or phytochemical content was not possible in this study (Chapter IV and V). Hemp is genetically divergent from drug-types, although there are some drug-types that share a high degree of ancestry with Hemp types (Chapter III). The High CBD samples bridge the genetic division between Hemp and drug-types (Chapter III). One of the most important discoveries was that "research grade marijuana" supplied for medical studies in the U.S. is substantially different from drug-type samples purchased through the legal market (Chapter III). This has serious implications because medical studies researching medical applications using federally supplied marijuana are inherently flawed as medical patients do not have access to and are not consuming similar products.

Cannabis phytochemicals are abundant and 120 terpenes to date have been identified (ElSohly et al. 2017) which contribute to the diverse aromas found in *Cannabis*. The discovery of genetic inconsistencies within strains led us to investigate if genetic differences were expressed through detectable aromatic differences (Chapter IV). We purposefully chose four strains based on availability and previously determined unique aromatic profiles (Gilbert and DiVerdi 2018). There was considerable variation in the aromatic profiles within samples of the same strain and with identical genotypes but purchased from different locations (Chapter IV). However, analyses revealed the aromatic profile of the genetic outliers not only differed substantially from those with the same genotype, but also the aromatic profile of the outliers were uncharacteristic of previously described profiles (Gilbert and DiVerdi 2018; Leafly 2018b). Detectable differences in aromatic profiles of a genetic imposter suggest that genetic inconsistencies

observed within strains may not be simply a case of mistaken identity, but perhaps an indication of mislabeling or relabeling.

In addition to terpenes, *Cannabis* also has 120 known cannabinoids (ElSohly et al. 2017) which are responsible for physiological and psychoactive effects (Andre et al. 2016). Laws on potency testing for several cannabinoids vary from state to state, but THCA/THC and CBDA/CBD are a standard requirement (e.g.: Colorado Department of Revenue 2017). In order to further examine the relationship between genetic variation and chemotype, we analyzed eight cannabinoids and found variation in levels of phytochemicals, as well as the relative ratios of cannabinoids among plants with identical genotypes (Chapter V). Also, it appears that two cannabinoids, THCA and CBGA are the contributing to the majority of the variation (Chapter V). Variation in cannabinoids did not appear to align with genetic variation (or lack of variation). These results indicate that cannabinoid levels are likely substantially influenced by environmental conditions that vary among different growing facilities.

Sensory perception of the various aromas found in *Cannabis* is attributed to levels and combinations of terpenes produced in the flower trichomes. Olfaction and perceptions of smell are personal and subjective, and the subjects who participated in this research were not sensory experts. However, additional scientific examination was possible by analyzing the terpene profiles of the samples in the olfactory study. A pairwise analysis between sensory perception and terpene levels found 33 significant correlations (Table 6.1). This analysis shows perceived scents produced by the terpenes can be categorized: a-caryophyllene and b-caryophyllene are contributing to “Earthy” smells including soil, buttery, nutty or roasted aromas; guaiol and isopulegol contribute to

the “Spicy” including spices and dried leafy scents; neridiol-1, ocimene, p-cymene and terpinolene contribute to “Floral” scents such as fresh plants/flowers; neridiol-2 is the main contributor to “Pungent” scents including diesel, ammonia and chemical smells.

Table 6.1. Significant pairwise Pearson correlation coefficients (R^2) between terpene levels (Chapter V) and scents detected in the olfactory study (Chapter IV).

Smell	Terpene	R^2	P
Butter	a-caryophyllene	0.648	0.0226
Cheese	a-caryophyllene	0.632	0.0274
Coffee	a-caryophyllene	0.623	0.0304
Earthy	a-caryophyllene	0.623	0.0304
Menthol	a-gurjunene	0.584	0.0461
Butter	b-caryophyllene	0.741	0.0059
Coffee	b-caryophyllene	0.638	0.0257
Nutty	b-caryophyllene	0.583	0.0468
Sage	guaiol	0.740	0.0059
Pepper	isopulegol	0.582	0.0472
Menthol	limonene	0.584	0.0460
Coffee	linalool	0.794	0.0021
Butter	myrcene	0.673	0.0165
Flowery	nerolidol-1	0.632	0.0304
Mint	nerolidol-1	0.678	0.0154
Rose	nerolidol-1	0.721	0.0082
Sweet	nerolidol-1	0.607	0.0363
Violet	nerolidol-1	0.735	0.0065
Ammonia	nerolidol-2	0.694	0.0123
Chemical	nerolidol-2	0.640	0.0250
Diesel	nerolidol-2	0.620	0.0316
Pungent	nerolidol-2	0.602	0.0385
Flowery	ocimene	0.617	0.0326
Mint	ocimene	0.925	< 0.0001
Flowery	p-cymene	0.732	0.0068
Lemon	p-cymene	0.629	0.0285
Mint	p-cymene	0.848	0.0005
Tropical fruit	p-cymene	0.600	0.0392
Flowery	terpinolene	0.753	0.0047
Mint	terpinolene	0.821	0.0011
Rose	terpinolene	0.721	0.0082
Tropical fruit	terpinolene	0.641	0.0247
Violet	terpinolene	0.605	0.0373

The abundance of phytochemicals present in *Cannabis* is impressive and are one of the main drivers of the artificial selection process for desirable aromas and physiological effects. There is an expected amount of variation among organisms of the same genetic lineage, as phenotypic variation is the product of genotype and environment. Therefore, individuals with identical genotypes are expected to have highly similar phenotypes. However, the extent to which environmental variation impacts phenotype is largely unknown. Analyses of both terpene and cannabinoid profiles among individuals with identical genotypes acquired from different sources indicate environmental variation has a substantial impact on phenotype in *Cannabis.* Considering these results, medical marijuana patients are unlikely to have access to consistent products, even if the genotype has been verified. Not only are potency levels variable, but variation in phytochemical constituents detected in the samples is evident.

The *Cannabis* industry needs a system to verify products to ensure consumers are provided the product as indicated by the name provided. There have been suggestions that due to strain name unreliability, describing products based on chemotype may be a solution. However, I would caution against this approach as there are 560 chemical constituents in *Cannabis* and analytical labs only test a small portion of cannabinoids and terpenes, ignoring the vast majority of the micro-chemotype which likely contribute to differences in aromatic profiles and effects (Amirault and Boyar 2019). Additionally, the variation in samples with identical genotypes indicates varying environmental conditions are influencing chemical profiles. *Cannabis* consumers, practitioners, breeders, and growers need to be able to communicate about varietals, and the most familiar method to

do this is to give varietals a name. It would be counter intuitive and incredibly difficult to discuss *Cannabis* flowers based on a partial chemotype that changes over time.

Future Directions

The *Cannabis* industry exists to produce products for a variety of markets. The industrial hemp industry grows crops for fiber, oils, and seeds which can be transformed into thousands of products. Because many industrial hemp products are produced from processed plant material, consistency among plants is not imperative. The marijuana *Cannabis* industry grows plants for human consumption, some of which are processed, such as isolates, tinctures and edibles. However, there is a large proportion of products on dispensary shelves that are sold under specific names and are non-processed flowers. These products are intended to be smoked and will have some effect on the consumer. Recreational adult use is associated with the psychotropic responses to partaking in smoking marijuana. Medicinal use may include psychotropic responses, but more importantly, the medical consumer is seeking to alleviate symptoms related to particular medical conditions. Medicine needs to be reliable and consistent, and current *Cannabis* products sampled in this study are not reliable or consistent. Producers strive to produce a consistent phenotype through cloning, but it is apparent that environmental influences have a large effect. A standard genotyping procedure is needed to confirm identity in addition to the chemotype tests currently in place that determine potency. Also, a set of standard growing conditions needs to be established, as well as standard harvesting and curing procedures.

Studies have shown that environmental stresses can lead to phenotypic changes in plants. I am unaware of any published work investigating how environmental stress

effects the epigenome and the phenotypic consequences of those changes in clonal plants. It is possible that the stress of the cloning process, on both the plant from which clones are cut as well as the resulting progeny, will result in phenotypic changes due to epigenetic changes. This is certainly a possible source for variation in *Cannabis* and warrants investigation. Additionally, a system to protect intellectual property of breeders who develop genetically stable lines (varietals) would limit variation from cloning and allow desirable lineages to persist into the future.

