Cannabis and Cannabinoid Research
Volume 5, Number 3, 2020
Mary Ann Liebert, Inc.
DOI: 10.1089/can.2018.0063

Cannabis Chemovar Nomenclature Misrepresents Chemical and Genetic Diversity; Survey of Variations in Chemical Profiles and Genetic Markers in Nevada Medical Cannabis Samples

Ulrich Reimann-Philipp,[1] Mark Speck,[2] Cindy Orser,[3] Steve Johnson,[3] Aaron Hilyard,[3] Helen Turner,[2] Alexander J. Stokes,[4] and Andrea L. Small-Howard[1,*]

Abstract

Introduction: Medical cannabis patients receive clinical benefits from the secondary metabolites of the plant, which contain a variety of cannabinoids and terpenoids in combinations that can be used to classify the chemovars. State-regulated medical cannabis programs rely on breeder-reported "strain" names both within diversion control systems and to describe the medical cannabis products that are sold to patients in medical cannabis dispensaries. In state-regulated medical cannabis programs, there is no conventional nomenclature system that correlates the breeder-reported names with their profiles of active ingredients, and these "strain" names are invalid as they refer to chemical differences properly referred to as to chemovars.

Materials and Methods: To determine the actual levels of chemical diversity represented in 2662 samples of *Cannabis* flower collected between January 2016 and June of 2017 in Nevada, chemical profile data were measured from these samples by a state-qualified third-party testing laboratory. Principal component analysis (PCA) was used to define clusters in data sets representing both cannabinoids and terpenoids, cannabinoids only, or terpenoids only.

Results: The PCA of the terpenoid only data set revealed three well-defined clusters. All three terpenoids only data clusters had high tetrahydrocannabinolic acid synthase, but the terpene profiles listed in reverse-order of abundance best defined these chemovars. The three chemovars in Nevada were labeled with 396 breeder-reported sample names, which overestimate the diversity and do not inform patients regarding chemical properties. Representative DNA samples were taken from each chemovar to determine whether the genetic diversity was greater than the chemical diversity. The limited genotyping experiment was based on DNA sequence polymorphisms. The genetic analysis revealed twelve distinct genetic clades, which still does not account for the entirety of the 396 reported sample names. The finite genotypes did not correlate with the chemotypes determined for the samples. This suggests that either the DNA-markers used were too narrowly restricted for factual separation or that environmental factors contributed more significantly to the chemical profiles of cannabis than genetics.

Conclusion: The three chemovars and twelve genotypes reflect low medical diversity on the market in Nevada during its "medical use only" phase. Furthermore, the 396 breeder-reported sample names within this set imply a false sense of diversity of products in Nevada dispensaries.

Keywords: cannabis; chemical profile; chemovars; genetic diversity; medical cannabis; nomenclature

[1]GB Sciences, Inc., Las Vegas, Nevada.
[2]Department of Biology, Chaminade University of Honolulu, Honolulu, Hawai'i.
[3]Digipath Labs, Las Vegas, Nevada.
[4]Department of Cell and Molecular Biology, John A. Burns School of Medicine, University of Hawai'i, Honolulu, Hawai'i.

*Address correspondence to: Andrea L. Small-Howard, PhD, GB Sciences, Inc., 3550 West Teco Avenue, Las Vegas, NV 89118, E-mail: andrea@gbsciences.com

© Ulrich Reimann-Phillip et al. 2019; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons Attribution Noncommercial License (http://creativecommons.org/licenses/by-nc/4.0/) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are cited.

Introduction

Legal medical marijuana usage is increasing in the United States due to the expansion of state-regulated medical cannabis programs and the broadening scope of existing state medical marijuana programs to include "adult use.[1]" Medical cannabis patients benefit from the secondary metabolites of the plant, which include a variety of cannabinoids and terpenoids in combinations that can be used to classify the varieties of *Cannabis*.[2,3] The major stakeholders in these state-run programs (i.e., cultivators, extractors, processors, retailers, testing labs, and state-regulating agencies) characterize cannabis products and regulate cannabis materials based on the breeder-reported name. Despite this, none of the stakeholders in medical cannabis programs is required to prove the chemovar-identity of the products.

The genus *Cannabis* contains a single species, *Cannabis sativa*.[4] Most of the cannabis available on the U.S. market are hybrid varieties defined by properties such as psychoactive and medicinal effects, appearance, yield, taste, and odor.[5] These varieties are called "strains" in the common vernacular, but should properly be referred to as chemovars as they do not meet the scientific definition of a strain such as that used for bacteria or viruses.[6,7] The vast majority of commercially cultivated *Cannabis* plants are produced through cloning.[5] Clonal propagation ensures that plants are genetically identical to the mother plant.[8] In contrast to varieties propagated through seeds, requiring a lengthy process of backcrossing and inbreeding to achieve consistency, new chemovars can be created much faster by clonal propagation. New chemovars are constantly generated and enter the market, resulting in thousands of different breeder-reported names without any scientific naming convention.

Most state-regulated medical marijuana programs rely on the breeder-reported names for tracking cannabis products, and these are also the primary cannabis product designation. Naming conventions for cannabis chemovars are poorly or not at all defined, and there are currently thousands of chemovars available in states that have legalized marijuana usage. This creates a confusing situation for patients, who depend on the identifications and potency data on the packaging. If the commercial names bear little consistent relationship to either genotype or chemical phenotype, then they should not be the primary basis provided to patients for decision-making.[3,9]

Some cannabis researchers have suggested that chemovars should be identified based on their chemical profiles, as informed by principal component analysis (PCA),[2,3,10] while others advocate for a genetically based labeling system.[11–14] Currently, genetically identical chemovars may be labeled in dispensaries with different names, or different genotypes might be sold under the same name in state licensed dispensaries. Furthermore, genetically identical samples that are grown under different environmental conditions may produce different profiles of active ingredients.[15]

Medical marijuana patients are entitled to consistent composition and potency of cannabinoids and terpenoids.[16] Some states require disclosure of complex chemical profile information; whereas others only require disclosure of tetrahydrocannabinol levels and cannabidiol levels.[1] The potential for mislabeling of chemovars, inconsistent chemical profiles of marijuana products, and often limited testing data make it difficult or impossible for many patients to obtain a consistent chemical profile of the product. The relationships between the breeder-reported names, genotype, and chemotype have not been rigorously evaluated.

In this study, we investigate variation among chemovars based on the variability of chemical profiles and the relationship of these chemical profiles to three genetic markers. We focused on the state of Nevada, which started to grant licenses for commercial medical marijuana in November 2014. Medical marijuana products sold in Nevada are tested by a third-party laboratory for cannabinoid and terpene content and for microbial and pesticide safety, thus, testing data are available for analysis. Chemical profiles of *Cannabis* flower samples collected between January 2016 and June 2017 were generated by a major Nevada testing laboratory. During the period of data collection, Nevada was a medical-only cannabis state, and so, all tested samples were destined for medically oriented use. Data cleaning consisted of correcting for misspellings and eliminating different names that were apparently used for the same samples and not counting samples only identified by numbers. PCA data for all samples were produced using algorithms to reveal probable clusters. We then compared the average content of individual cannabinoids and terpenes within the resultant clusters. Finally, representative samples were selected for DNA sequence analysis to reveal the genetic diversity within the cannabis samples. DNA sequence data from fragments of three genes involved in the biosynthesis of cannabinoids and terpenes were the basis of the genotypic analysis. The variation in these sequences was then used to establish the genetic relationship of

the samples and the relationship of the genetic data to the chemical profiles.

Materials and Methods

Analysis of cannabinoid and terpenoid chemical profiles

Cannabinoid assays were run on an Agilent Technologies 1260 UPLC system (Santa Clara, CA) equipped with a G4212A DAD, G1316C temperature-controlled column compartment, G4226A autosampler, and G4204A quaternary pump. Cannabinoids were separated on a Poroshell 120 EC-C18 column with a Poroshell 120 EC-C18 guard column. The UPLC runs OpenLab CDS ChemStation Rev C. 01.06 (61) software (Agilent Technologies). Terpenoid analyses were carried out on an Agilent 7890B GC/7697A Headspace/5977A mass spectrophotometer with a DB-624UI and Agilent 5181–8818 split/splitless liner. Injector port temperature was 250°C with a transfer line, valve oven, and needle temperature of 180°C. Carrier gas was helium at a flow of 33.0 cm/sec. The mass spectrometer detector was set to scan with a range from 50 to 300 m/z. The instrument was controlled by Agilent MassHunter Quantitative Analysis (Vers. B.08.00 Build 8.0.593.0). Certified reference standards were from Restek (Bellefonte, PA) with MassHunter library confirmation.

Data and statistical analysis of chemical profiles for multivariate analyses

The analyte values from the cannabis samples were normalized before performing multivariate analysis.[17] Scaling was performed using the scale function, which is part of the base R statistical software, using the *root-mean-square* method. Cluster estimation was performed using the NbClust function provided by the NbClust package for the R software. NbClust utilizes 25 different cluster estimation indices to generate a majority rules number of clusters for the given variable set. Clustering was done using the k-means hierarchical clustering method available in the R Stats package. PCA was used as a statistical tool for exploratory data analysis to infer predictive models.[17] PCAs were conducted in R version 3.0.2, and then visualized with the package ggplot2. Mean terpenoid and cannabinoid concentrations were displayed in bar charts to visualize the individual analyte profiles.