The results of these studies suggest the industry should implement regulatory checks in the form of genetic testing in order to provide consistency, especially for medical applications. Current required testing includes pesticide and potency analysis and reporting, and genetic tests could be implemented to verify products. In order to provide consumers consistent products, it is imperative to understand sources of variation. Phenotypic variation is unavoidable when genotype has not been verified. Following genotypic confirmation, it is possible to create phenotypic consistency if standard growing conditions can be established. This would not stifle the ability for growers to develop alternative conditions to produce different desirable characters such as larger flowers or higher terpene content, but it would allow for the production of consistent products if standard conditions are met. Following verification and established standards in protocols, deviations resulting in phenotypic changes should be disclosed to consumers so they are aware that there may be deviations from expected effects. Consumers deserve to be provided with quality consistent products as the industry continues to thrive on a global scale.

REFERENCES

75th Congress of the United States. 1937. Marihuana Tax Act of 1937. 75th Congress, United States of America. pp. 551-557.

Abel, E.L. 2013. Marihuana: the first twelve thousand years. Springer Science & Business Media.

Abuhasira, R., Schleider, L.B.-L., Mechoulam, R., and Novack, V. 2018. Epidemiological Characteristics, Safety and Efficacy of Medical Cannabis in the Elderly. European Journal of Internal Medicine **49**: 44-50. doi:10.1016/j.ejim.2018.01.019.

Aizpurua-Olaizola, O., Omar, J., Navarro, P., Olivares, M., Etxebarria, N., and Usobiaga, A. 2014. Identification and quantification of cannabinoids in *Cannabis sativa* L. plants by high performance liquid chromatography-mass spectrometry. Analytical and Bioanalytical Chemistry **406**(29): 7549-7560. doi:10.1007/s00216-014-8177-x.

Aizpurua-Olaizola, O., Soydaner, U., Ozturk, E., Schibano, D., Simsir, Y., Navarro, P., Etxebarria, N., and Usobiaga, A. 2016. Evolution of the Cannabinoid and Terpene Content during the Growth of *Cannabis sativa* Plants from Different Chemotypes. Journal of Natural Products **79**(2): 324-331. doi:10.1021/acs.jnatprod.5b00949.

Alexander, S.P.H. 2016. Therapeutic potential of cannabis-related drugs. Progress in Neuro-Psychopharmacology & Biological Psychiatry **64**: 157-166. doi:10.1016/j.pnpbp.2015.07.001.

Alghanim, H.J., and Almirall, J.R. 2003. Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Analytical and Bioanalytical Chemistry **376**(8): 1225-1233. doi:10.1007/s00216-003-1984-0.

Amirault, B., and Boyar, K. 2019. The Intersection of Blockchain and Genetics: Protecting Your Cannabis IP. *In* Terpenes & Testing; The Horticultural Issue. MACE Media Group, Pismo Beach, California. pp. 42-45.

Anderson, L.C. 1980. Leaf variation among *Cannabis* species from a controlled garden. Botanical Museum leaflets. Harvard University Botanical Museum Leaflets, Boston, Massachusetts.

Andre, C.M., Hausman, J.F., and Guerriero, G. 2016. Cannabis sativa: The Plant of the Thousand and One Molecules. Frontiers in Plant Science **7**. doi:10.3389/fpls.2016.00019.

Baldoni, L., Cultrera, N.G., Mariotti, R., Ricciolini, C., Arcioni, S., Vendramin, G.G., Buonamici, A., Porceddu, A., Sarri, V., Ojeda, M.A., Trujillo, I., Rallo, L., Belaj, A., Perri, E., Salimonti, A., Muzzalupo, I., Casagrande, A., Lain, O., Messina, R., and Testolin, R. 2009. A consensus list of microsatellite markers for olive genotyping. Molecular Breeding **24**(3): 213-231. doi:10.1007/s11032-009-9285-8.

Banerjee, S.P., Snyder, S.H., and Mechoulam, R. 1975. Cannabinoids - Influence on Neurotransmitter Uptake in Rat-Brain Synaptosomes. Journal of Pharmacology and Experimental Therapeutics **194**(1): 74-81.

Baron, E.P. 2018. Medicinal Properties of Cannabinoids, Terpenes, and Flavonoids in *Cannabis*, and Benefits in Migraine, Headache, and Pain: An Update on Current

Evidence and Cannabis Science. Headache **58**(7): 1139-1186. doi:10.1111/head.13345.

Belaj, A., Cipriani, G., Testolin, R., Rallo, L., and Trujillo, I. 2004. Characterization and identification of the main Spanish and Italian olive cultivars by simple-sequence-repeat markers. Hortscience **39**(7): 1557-1561.

Ben-Shabat, S., Fride, E., Sheskin, T., Tamiri, T., Rhee, M.H., Vogel, Z., Bisogno, T., De Petrocellis, L., Di Marzo, V., and Mechoulam, R. 1998. An entourage effect: inactive endogenous fatty acid glycerol esters enhance 2-arachidonoyl-glycerol cannabinoid activity. European Journal of Pharmacology **353**(1): 23-31. doi:10.1016/s0014-2999(98)00392-6.

Booth, M. 2004. Cannabis: a history. St Martin's Press, New York.

Borgelt, L.M., Franson, K.L., Nussbaum, A.M., and Wang, G.S. 2013. The Pharmacologic and Clinical Effects of Medical Cannabis. Pharmacotherapy **33**(2): 195-209. doi:10.1002/phar.1187.

Brighenti, V., Pellati, F., Steinbach, M., Maran, D., and Benvenuti, S. 2017. Development of a new extraction technique and HPLC method for the analysis of non-psychoactive cannabinoids in fibre-type *Cannabis sativa* L. (hemp). Journal of Pharmaceutical and Biomedical Analysis **143**: 228-236. doi:10.1016/j.jpba.2017.05.049.

Callaway, J.C. 2004. Hempseed as a nutritional resource: An overview. Euphytica **140**(1-2): 65-72. doi:10.1007/s10681-004-4811-6.

Casano, S., Grassi, G., Martini, V., and Michelozzi, M. 2011. Variations in Terpene

Profiles of Different Strains of *Cannabis sativa* L. *In* ISHS Acta Horiculturae 925: XXVIII International Horticultural Congress on Science and Horticulature for People (IHC2010): A New Look at Medicinal and Aromatic Plants Seminar. pp. 115-121.

Cervantes, J. 2006. Marijuana Horticulture: The indoor/outdoor medical grower's bible. Van Patten Publishing.

Cipriani, G., Marrazzo, M.T., Marconi, R., Cimato, A., and 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**(2-3): 223-228. doi:10.1007/s001220100685.

Citti, C., Braghiroli, D., Vandelli, M.A., and Cannazza, G. 2018. Pharmaceutical and biomedical analysis of cannabinoids: A critical review. Journal of Pharmaceutical and Biomedical Analysis **147**: 565-579. doi:10.1016/j.jpba.2017.06.003.

Clarke, R.C., and Merlin, M.D. 2013. Cannabis: evolution and ethnobotany. University of California Press, Berkeley.

Clarke, R.C., and Merlin, M.D. 2015. Letter to the Editor: Small, Ernest. 2015. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. Botanical Review 81(3): 189-294. Botanical Review **81**(4): 295-305. doi:10.1007/s12229-015-9158-2.

Clarke, R.C., and Merlin, M.D. 2016. *Cannabis* Domestication, Breeding History, Present-day Genetic Diversity, and Future Prospects. Critical Reviews in Plant Sciences **35**(5-6): 293-327. doi:10.1080/07352689.2016.1267498.

Colorado Department of Revenue. 2017. **M 712 – Medical Marijuana Testing Facilities: Sampling and Testing Program**. *Edited by* M.E. Division, Colorado.