Supplemental clustering

Starting with subsetting and scaling the different analyte components, clustering estimation and k-means clustering were performed on (1) all cannabinoids, (2) all terpenes, and (3) all tested components (combined cannabinoids and terpenes). Grouped box plot of analytes was rendered to compare the differences in these variables among the clusters.

DNA sequencing

DNA sequence determination was performed using Illumina® next-generation targeted sequencing technology. The sequencing was performed by Medicinal Genomics, which used a *Reduced Representation Shotgun* approach.[18] Targeted sequence analysis was carried out on 250–350 base-pair fragments generated by *Nsp*I restriction endonuclease digestion. Short read and single-nucleotide polymorphism sequencing results were available as standard FASTQ and VCF formatted files, respectively. For the purpose of this work, the FASTQ were used for functional gene fragment alignments. Consensus sequences were assembled using pileups from the FASTQ files that ranged between 50 and 500 copies depending on the chemovar. This provided sufficient depth of coverage for constructing contiguous sequences.

Selection of target genes and DNA sequence alignment

Researchers have successfully used marker sequences from the major cannabinoid synthetic pathways in the genetic classification of *Cannabis* chemovars.[12] The biosynthetic pathways involved in the production of *Cannabis* bioactive compounds have been described.[19] We used the Geneious v6.1.8 software to identify conserved regions in three genes encoding key enzymes involved in the biosynthesis of cannabinoids and terpenoids: isopentenyl-diphosphate delta isomerase (IDI), geranyl diphosphate synthase small subunit 2 (GPP), and tetrahydrocannabinolic acid synthase (THC). The reference gene sequences, THC (AB212829.1), IDI (KY014569.1), and GPP (GPPS KY014583.1), were downloaded into *Geneious* from the National Center for Biotechnology Information's (NCBI) GenBank database. The three reference genes were used to align the sequence fragments from the FASTQ files. Since full sequences of all three genes were not available in the sequence FASTQ files, conserved consensus fragments of these genes were used to build a concatenated gene sequence for each cultivar of about 722 base pairs each. The concatenated genes were aligned using the ClustalW multiple sequence alignment algorithm.

Construction of phylogenetic trees

The MrBayes v3.2.1 algorithm[20] was used for Bayesian inference of the cultivar relationships because it proved to produce the most rigorous and likely results. Markov chain Monte Carlo method used a 1,000,100 chain length, sampling posterior trees every 200 generations. The first 10% were discarded as burn-in during tree and parameter summarization. The majority-rule consensus tree contains node values that represent the probability that particular clades form in the posterior tree distributions.

Results

Variations in cannabinoid and terpenoid content

Testing results of 2662 dried flower samples were analyzed. The content of 11 cannabinoids and 19 terpenoids was recorded as % (w/w) of dried flower. The samples were collected by the third-party laboratory at cultivation sites in the Las Vegas area. All samples were labeled with the breeder-reported name of the *Cannabis* for submission to the testing laboratory. These names appeared on the official test reports, the state registry, and on the final product labels. There were 637 different names associated with the samples. We corrected for obvious misspellings and likely inconsistent labeling of the same strain (e.g., Blueberry and Blueberry Kush) and eliminated samples that were only labeled with numbers. This left 396 different names. The goal of the PCA was to detect likely clusters of chemical profiles (chemovars) within the sample population to answer the question: How many different chemovars were available to medical cannabis patients in Nevada? Subsets of data were statistically analyzed to identify the most complete and valid number of clusters. Figure 1 shows the distribution of the chemovars for (1) cannabinoids only (Fig. 1A, B); (2) terpenoids only (Fig. 1C, D); and (3) cannabinoids and terpenoids combined (Fig. 1D, E). The numbers of clusters for the different data sets and the numbers of samples within the clusters are shown in Figure 1A–E. The number of predicted clusters was small: two clusters for cannabinoids and terpenoids combined, three clusters for cannabinoids only, and three clusters for terpenoids alone. The components and abundances within the chemical profiles among the samples were mostly not significantly different from each other. Most samples in the cannabinoid group were contained in one cluster (93%). There was more variability when the analysis was done on the terpenoid group alone. It resulted in three clusters, representing 59%, 33%, and 8% of all samples.

Chemical compositions in the chemovars

We examined the distribution of abundances of each cannabinoid/terpene component in each sample. We used Tukey plots to display the range of abundances (as % of dry mass) for each component. Figure 2A shows the relative content of the 11 cannabinoids in the 3 cannabinoid clusters. Figure 2B shows the relative content of the terpenes in the three clusters. Figure 3A–AD displays the range of abundancies for each compound by cluster.

THC in the acidic form is the overall most abundant cannabinoid in all three groups, followed by decarboxylated THC (Cluster 1), cannabigerolic acid (CBGA) (Cluster 2), and cannabdiolic acid (CBDA; Cluster 3). All other cannabinoids are only present in very small amounts. The 47 samples in Cluster 3 contained more than 1% CBDA, ranging from 3.8% to 16.7%. A few samples also contained higher amounts of individual "minor" cannabinoids. While most of the samples came from high-THC chemovars, the terpenoid profiles were more diverse. The terpenoid profiles can best be described by the order of abundance of the individual terpenes.

DNA sequence analysis

As shown above, three clusters with distinct terpenoid profiles were predicted by the PCA: Cluster 1 (1582 samples): β-myrcene, limonene, α-pinene, β-caryophyllene, and α-humulene; Cluster 2 (200 samples): γ-terpinene, terpinolene, β-myrcene, α-humulene, and β-pinene; and Cluster 3 (880 samples): limonene, β-myrcene, β-caryophyllene, nerolidol, and limonene. The data showed that samples could be better separated into clusters based on their terpenoid content than on their cannabinoid content. To study DNA sequence polymorphisms and their possible link to the chemical profiles, we selected the samples based on their terpenoid profiles.

DNA sequence polymorphism, or the number of nucleotide changes within a certain fragment of the genome, can be used to establish phylogenetic trees and identify and distinguish varieties within one species. Exchanges of single or multiple nucleotides within defined "marker" fragments are also used to facilitate breeding of new varieties. Combinations of markers that can be linked to specific traits are a way of identifying proprietary genetic resources in legal disputes.

We explored the suitability of three gene fragments to be used as markers. The fragments were contained in three genes that encode enzymes involved in the biosynthesis of terpenoids and cannabinoids. These genes