Costa, B., Trovato, A.E., Comelli, F., Giagnoni, G., and Colleoni, M. 2007. The non-psychoactive cannabisconstituent cannabidiol is an orally effective therapeutic agent in rat chronic inflammatory and neuropathic pain. European Journal of Pharmacology **556**: 75-83.

Costantini, L., Monaco, A., Vouillamoz, J.F., Forlani, M., and Grando, M.S. 2005. Genetic relationships among local *Vitis vinifera* cultivars from Campania (Italy). Vitis **44**(1): 25-34.

Cousijn, J., Nunez, A.E., and Filbey, F.M. 2018. Time to acknowledge the mixed effects of cannabis on health: a summary and critical review of the NASEM 2017 report on the health effects of cannabis and cannabinoids. Addiction **113**(5): 958-966. doi:10.1111/add.14084.

Coyle, H.M., Palmbach, T., Juliano, N., Ladd, C., and Lee, H.C. 2003. An overview of DNA methods for the identification and individualization of marijuana. Croatian Medical Journal **44**(3): 315-321.

Danko, D. 2016. 25 Years of Chem Dog. Retrieved from https://hightimes.com/grow/25-years-of-chem-dog/. [accessed August 2018]

Datwyler, S.L., and Weiblen, G.D. 2006. Genetic variation in hemp and marijuana (*Cannabis sativa* L.) according to amplified fragment length polymorphisms. Journal of Forensic Sciences **51**(2): 371-375. doi:10.1111/j.1556-4029.2006.00061.x.

De Backer, B., Debrus, B., Lebrun, P., Theunis, L., Dubois, N., Decock, L., Verstraete, A., Hubert, P., and Charlier, C. 2009. Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **877**(32): 4115-4124. doi:10.1016/j.jchromb.2009.11.004.

de Boubee, D.R., Van Leeuwen, C., and Dubourdieu, D. 2000. Organoleptic impact of 2-methoxy-3-isobutylpyrazine on red Bordeaux and Loire wines. Effect of environmental conditions on concentrations in grapes during ripening. Journal of Agricultural and Food Chemistry **48**(10): 4830-4834.

de Lamarck, J.D. 1785. Botanique. Encyclopedie methodique, Panckoucke, Paris. pp. 694–695.

De Meijer, E., and Keizer, L. 1996. Patterns of diversity in *Cannabis*. Genetic resources and crop evolution **43**(1): 41-52.

de Meijer, E.P.M., Vanderkamp, H.J., and Vaneeuwijk, F.A. 1992. Characterization of *Cannabis* accessions with regard to cannabinoid content in relation to other plant characters. Euphytica **62**(3): 187-200. doi:10.1007/bf00041753.

de Meijer, E.P.M., and Hammond, K.M. 2005. The inheritance of chemical phenotype in *Cannabis sativa* L. (II): Cannabigerol predominant plants. Euphytica **145**(1-2): 189-198. doi:10.1007/s10681-005-1164-8.

de Meijer, E.P.M., Hammond, K.M., and Sutton, A. 2009a. The inheritance of chemical phenotype in *Cannabis sativa* L. (IV): cannabinoid-free plants. Euphytica **168**(1): 95-112. doi:10.1007/s10681-009-9894-7.

de Meijer, E.P.M., Hammond, K.M., and Micheler, M. 2009b. The inheritance of chemical phenotype in *Cannabis sativa* L. (III): variation in cannabichromene proportion. Euphytica **165**(2): 293-311. doi:10.1007/s10681-008-9787-1.

de Oliveira, G.L., Voloch, M.H., Sztulman, G.B., Negrini Neto, O., and Yonamine, M. 2008. Cannabinoid contents in cannabis products seized in Sao Paulo, Brazil, 2006-2007. Forensic Toxicology **26**(1): 31-35. doi:10.1007/s11419-008-0046-x.

De Padua, L.S., Bunyapraphatsō̦n, N., and Lemmens, R.H.M.J. 1999. Medicinal and poisonous plants. Plant resources of South-East Asia. No. 12. Backhuys Publishers, Leiden, Netherlands. pp. 169.

Desjardins, A.E. 2008. Natural product chemistry meets genetics: when is a genotype a chemotype? Journal of agricultural and food chemistry **56**(17): 7587-7592.

Doyle, J.J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue Phytochemical Bulletin **4(2)**: 359-361.

Dufresnes, C., Jan, C., Bienert, F., Goudet, J., and Fumagalli, L. 2017. Broad-Scale Genetic Diversity of *Cannabis* for Forensic Applications. Plos One **12**(1). doi:10.1371/journal.pone.0170522.

Duran, M., Perez, E., Abanades, S., Vidal, X., Saura, C., Majem, M., Arriola, E., Rabanal, M., Pastor, A., Farre, M., Rams, N., Laporte, J.-R., and Capella, D. 2010. Preliminary efficacy and safety of an oromucosal standardized cannabis extract in chemotherapy-induced nausea and vomiting. British Journal of Clinical Pharmacology **70**(5): 656-663. doi:10.1111/j.1365-2125.2010.03743.x.

Earl, D.A., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno

method. Conservation Genetics Resources **4**(2): 359-361. doi:10.1007/s12686-011-9548-7.

ElSohly, H., Turner, C.E., Clark, A.M., and ElSohly, M.A. 1982. Synthesis and antimicrobial properties of certain cannabichrome and cannabigerol related compounds. **71**: 1319–1323.

ElSohly, M.A., and Slade, D. 2005. Chemical constituents of marijuana: The complex mixture of natural cannabinoids. Life Sciences **78**(5): 539-548. doi:10.1016/j.lfs.2005.09.011.

ElSohly, M.A. 2007. Marijuana & the Cannabinoids. Humana Press, Totowa, New Jersey.

ElSohly, M.A., Mehmedic, Z., Foster, S., Gon, C., Chandra, S., and Church, J.C. 2016. Changes in *Cannabis* Potency Over the Last 2 Decades (1995-2014): Analysis of Current Data in the United States. Biological Psychiatry **79**(7): 613-619. doi:10.1016/j.biopsych.2016.01.004.

ElSohly, M.A., Radwan, M.M., Gul, W., Chandra, S., and Galal, A. 2017. Phytochemistry of *Cannabis sativa* L. Progress in the chemistry of organic natural products **103**: 1-36. doi:10.1007/978-3-319-45541-9_1.

Elzinga, S., Fischedick, J., Podkolinski, R., and Raber, J.C. 2015. Cannabinoids and terpenes as chemotaxonomic markers in cannabis. Natural Products Chemistry & Research **3**(181): 2.

Emboden, W.A. 1974. *Cannabis*—a polytypic genus. Economic Botany **28**(3): 304-310.

Emboden, W.A. 1977. Taxonomy For *Cannabis*. Taxon **26**(1): 110-110. doi:10.2307/1220203.

Emboden, W.A. 1981. The Genus *Cannabis* and the Correct Use of Taxonomic Categories. Journal of Psychoactive Drugs **13**(1): 15-21. doi:10.1080/02791072.1981.10471446.

Escondido, N. 2014. 25 Greatest Strains of All Time. Available from https://hightimes.com/grow/25-greatest-strains-of-all-time/.

European Monitoring Centre for Drugs and Drug Addiction. 2018. *Cannabis* legislation in Europe: An Overview. *Edited by* B. Hughes. Publications Office of the European Union, Luxembourg. pp. 1-30.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**(8): 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x.

Faircloth, B.C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources **8**(1): 92-94. doi:10.1111/j.1471-8286.2007.01884.x.

Farrimond, J.A., Whalley, B.J., and Williams, C.M. 2012. Cannabinol and cannabidiol exert opposing effects on rat feeding patterns. Psychopharmacology **223**(1): 117-129. doi:10.1007/s00213-012-2697-x.

Faux, A.M., Berhin, A., Dauguet, N., and Bertin, P. 2014. Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica **196**(2): 183-197. doi:10.1007/s10681-013-1023-y.