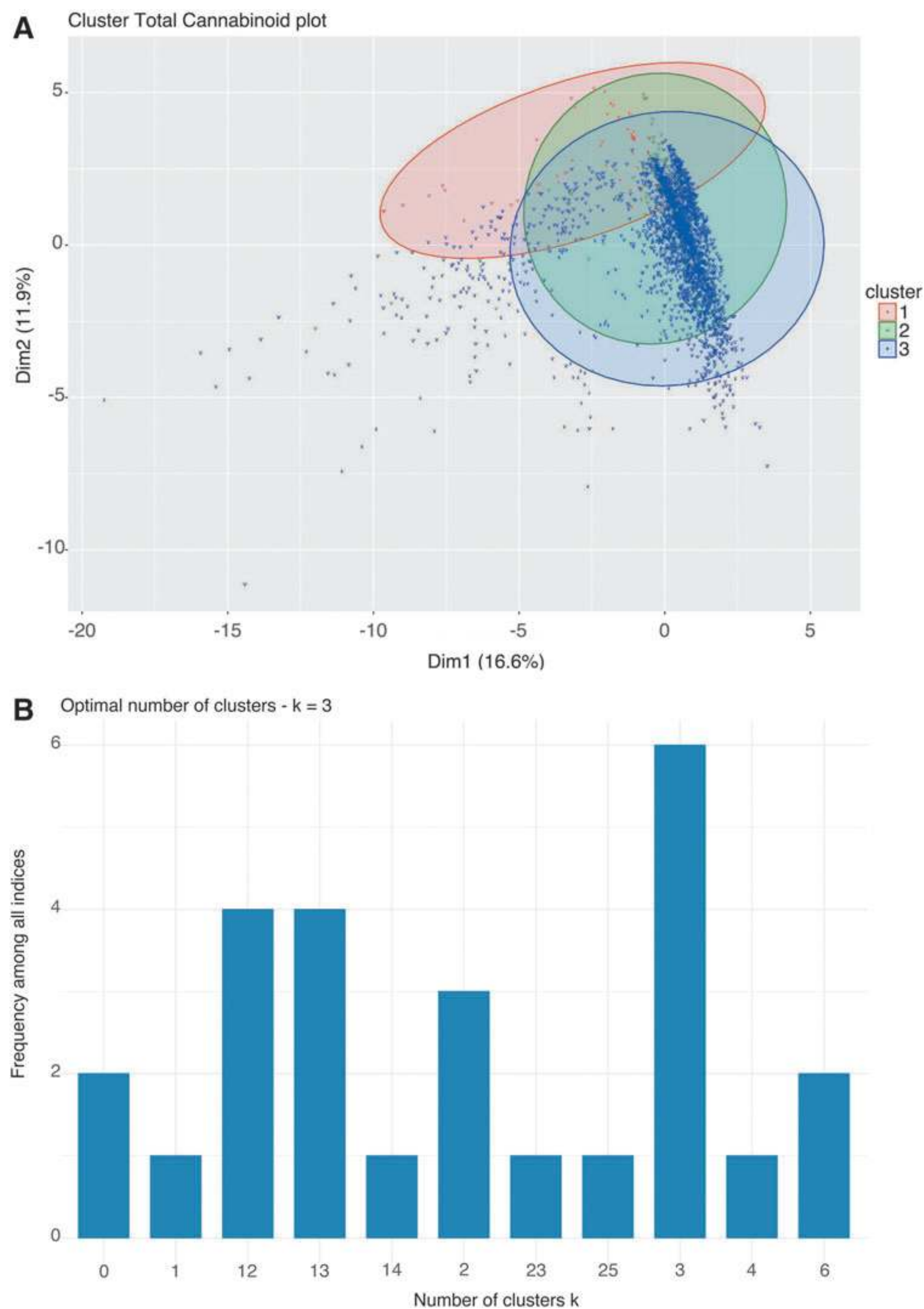

FIG. 1. Principal component analysis of *Cannabis* flower test results. Cluster analysis was performed on cannabinoid only, terpenoid only, and combined cannabinoid and terpenoid data sets. Upper panels show multivariate analysis of groupings. The groupings predicted by the highest number of algorithms are shown for cannabinoid only **(A, B)**, terpenoid only **(C, D)**, and combined cannabinoid and terpenoid data **(E, F)**. Most likely numbers of clusters were predicted using an array of 26 algorithms and are presented as histogram analyses (lower panels).

(*Figure continued*→)

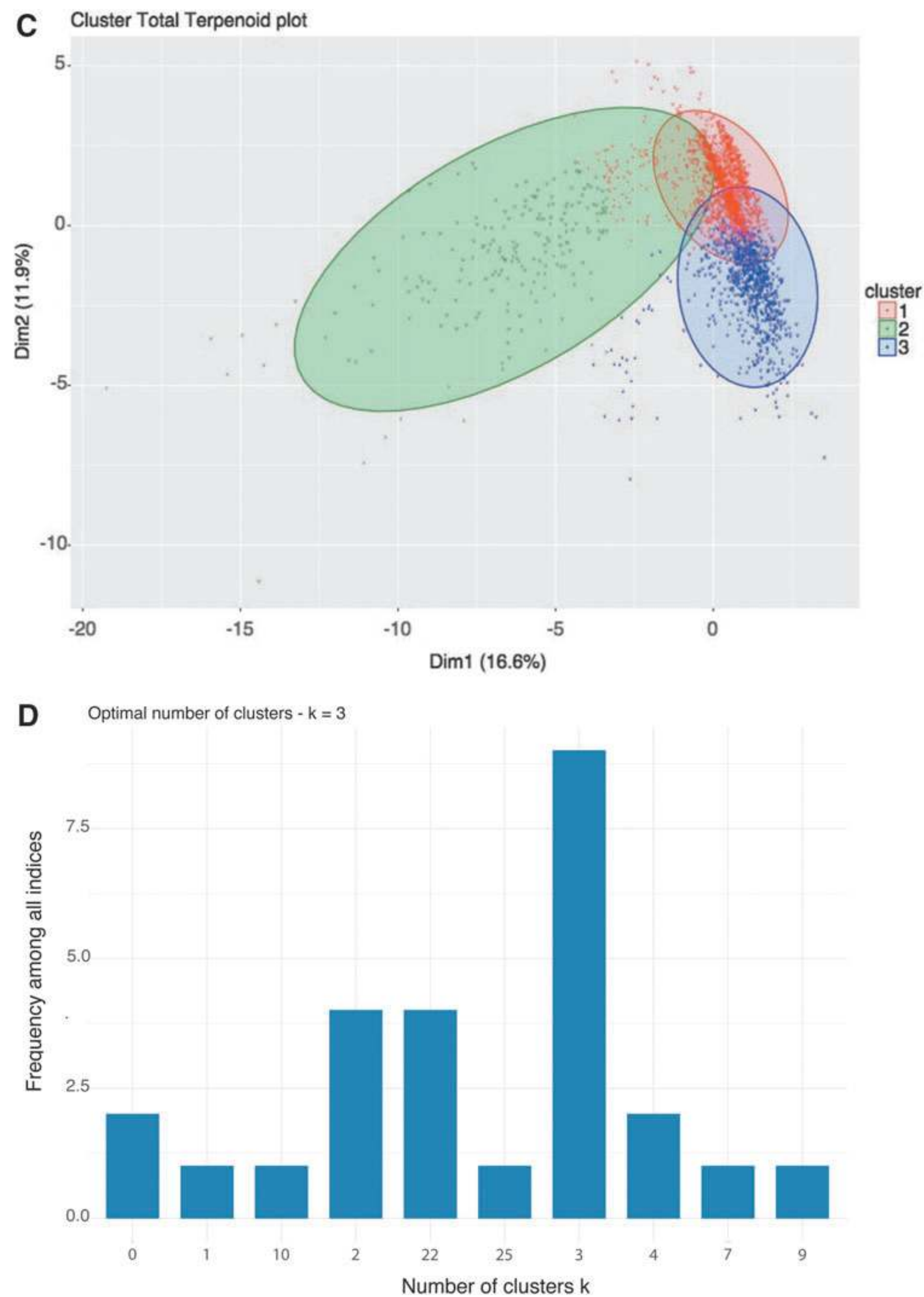

FIG. 1. Continued.