Faux, A.M., Draye, X., Flamand, M.C., Occre, A., and Bertin, P. 2016. Identification of QTLs for sex expression in dioecious and monoecious hemp (*Cannabis sativa* L.). Euphytica **209**(2): 357-376. doi:10.1007/s10681-016-1641-2.

Figueiredo, A.C., Barroso, J.G., Pedro, L.G., and Scheffer, J.J.C. 2008. Factors affecting secondary metabolite production in plants: volatile components and essential oils. Flavour and Fragrance Journal **23**(4): 213-226. doi:10.1002/ffj.1875.

Fischedick, J.T., Hazekamp, A., Erkelens, T., Choi, Y.H., and Verpoorte, R. 2010. Metabolic fingerprinting of *Cannabis sativa* L, cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. Phytochemistry **71**(17-18): 2058-2073. doi:10.1016/j.phytochem.2010.10.001.

Forapani, S., Carboni, A., Paoletti, C., Cristiana, V.M., Ranalli, N.P., and Mandolino, G. 2001. Comparison of hemp varieties using random amplified polymorphic DNA markers. Crop Science **41**(6): 1682-1689. doi:10.2135/cropsci2001.1682.

Gabriel, W., Lynch, M., and Burger, R. 1993. Mullers Ratchet and Mutational Meltdowns. Evolution **47**(6): 1744-1757. doi:10.2307/2410218.

Gallily, R., Even-Chen, T., Katzavian, G., Lehmann, D., Dagan, A., and Mechoulam, R. 2003. gamma-irradiation enhances apoptosis induced by cannabidiol, a non-psychotropic cannabinoid, in cultured HL-60 myeloblastic leukemia cells. Leukemia & Lymphoma **44**(10): 1767-1773. doi:10.1080/1042819031000103917.

Gao, C.S., Xin, P.F., Cheng, C.H., Tang, Q., Chen, P., Wang, C.B., Zang, G.G., and Zhao, L.N. 2014. Diversity Analysis in *Cannabis sativa* Based on Large-Scale Development of Expressed Sequence Tag-Derived Simple Sequence Repeat Markers. Plos One **9**(10). doi:10.1371/journal.pone.0110638.

Gerner, M. 2015. Uruguay's Advances in Drug Policy. International Journal on Human Rights **21**: 1-4. Available from https://sur.conectas.org/en/uruguays-advances-drug-policy/.

Giese, M.W., Lewis, M.A., Giese, L., and Smith, K.M. 2015. Development and Validation of a Reliable and Robust Method for the Analysis of Cannabinoids and Terpenes in *Cannabis*. Journal of Aoac International **98**(6): 1503-1522. doi:10.5740/jaoacint.15-116.

Gilbert, A.N., and DiVerdi, J.A. 2018. Consumer perceptions of strain differences in *Cannabis* aroma. Plos One **13**(2). doi:10.1371/journal.pone.0192247.

Gilmore, S., and Peakall, R. 2003. Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana). Molecular Ecology Notes **3**(1): 105-107. doi:10.1046/j.1471-8286.2003.00367.x.

Gilmore, S., Peakall, R., and Robertson, J. 2003. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Science International **131**(1): 65-74. doi:10.1016/s0379-0738(02)00397-3.

Grassa, C.J., Wenger, J.P., Danny, C., Poplawski, S.G., Motley, T.S., Michael, T.P., Schwartz, C.J., and Weiblen, G.D. 2018. A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. bioRxiv preprint. doi: https://doi.org/10.1101/458083.

Green, G. 2005. The *Cannabis* Breeder's Bible. Green Candy Press, San Francisco.

Green House Seed Company. 2018. The Flavor Wheel. Available from https://shop.greenhouseseeds.nl/flavour-wheel.html?sl=en.

Griess, T. 2016. Cultivar vs. Variety. Yard and Garden. Available from https://extension.unl.edu/statewide/buffalo/Yard/Cultivar%20vs%20Variety%200 1-30-2016.pdf [accessed January 12 2019].

Guilford, P., Prakash, S., Zhu, J.M., Rikkerink, E., Gardiner, S., Bassett, H., and Forster, R. 1997. Microsatellites in Malus X domestica (apple): Abundance, polymorphism and cultivar identification. Theoretical and Applied Genetics **94**(2): 249-254. doi:10.1007/s001220050407.

Guindon, J., and Hohmann, A.G. 2008. Cannabinoid CB2 receptors: a therapeutic target for the treatment of inflammatory and neuropathic pain. British Journal of Pharmacology **153**(2): 319-334. doi:10.1038/sj.bjp.0707531.

Gul, W., Gul, S.W., Radwan, M.M., Wanas, A.S., Mehmedic, Z., Khan, I.I., Sharaf, M.H.M., and ElSohly, M.A. 2015. Determination of 11 Cannabinoids in Biomass and Extracts of Different Varieties of *Cannabis* Using High-Performance Liquid Chromatography. Journal of Aoac International **98**(6): 1523-1528. doi:10.5740/jaoacint.15-095.

Hakki, E.E., Kayis, S.A., Pinarkara, E., and Sag, A. 2007. Inter simple sequence repeats separate efficiently hemp from marijuana (*Cannabis sativa* L.). Electronic Journal of Biotechnology **10**(4): 570-581. doi:10.2225/vol10-issue4-fulltext-4.

Halford, J.C.G., and Harrold, J.A. 2008. Neuropharmacology of human appetite expression. Developmental Disabilities Research Reviews **14**(2): 158-164. doi:10.1002/ddrr.20.

Han, B.H., and Palamar, J.J. 2018. Marijuana use by middle-aged and older adults in the United States, 2015-2016 (vol 191, pg 374, 2018). Drug and Alcohol Dependence **192**: 171-171. doi:10.1016/j.drugalcdep.2018.09.001.

Hazekamp, A., Peltenburg, A., Verpoorte, R., and Giroud, C. 2005. Chromatographic and spectroscopic data of cannabinoids from *Cannabis sativa* L. Journal of Liquid

Chromatography & Related Technologies **28**(15): 2361-2382. doi:10.1080/10826070500187558.

Hazekamp, A., and Fischedick, J.T. 2012. *Cannabis* - from cultivar to chemovar. Drug Testing and Analysis **4**(7-8): 660-667. doi:10.1002/dta.407.

Henry, P. 2015. Genome-wide analyses reveal clustering in *Cannabis* cultivars: the ancient domestication trilogy of a panacea. Peer J Preprints https://peerj.com/preprints/1553.pdf.

Hill, K.P. 2015. Medical Marijuana for Treatment of Chronic Pain and Other Medical and Psychiatric Problems A Clinical Review. Jama-Journal of the American Medical Association **313**(24): 2474-2483. doi:10.1001/jama.2015.6199.

Hillig, K.W., and Mahlberg, P.G. 2004. A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). American Journal of Botany **91**(6): 966-975. doi:10.3732/ajb.91.6.966.

Hillig, K.W. 2005. Genetic evidence for speciation in *Cannabis* (Cannabaceae). Genetic Resources and Crop Evolution **52**(2): 161-180. doi:10.1007/s10722-003-4452-y.

Hojsgaard, D., and Horandl, E. 2015. A little bit of sex matters for genome evolution in asexual plants. Frontiers in plant science **6**: 82-82. doi:10.3389/fpls.2015.00082.

Hokanson, S.C., Szewc-McFadden, A.K., Lamboy, W.F., and McFerson, J.R. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* x *domestica* borkh. core subset collection. Theoretical and Applied Genetics **97**(5-6): 671-683. doi:10.1007/s001220050943.

Houston, R., Birck, M., Hughes-Stamm, S., and Gangitano, D. 2016. Evaluation of a 13-loci STR multiplex system for *Cannabis sativa* genetic identification.

International Journal of Legal Medicine **130**(3): 635-647. doi:10.1007/s00414-015-1296-x.

IBM Corp. 2016. IBM SPSS Statistics for Windows. IBM Corp., Armonk, NY.

Integrated Taxonomic Information System. 2014. Cannabaceae. *In* Taxonomic Serial No: 19118.