encode key enzymes involved in the biosynthesis of cannabinoids and terpenoids: *IDI*, *GPP*, and *THC(s)*. The reference gene sequences, *THC(s)* (AB212829.1), *IDI* (KY014569.1), and *GPP* (GPPS KY014583.1), were downloaded into *Geneious* from NCBI. Polymorphisms in these sequences are more likely linked to the chemical profiles represented in the different chemovar clusters. The 72 samples that were selected for sequencing represented the three different terpenoid patterns typical for each cluster.

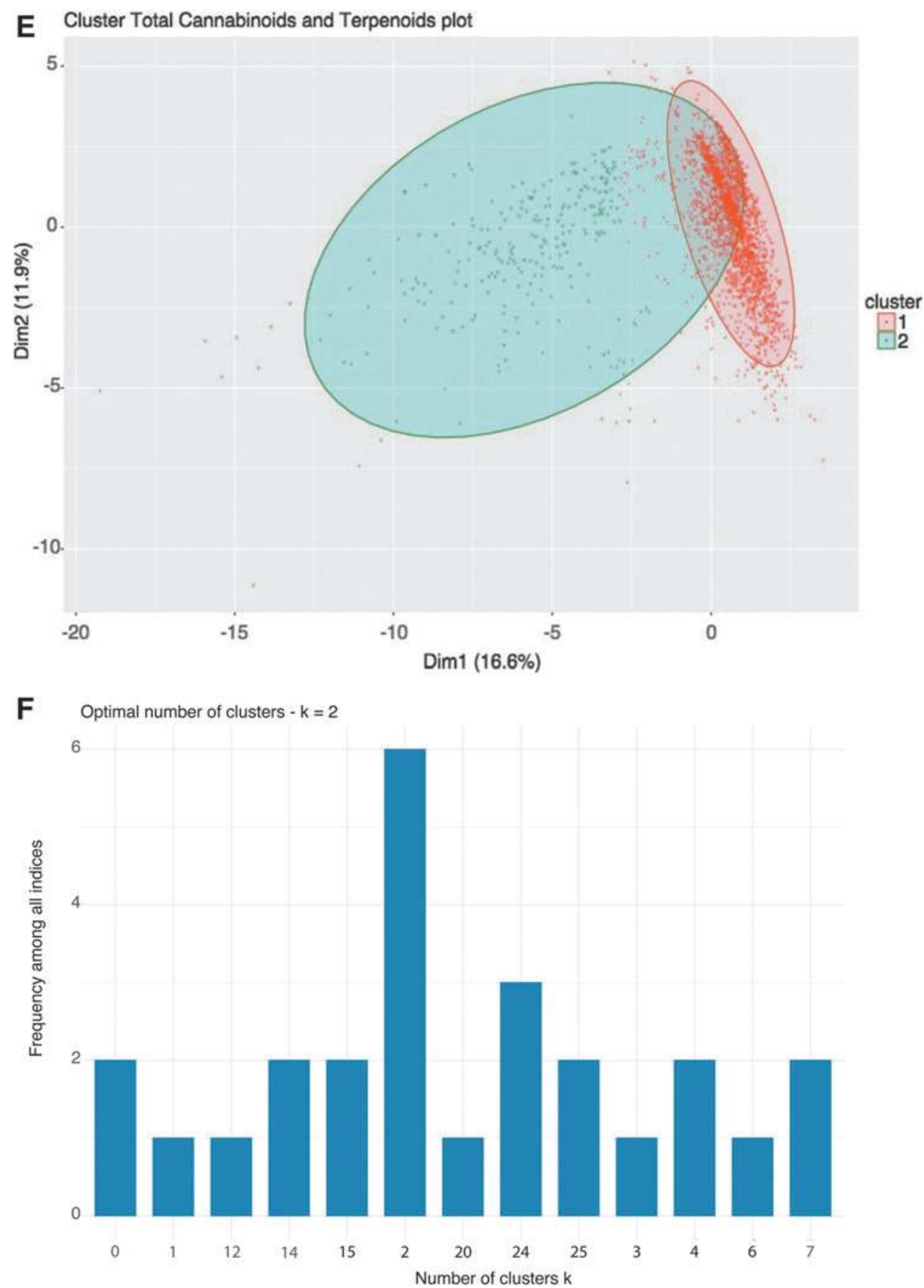

FIG. 1. Continued.

Figure 4 shows three projections of the phylogenetic tree that was constructed based on the combined polymorphisms of the three gene fragments. The other samples connected to the consensus sequence with a straight line also are different at one location, but each location is different from the other samples. The two samples KimboSlice-RSP10949 and AgentOrangeRSP10605 share the same single nucleotide change from the theoretical "consensus" sequence but have an additional change

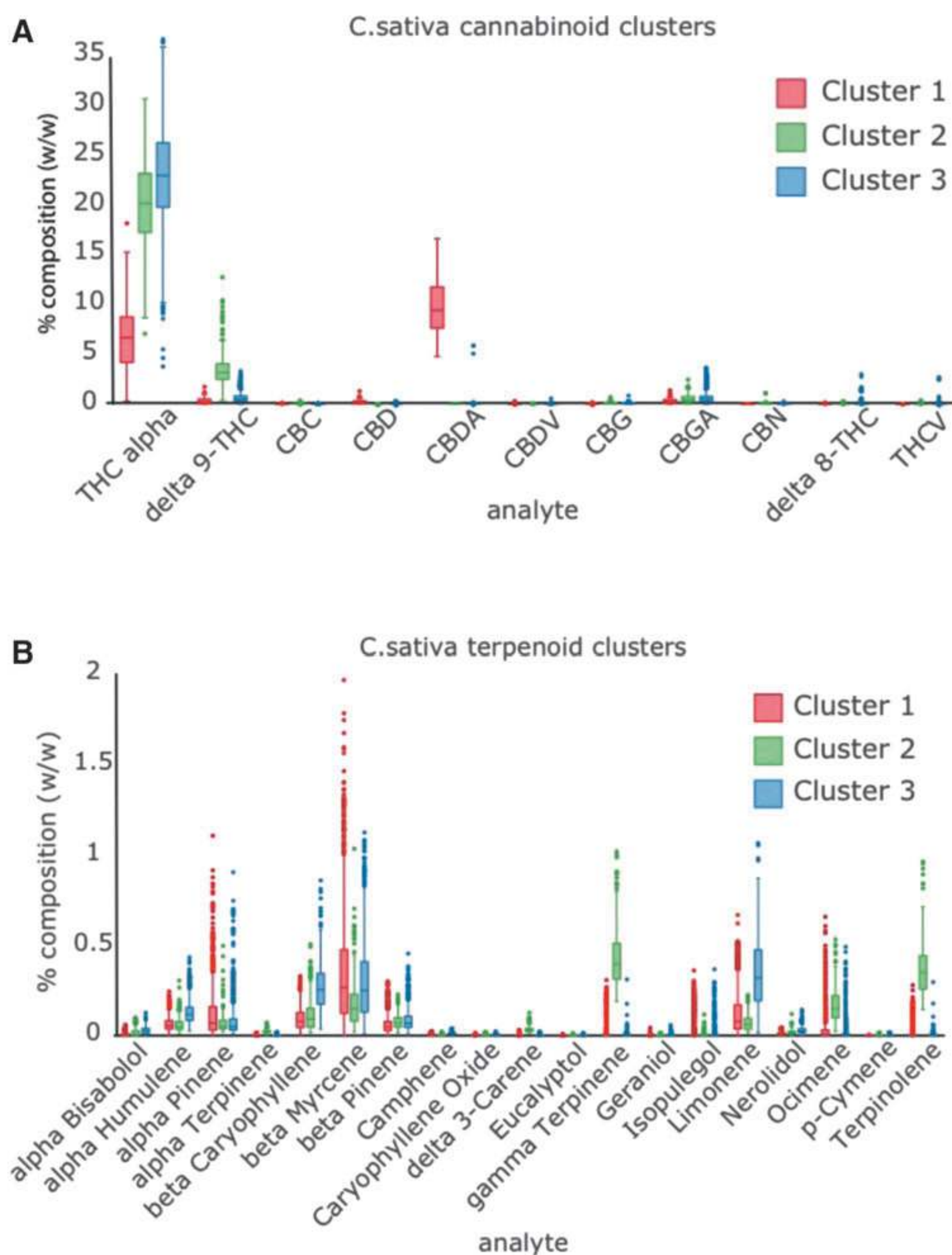

FIG. 2. Distribution of content values of cannabinoids and terpenoids in predicted clusters of *Cannabis* samples. The box plots display the range and distribution of each analyte in terms of %w/w. The line bisecting each box represents the median for that distribution. The lower and upper lines show the minimum and maximum values of the lower and upper quartiles, respectfully. The points show the outliers for the sampled ranges. The y-axis values represent % (w/w) of dried flower for each of the indicated cannabinoid species. **(A)** Cannabinoid data including THCA, which are dominant in terms of abundance. **(B)** Terpenoid data grouped by cluster assignments. THCA, tetrahydrocannabinolic acid; delta-9THC, delta 9 tetrahydrocannabinolic acid; CBC, cannabichromene; CBD, cannabdiol; CBDA, cannabdiolic acid; CBDV, cannabivarin; CBG, cannabigerol; CBN, cannabinol; delta 8 THC, delta 8 tetrahydrocannabinolic acid; THCV, tetrahydrocannabivarin.