Jackson, D.I., and Lombard, P.B. 1993. Environmental and Management-Practices Affecting Grape Composition and Wine Quality - A Review. American Journal of Enology and Viticulture **44**(4): 409-430.

Jaeger, K. 2018. Surgeon General Says Marijuana's Schedule 1 Status Hinders Research. Available from https://www.marijuanamoment.net/surgeon-general-says-marijuanas-schedule-i-status-hinders-research/ [accessed January 7, 2019 2018].

Jensen, B., Chen, J., Furnish, T., and Wallace, M. 2015. Medical Marijuana and Chronic Pain: a Review of Basic Science and Clinical Evidence. Current Pain and Headache Reports **19**(10). doi:10.1007/s11916-015-0524-x.

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**. doi:10.1038/s41598-018-22755-2.

Johnson, Bankole, and A. 2010. Addiction Medicine: Science and Practice. Springer Science & Business Media, Springer Science & Business Media. p. 303.

Khan, I. 2018. Criticism of U. of Mississippi Marijuana Quality Unfair. Available from https://www.chronicle.com/blogs/letters/criticism-of-u-of-mississippi-marijuana-quality-unfair/ [accessed April 2, 2017].

Kivela, M., Arnaud-Haond, S., and Saramaki, J. 2015. EDENetworks: A user-friendly software to build and analyse networks in biogeography, ecology and population genetics. Molecular Ecology Resources **15**(1): 117-122. doi:10.1111/1755-0998.12290.

Knight, G., Hansen, S., Connor, M., Poulsen, H., McGovern, C., and Stacey, J. 2010. The results of an experimental indoor hydroponic *Cannabis* growing study, using the 'Screen of Green' (ScrOG) method-Yield, tetrahydrocannabinol (THC) and DNA analysis. Forensic Science International **202**(1-3): 36-44. doi:10.1016/j.forsciint.2010.04.022.

Kojoma, M., Seki, H., Yoshida, S., and Muranaka, T. 2006. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" *Cannabis sativa* L. Forensic Science International **159**(2-3): 132-140. doi:10.1016/j.forsciint.2005.07.005.

Konopasek, M. 2019. Local hemp manufacturer claims $500K-worth of legal hemp confiscated in Oklahoma. Fox 3, Denver.

Laverty, K.U., Stout, J.M., Sullivan, M.J., Shah, H., Gill, N., Holbrook, L., Deikus, G., Sebra, R., Hughes, T.R., Page, J.E., and van Bakel, H. 2019. A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Research **29**(1): 146-156. doi:10.1101/gr.242594.118.

Leafly. 2018a. Medical Marijuana Patient Numbers. Available from https://www.mpp.org/issues/medical-marijuana/state-by-state-medical-marijuana-laws/medical-marijuana-patient-numbers/ [accessed September 12 2018].

Leafly. 2018b. *Cannabis* Strain Explorer. Available from https://www.leafly.com/explore [accessed September 12 2018].

Ligresti, A., Moriello, A.S., Starowicz, K., Matias, I., Pisanti, S., De Petrocellis, L., Laezza, C., Portella, G., Bifulco, M., and Di Marzo, V. 2006. Antitumor activity of plant cannabinoids with emphasis on the effect of cannabidiol on human breast carcinoma. Journal of Pharmacology and Experimental Therapeutics **318**(3): 1375-1387. doi:10.1124/jpet.106.105247.

Lynch, M., and Ritland, K. 1999. Estimation of pairwise relatedness with molecular markers. Genetics **152**(4): 1753-1766.

Lynch, R.C., Vergara, D., Tittes, S., White, K., Schwartz, C.J., Gibbs, M.J., Ruthenburg, T.C., deCesare, K., Land, D.P., and Kane, N.C. 2016. Genomic and Chemical Diversity in *Cannabis*. Critical Reviews in Plant Sciences **35**(5-6): 349-363. doi:10.1080/07352689.2016.1265363.

Maa, E., and Figi, P. 2014. The case for medical marijuana in epilepsy. Epilepsia **55**(6): 783-786. doi:10.1111/epi.12610.

Mandolino, G., Carboni, A., Forapani, S., Faeti, V., and Ranalli, P. 1999. Identification of DNA markers linked to the male sex in dioecious hemp (*Cannabis sativa* L.). Theoretical and Applied Genetics **98**(1): 86-92. doi:10.1007/s001220051043.

Marijuana Business Daily. 2017. Marijuana Business Factbook 2017. Available from https://mjbizdaily.com/wp-content/uploads/2017/05/Factbook2017ExecutiveSummary.pdf Anne Holland Ventures Inc.

Marijuana strains database. 2019 Available from https://www.cannabis.info/en/strains

[accessed January 2019].

Mariotti, K.D., Marcelo, M.C.A., Ortiz, R.S., Borille, B.T., dos Reis, M., Fett, M.S., Ferrao, M.F., and Limberger, R.P. 2016. Seized cannabis seeds cultivated in greenhouse: A chemical study by gas chromatography-mass spectrometry and chemometric analysis. Science & Justice **56**(1): 35-41. doi:10.1016/j.scijus.2015.09.002.

McCune, B., and Mefford, M.J. 1999. PC-ORD: multivariate analysis of ecological data; Version 4 for Windows; User's Guide. MjM software design.

McKallip, R.J., Jia, W., Schlomer, J., Warren, J.W., Nagarkatti, P.S., and Nagarkatti, M. 2006. Cannabidiol-induced apoptosis in human leukemia cells: A novel role of cannabidiol in the regulation of p22(phox) and Nox4 expression. Molecular Pharmacology **70**(3): 897-908. doi:10.1124/mol.106.023937.

McPartland, J., Di Marzo, V., De Petrocellis, L., Mercer, A., and Glass, M. 2001. Cannabinoid receptors are absent in insects. Journal of Comparative Neurology **436**(4): 423-429. doi:10.1002/cne.1078.

McPartland, J.M. 2006. Commentary on: Datwyler SL, Weiblen GD. Weiblen. Genetic variation in hemp and marijuana (*Cannabis sativa* L.) according to amplified fragment length polymorphisms. Journal of Forensic Sciences **51**(6): 1405-1405. doi:10.1111/j.1556-4029.2006.00276.x.

McPartland, J.M. 2017. *Cannabis sativa* and *Cannabis indica* versus "Sativa" and "Indica". *In* Botany and Biotechnology. *Edited by* S. Chandra and H. Lata and M. ElSohly. Springer, Cham.

McPartland, J.M., and Guy, G.W. 2017. Models of *Cannabis* Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names. Botanical Review **83**(4): 327-381. doi:10.1007/s12229-017-9187-0.

McPartland, P. 2018. *Cannabis* Systematics at the Levels of Family, Genus, and Species. Cannabis and Cannabinoid Research. Published online 2018 Oct 1. doi: 10.1089/can.2018.0039.

Mechoulam, R., Benshabat, S., Hanus, L., Ligumsky, M., Kaminski, N.E., Schatz, A.R., Gopher, A., Almog, S., Martin, B.R., Compton, D.R., Pertwee, R.G., Griffin, G., Bayewitch, M., Barg, J., and Vogel, Z. 1995. Identification of an Endogenous 2-Monoglyceride, Present in Canine Gut, That Binds to Cannabinoid Receptors. Biochemical Pharmacology **50**(1): 83-90. doi:10.1016/0006-2952(95)00109-d.

Meier, C., and Mediavilla, V. 1998. Factors influencing the yield and the quality of hemp (*Cannabis sativa* L.) essential oil. **5(1)**: 16-20.

Minkin, R.M. 2014. Cannabis Pharmacy: The Practical Guide to Medical Marijuana. Library Journal **139**(15): 98-98.

Mississippi, T.U.o. 2017. Marijuana Research NIDA Contract. Available from http://pharmacy.olemiss.edu/marijuana/nida-contract/ [accessed December 6 2018].

Moliterni, V.M.C., Cattivelli, L., Ranalli, P., and Mandolino, G. 2004. The sexual differentiation of *Cannabis sativa* L.: A morphological and molecular study. Euphytica **140**(1-2): 95-106. doi:10.1007/s10681-004-4758-7.

Mostyn, B., Gibbon, H., and Cowdery, N. 2012. The criminalisation of drugs and the search for alternative approaches. Current Issues in Criminal Justice **24(2)**: 261-272.

Moura, A.E., Natoli, A., Rogan, E., and Hoelzel, A.R. 2013. Atypical panmixia in a European dolphin species (*Delphinus delphis*): implications for the evolution of diversity across oceanic boundaries. Journal of Evolutionary Biology **26**(1): 63-75. doi:10.1111/jeb.12032.

Moyston, L. 2013. The ganja law of 1913: 100 years of oppressive injustice. Jamaica Observer.

Musty, R.E.K., I.G., Shirikawa, I.T., R. N., and Knobel, E. 1976. Interactions of delta-9-tetrahydrocannabinol and cannabinol in man. Pharmacology of marihuana, New York, New York. pp. 559-563.

Muzzalupo, I., Stefanizzi, F., and Perri, E. 2009. Evaluation of Olives Cultivated in Southern Italy by Simple Sequence Repeat Markers. Hortscience **44**(3): 582-588.

Naftali, T., Schleider, L.B.-L., Dotan, I., Lansky, E.P., Benjaminov, F.S., and Konikoff, F.M. 2013. *Cannabi*s Induces a Clinical Response in Patients With Crohn's Disease: A Prospective Placebo-Controlled Study. Clinical Gastroenterology and Hepatology **11**(10): 1276-1280. doi:10.1016/j.cgh.2013.04.034.

National Academies of Sciences, E., and Medicine,. 2017. The Health Effects of *Cannabis* and Cannabinoids: The Current State of Evidence and Recommendations for Research. The National Academies Press, Washington, DC.

National Conference of State Legislatures. 2018. Marijuana Overview: Legislations. Available from http://www.ncsl.org/research/civil-and-criminal-justice/marijuana-overview.aspx.

National Institute of Health, and National Institute on Drug Abuse. 2018a. Funding: Project Listing by Category, Cannabinoid Research. Available from https://report.nih.gov/categorical_spending_project_listing.aspx?FY=2016&ARRA=N&DCat=Cannabinoid+Research [accessed November 15, 2018].

National Institute of Health, and National Institute on Drug Abuse. 2018b. Funding: Project Listing by Category, Cannabidiol Research. Available from https://report.nih.gov/categorical_spending_project_listing.aspx?FY=2015&ARRA=N&DCat=Cannabidiol%20Research [accessed November 15, 2018].

National Institute on Drug Abuse. 2018. NIDA's Role in Providing Marijuana for Research. Available from https://www.drugabuse.gov/drugs-abuse/marijuana/nidas-role-in-providing-marijuana-research [accessed November 15, 2018].

NCSM: Nederlandse Associatie voor legale Cannabis. 2018. Strains. from https://www.ncsm.nl/strain [accessed March, 2019].

Norman, A.J., Stronen, A.V., Fuglstad, G.-A., Ruiz-Gonzalez, A., Kindberg, J., Street, N.R., and Spong, G. 2017. Landscape relatedness: detecting contemporary fine-scale spatial structure in wild populations. Landscape Ecology **32**(1): 181-194. doi:10.1007/s10980-016-0434-2.

Ogborne, A.C., Smart, R.G., Weber, T., and Birchmore-Timney, C. 2000. Who is using cannabis as a medicine and why: an exploratory study. Journal of psychoactive drugs **32**(4): 435-443.

Okazaki, H., Kobayashi, M., Momohara, A., Eguchi, S., Okamoto, T., Yanagisawa, S., Okubo, S., and Kiyonaga, J. 2011. Early Holocene coastal environment change inferred from deposits at Okinoshima archeological site, Boso Peninsula, central Japan. Quaternary International **230**(1-2): 87-94. doi:10.1016/j.quaint.2009.11.002.

Onofri, C., and Mandolino, G. 2017. Genomics and Molecular Markers in *Cannabis sativa* L. *In* Cannabis sativa L. -Botany and Biotechnology. *Edited by* S. Chandra and H. Lata and M.A. ElSohly. Springer International Publishing, Cham, Switzerland. p. 474.

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., and Mandolino, G. 2006. Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. Molecular Breeding **17**(3): 257-268. doi:10.1007/s11032-005-5681-x.

Paduch, R., Kandefer-Szerszen, M., Trytek, M., and Fiedurek, J. 2007. Terpenes: substances useful in human healthcare. Archivum Immunologiae Et Therapiae Experimentalis **55**(5): 315-327. doi:10.1007/s00005-007-0039-1.

Parliament of Western Australia. 2004. Industrial Hemp Act 2004. Available from https://www.legislation.wa.gov.au/legislation/statutes.nsf/main_mrtitle_449_homepage.html [accessed January 17 2019]

Patzak, J., Nesvadba, V., Krofta, K., Henychova, A., Marzoev, A.I., and Richards, K. 2010. Evaluation of genetic variability of wild hops (*Humulus lupulus* L.) in

Canada and the Caucasus region by chemical and molecular methods. Genome **53**(7): 545-557. doi:10.1139/g10-024.

Pavlovic, V., Pavlovic, M., Cerenak, A., Kosir, I.J., Ceh, B., Rozman, C., Turk, J., Pazek, K., Krofta, K., and Gregoric, G. 2012. Environment and weather influence on quality and market value of hops. Plant Soil and Environment **58**(4): 155-160. doi:10.17221/499/2011-pse.

PCRlabs. 2019. The Cannabis Review. Available from https://pcr-labs.com/the-cannabis-review/terpenes/ [accessed February 18 2019].

Peakall, R., and Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**(1): 288-295. doi:10.1111/j.1471-8286.2005.01155.x.

Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**(19): 2537-2539. doi:10.1093/bioinformatics/bts460.

Peil, A., Flachowsky, H., Schumann, E., and Weber, W.E. 2003. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). Theoretical and Applied Genetics **107**(1): 102-109. doi:10.1007/s00122-003-1212-5.

Pellegrini, M., Marchei, E., Pacifici, R., and Pichini, S. 2005. A rapid and simple procedure for the determination of cannabinoids in hemp food products by gas chromatography-mass spectrometry. Journal of Pharmaceutical and Biomedical Analysis **36**(5): 939-946. doi:10.1016/j.jpba.2004.07.035.

Pellerone, F.I., Edwards, K.J., and Thomas, M.R. 2001. Grapevine microsatellite repeats: Isolation, characterisation and use for genotyping of grape germplasm from Southern Italy. Vitis **40**(4): 179-186.

Pertwee, R.G. 2001. Cannabinoid receptors and pain. Progress in Neurobiology **63**(5): 569-611. doi:10.1016/s0301-0082(00)00031-9.

Pertwee, R.G. 2005. Pharmacological actions of cannabinoids. Handbook of experimental pharmacology (168): 1-51.

Pertwee, R.G. 2008. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ9- tetrahydrocannabinol, cannabidiol and Δ9- tetrahydrocannabivarin. British Journal of Pharmacology **153(2)**: 199-215.

Pertwee, R.G. 2009. Emerging strategies for exploiting cannabinoid receptor agonists as medicines. British Journal of Pharmacology **156**(3): 397-411. doi:10.1111/j.1476-5381.2008.00048.x.

Piluzza, G., Delogu, G., Cabras, A., Marceddu, S., and Bullitta, S. 2013. Differentiation between fiber and drug types of hemp (*Cannabis sativa* L.) from a collection of wild and domesticated accessions. Genetic Resources and Crop Evolution **60**(8): 2331-2342. doi:10.1007/s10722-013-0001-5.

Piomelli, D., and Russo, E.B. 2016. The *Cannabis sativa* Versus *Cannabis indica* Debate: An Interview with Ethan Russo, MD. Cannabis and cannabinoid research **1**(1): 44-46. doi:10.1089/can.2015.29003.ebr.

Pisupati, R., Vergara, D., and Kane, N.C. 2018. Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. BMC Genomics **19**. doi:10.1186/s12864-018-4494-3.

Poljuha, D., Sladonja, B., Setic, E., Milotic, A., Bandelj, D., Jakse, J., and Javornik, B. 2008. DNA fingerprinting of olive varieties in Istria (Croatia) by microsatellite markers. Scientia Horticulturae **115**(3): 223-230. doi:10.1016/j.scienta.2007.08.018.

PotGuide.com. 2018. Marijuana Strain Profiles. Retrieved from https://potguide.com/strain-profiles/ [accessed March 2019].

Potter, D.J., Clark, P., and Brown, M.B. 2008. Potency of Delta(9)-THC and other cannabinoids in *Cannabis* in England in 2005: Implications for psychoactivity and pharmacology. Journal of Forensic Sciences **53**(1): 90-94. doi:10.1111/j.1556-4029.2007.00603.x.

Prichard, R. 2014. Five questions: Why your marijuana isn’t always what budtenders say it is. *In* The Cannabist. Available from https://www.thecannabist.co/2014/07/21/marijuana-five-questions/11874/. [accessed March 31 2019]

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics **155**(2): 945-959.

ProCon. 2016a. For Which Symptoms or Conditions Might Marijuana Provide Relief? Available from https://medicalmarijuana.procon.org/view.answers.php?questionID=000087. [accessed May 21 2016].

ProCon. 2016b. States with Legislation or Ballot Measures to Legalize Medical Marijuana. Available from

https://medicalmarijuana.procon.org/view.resource.php?resourceID=002481. [accessed May 1 2016]

ProCon. 2018a. Number of Legal Medical Marijuana Patients. Available from https://medicalmarijuana.procon.org/view.resource.php?resourceID=005889. [accessed May 17 2018].

ProCon. 2018b. 33 Legal Medical Marijuana States and DC. Available from https://medicalmarijuana.procon.org/view.resource.php?resourceID=000881. [accessed November 12 2018].

R Studio Team. 2015. R Studio: integrated development for R,. RStudio, Inc., Boston, MA.

Raharjo, T.J., and Verpoorte, R. 2004. Methods for the analysis of cannabinoids in biological materials: a review. Phytochemical Analysis **15**(2): 79-94. doi:10.1002/pca.753.

Rahn, B. 2016a. Top 10 *Cannabis* Strains in California. Available from https://www.leafly.com/news/strains-products/top-10-cannabis-strains-in-california. [accessed May 2017].

Rahn, B. 2016b. Top 10 *Cannabis* Strains in Colorado. Available from https://www.leafly.com/news/strains-products/the-top-10-cannabis-strains-in-colorado. [accessed May 2017].

Rahn, B., Pearson, B.J., Trigiano, R.N., and Gray, D.J. 2016. The Derivation of Modern *Cannabis* Varieties. Critical Reviews in Plant Sciences **35**(5-6): 328-348. doi:10.1080/07352689.2016.1273626.

Rahn, E.J., and Hohmann, A.G. 2009. Cannabinoids as Pharmacotherapies for Neuropathic Pain: From the Bench to the Bedside. Neurotherapeutics **6**(4): 713-737. doi:10.1016/j.nurt.2009.08.002.

Raymond, M., and Rousset, F. 1995. GENEPOP (version-1.2) - population-genetics software for exact tests and ecumenicism. Journal of Heredity **86**(3): 248-249. doi:10.1093/oxfordjournals.jhered.a111573.

Razumova, O.V., Alexandrov, O.S., Divashuk, M.G., Sukhorada, T.I., and Karlov, G.I. 2016. Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution. Protoplasma **253**(3): 895-901. doi:10.1007/s00709-015-0851-0.

Richins, R.D., Rodriguez-Uribe, L., Lowe, K., Ferral, R., and O'Connell, M.A. 2018. Accumulation of bioactive metabolites in cultivated medical *Cannabis*. Plos One **13**(7). doi:10.1371/journal.pone.0201119.

Riggs, T.J. 1988. Breeding F1-hybrid varieties of vegetables. Journal of Horticultural Science **63**(3): 369-382. doi:10.1080/14620316.1988.11515871.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U., and Cregan, P.B. 1995. The use of microsatellite dna markers for soybean genotype identification. Theoretical and Applied Genetics **90**(1): 43-48.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8**(1): 103-106. doi:10.1111/j.1471-8286.2007.01931.x.

Russo, E.B. 2007. History of *Cannabis* and its preparations in saga, science, and sobriquet. Chemistry & Biodiversity **4**(8): 1614-1648. doi:10.1002/cbdv.200790144.

Rustichelli, C., Ferioli, V., Vezzalini, F., and Gamberini, G. 1996. Simultaneous separation and identification of hashish constituents by coupled liquid chromatography mass spectrometry (HPLC-MS). Chromatographia **43**(3-4): 129-134. doi:10.1007/bf02292940.

Rustichelli, C., Ferioli, V., Baraldi, M., Zanoli, P., and Gamberini, G. 1998. Analysis of cannabinoids in fiber hemp plant varieties (*Cannabis sativa* L.) by high-performance liquid chromatography. Chromatographia **48**(3-4): 215-222. doi:10.1007/bf02467674.

Sarri, V., Baldoni, L., Porceddu, A., Cultrera, N.G.M., Contento, A., Frediani, M., Belaj, A., Trujillo, I., and Cionini, P.G. 2006. Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. Genome **49**(12): 1606-1615. doi:10.1139/g06-126.

Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E., and Myles, S. 2015. The Genetic Structure of Marijuana and Hemp. Plos One **10**(8). doi:10.1371/journal.pone.0133292.

Schultes, R.E., Klein, W.M., Plowman, T., and Lockwood, T.E. 1974. *Cannabis*: an example of taxonomic neglect. Botanical Museum Leaflets, Harvard University **23**(9): 337-367.

Schwabe, A.L., and McGlaughlin, M.E. 2018. Genetic tools weed out misconceptions of

strain reliability in *Cannabis sativa*: Implications for a budding industry. bioRxiv. doi: https://doi.org/10.1101/332320.

Schwabe, A.L., Neale, J.R., and McGlaughlin, M.E. 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**(2): 443-457. doi:10.1007/s10592-014-0671-3.

SeedFinder. 2018. Clone Only Strains. Available from http://en.seedfinder.eu/database/strains/cloneonly/ [accessed May 31 2017].

Sharp, D.C., Townsend, M.S., Qian, Y.P., and Shellhammer, T.H. 2014. Effect of Harvest Maturity on the Chemical Composition of Cascade and Willamette Hops. Journal of the American Society of Brewing Chemists **72**(4): 231-238. doi:10.1094/asbcj-2014-1002-01.

Small, E. 1997. Cannabaceae. Flora of North America, North of Mexico. Oxford University Press, New York. pp. 381–387.

Small, E. 2015a. Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. The Botanical Review **81**(3): 189-294.

Small, E. 2015b. Response to the erroneous critique of my *Cannabis* monograph by RC Clarke and MD Merlin. The Botanical Review. pp. 306-316.

Small, E. 2016. *Cannabis*: a complete guide. CRC Press.

Small, E. 2017. Classification of *Cannabis sativa* L. in Relation to Agricultural, Biotechnological, Medical and Recreational Utilization. *In Cannabis sativa L.- Botany and Biotechnology*. Springer. pp. 1-62.

Small, E., Jui, P.Y., and Lefkovitch, L.P. 1976. A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. Systematic Botany: 67-84.

Soler, S., Gramazio, P., Figas, M.R., Vilanova, S., Rosa, E., Llosa, E.R., Borras, D., Plazas, M., and 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**: 171-178. doi:10.1016/j.indcrop.2017.04.043.

Staginnus, C., Zorntlein, S., and de Meijer, E. 2014. A PCR marker Linked to a THCA synthase Polymorphism is a Reliable Tool to Discriminate Potentially THC-Rich Plants of *Cannabis sativa* L. Journal of Forensic Sciences **59**(4): 919-926. doi:10.1111/1556-4029.12448.

Stajner, N., Rusjan, D., Korosec-Koruza, Z., and Javornik, B. 2011. Genetic Characterization of Old Slovenian Grapevine Varieties of *Vitis vinifera* L. by Microsatellite Genotyping. American Journal of Enology and Viticulture **62**(2): 250-255. doi:10.5344/ajev.2011.10011.

Swift, W., Wong, A., Li, K.M., Arnold, J.C., and McGregor, I.S. 2013. Analysis of *Cannabis* Seizures in NSW, Australia: Cannabis Potency and Cannabinoid Profile. Plos One **8**(7). doi:10.1371/journal.pone.0070052.

Sytsma, K.J., Morawetz, J., Pires, J.C., Nepokroeff, M., Conti, E., Zjhra, M., Hall, J.C., and Chase, M.W. 2002. Urticalean rosids: Circumscription, rosid ancestry, and phylogenetics based on rbcL, trnL-F, and ndhF sequences. American Journal of Botany **89**(9): 1531-1546. doi:10.3732/ajb.89.9.1531.

Techen, N., Chandra, S., Lata, H., ElSohly, M.A., and Khan, I.A. 2010. Genetic Identification of Female *Cannabis sativa* Plants at Early Developmental Stage. Planta Medica **76**(16): 1938-1939. doi:10.1055/s-0030-1249978.

The American Academy of Pain Medicine. 2019. AAPM Facts and Figures on Pain. Available from http://www.painmed.org/patientcenter/facts_on_pain.aspx [accessed January 16 2019].

The Holden Company, and Western Cultured. 2016. Terpene Flavor Wheel. Available from http://www.westerncultured.com/flavor-wheel/ [accessed January 21 2019].

Plant List. 2013. Cannabaceae. Retrieved from http://www.theplantlist.org/1.1/browse/A/Cannabaceae/ [accessed November 2018].

The United Nations. 1961. Single Convention On Narcotic Drugs 1961,. p. 25. Available from https://www.unodc.org/pdf/convention_1961_en.pdf. [accessed January 12 2019].

Tomida, I., Pertwee, R.G., and Azuara-Blanco, A. 2004. Cannabinoids and glaucoma. British journal of ophthalmology **88**(5): 708-713.

Turner, C.E., Elsohly, M.A., and Boeren, E.G. 1980. Constituents of *Cannabis sativa* L .17. A Review Of The Natural Constituents. Journal of Natural Products **43**(2): 169-234. doi:10.1021/np50008a001.

United States Congress. 1970. Comprehensive Drug Abuse Prevention and Control Act of 1970. Public Law, United States of America. Available from https://www.govinfo.gov/content/pkg/STATUTE-84/pdf/STATUTE-84-Pg1236.pdf.

United States Department of Agriculture. 1970. United States Plant Variety Protection Act. Available from https://www.ams.usda.gov/sites/default/files/media/Plant%20Variety%20Protection%20Act.pdf.

United States Department of Agriculture. 2014. Specialty Crop Competitiveness. *Edited by* D.o. Agriculture. Available from https://www.ams.usda.gov/services/grants/scbgp/specialty-crop

United States Department of Agriculture. 2015. Application for Plant Variety Protection Certificate. USDA. Available from https://www.ams.usda.gov/services/plant-variety-protection/pvpo-application-requirements.

United States Department of Agriculture, and 113th United States Congress. 2018. Agricultural Act of 2014. *In* 2014 U.S. Farm Bill, Washington. Available from https://www.congress.gov/bill/113th-congress/house-bill/2642.

United States Department of Health and Human Services, Food and Drug Administration, Office of Medical Products and Tobacco, Center for Drug Evaluation and Research, Office of Generic Drugs, and Office of Generic Drug Policy. 2018. Approved Drug Products with Therapeutic Equivalence Evaluations, 38th Edition.

United States Food and Drug Administration. 2018. FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy. June 25 2018. Accessed September 13 2018. Available from https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm611046.htm

University of Mississippi. 2017. Marijuana Research, History. Available from http://pharmacy.olemiss.edu/marijuana/history/ [accessed January 2019].

Vaccani, A., Massi, P., Colombo, A., Rubino, T., and Parolaro, D. 2005. Cannabidiol inhibits human glioma cell migration through a cannabinoid receptor-independent mechanism. British Journal of Pharmacology **144**(8): 1032-1036. doi:10.1038/sj.bjp.0706134.

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., and Page, J.E. 2011. The draft genome and transcriptome of *Cannabis sativa*. Genome Biology **12**(10). doi:10.1186/gb-2011-12-10-r102.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**(3): 535-538. doi:10.1111/j.1471-8286.2004.00684.x.

Vergara, D., Baker, H., Clancy, K., Keepers, K.G., Mendieta, J.P., Pauli, C.S., Tittes, S.B., White, K.H., and Kane, N.C. 2016. Genetic and Genomic Tools for *Cannabis sativa*. Critical Reviews in Plant Sciences **35**(5-6): 364-377. doi:10.1080/07352689.2016.1267496.

Vergara, D., Bidwell, L.C., Gaudino, R., Torres, A., Du, G., Ruthenburg, T.C., deCesare, K., Land, D.P., Hutchison, K.E., and Kane, N.C. 2017. Compromised External Validity: Federally Produced *Cannabis* Does Not Reflect Legal Markets. Scientific Reports **7**. doi:10.1038/srep46528.

Verity, R., and Nichols, R.A. 2016. Estimating the Number of Subpopulations (K) in Structured Populations. Genetics **203**(4): 1827-+. doi:10.1534/genetics.115.180992.

Weedmaps. 2018. Weedmaps. Retrieved from https://weedmaps.com/ [accessed September 2018].

Welling, M.T., Liu, L., Shapter, T., Raymond, C.A., and King, G.J. 2016. Characterisation of cannabinoid composition in a diverse *Cannabis sativa* L. germplasm collection. Euphytica **208**(3): 463-475. doi:10.1007/s10681-015-1585-y.

White, K.H., Vergara, D., Keepers, K.G., and Kane, N.C. 2016. The complete mitochondrial genome for *Cannabis sativa*. Mitochondrial DNA Part B-Resources **1**: 715-716. doi:10.1080/23802359.2016.1155083.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer, Verlag, New York

Wikileaf. 2018. Cannabis Strains: Strain Library. Retrieved from https://www.wikileaf.com/strains/ [accessed January 2019].

Wodak, A., and Owens, R. 1996. Drug Prohibition: A Call for Change. UNSW Press Ltd ISBN 978-0-86840-175-1., Sydney, New South Wales, Australia.

World Health Organization. 2018. Management of substance abuse, *Cannabis*. Available from http://www.who.int/substance_abuse/facts/cannabis/en/ [accessed November 15, 2018].

Xiaoyan, W.D.M., Searfoss, J., and Telepchak, M. 2016. Determination of cannabinoid content and pesticide residues in *Cannabis* edible beverages. LCGC. Available from http://www.chromatographyonline.com/determination-cannabinoid-content-and-pesticide-residues-cannabis-edibles-and-beverages.

Yang, M.-Q., van Velzen, R., Bakker, F.T., Sattarian, A., Li, D.-Z., and Yip, T.-S. 2013. Molecular phylogenetics and character evolution of Cannabaceae. Taxon **62**(3): 473-485. doi:10.12705/623.9.

Yates, T. 2015. David Griffiths and the missionary 'History of Madagascar'. Journal of Ecclesiastical History **66**(4): 904-906. doi:10.1017/s0022046915000895.

Zimmerman, D.W., and Zumbo, B.D. 1993. Relative Power of the Wilcoxon Test, the Friedman Test, and Repeated-Measures Anova on Ranks. Journal of Experimental Education **62**(1): 75-86. doi:10.1080/00220973.1993.9943832.

Zygmunt, P.M., Andersson, D.A., and Hogestatt, E.D. 2002. Delta(9)-tetrahydrocannabinol and cannabinol activate capsaicin-sensitive sensory nerves via a CB1 and CB2 cannabinoid receptor-independent mechanism. Journal of Neuroscience **22**(11): 4720-4727.