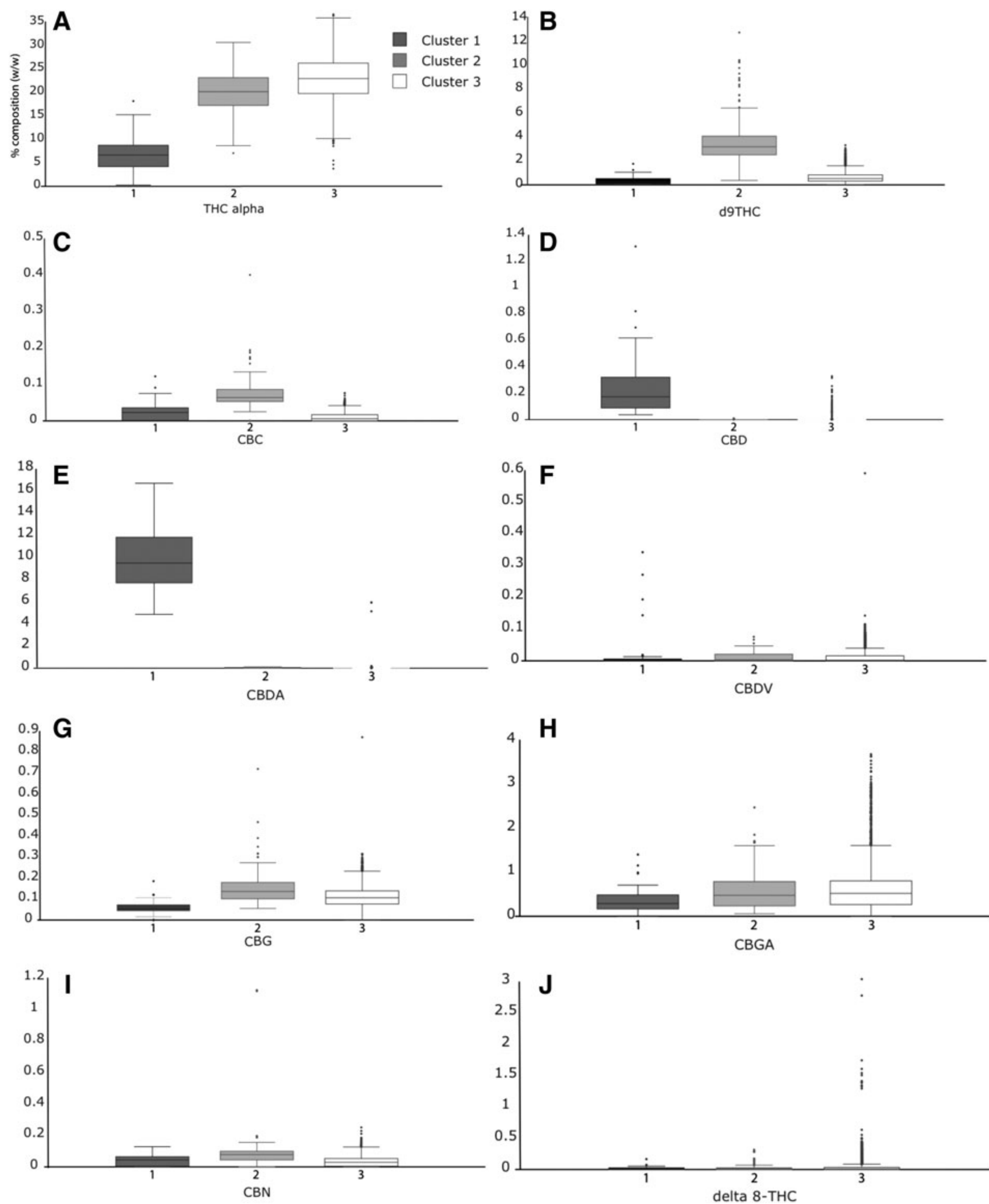

FIG. 3. Individual box plots of each of the analyzed cannabinoids and terpenes divided by cluster assignments. The y-axis values represent % (w/w) of dried flower for each of the indicated cannabinoids **(A–K)** and terpene **(L–AD)** species.

(*Figure continued*→)

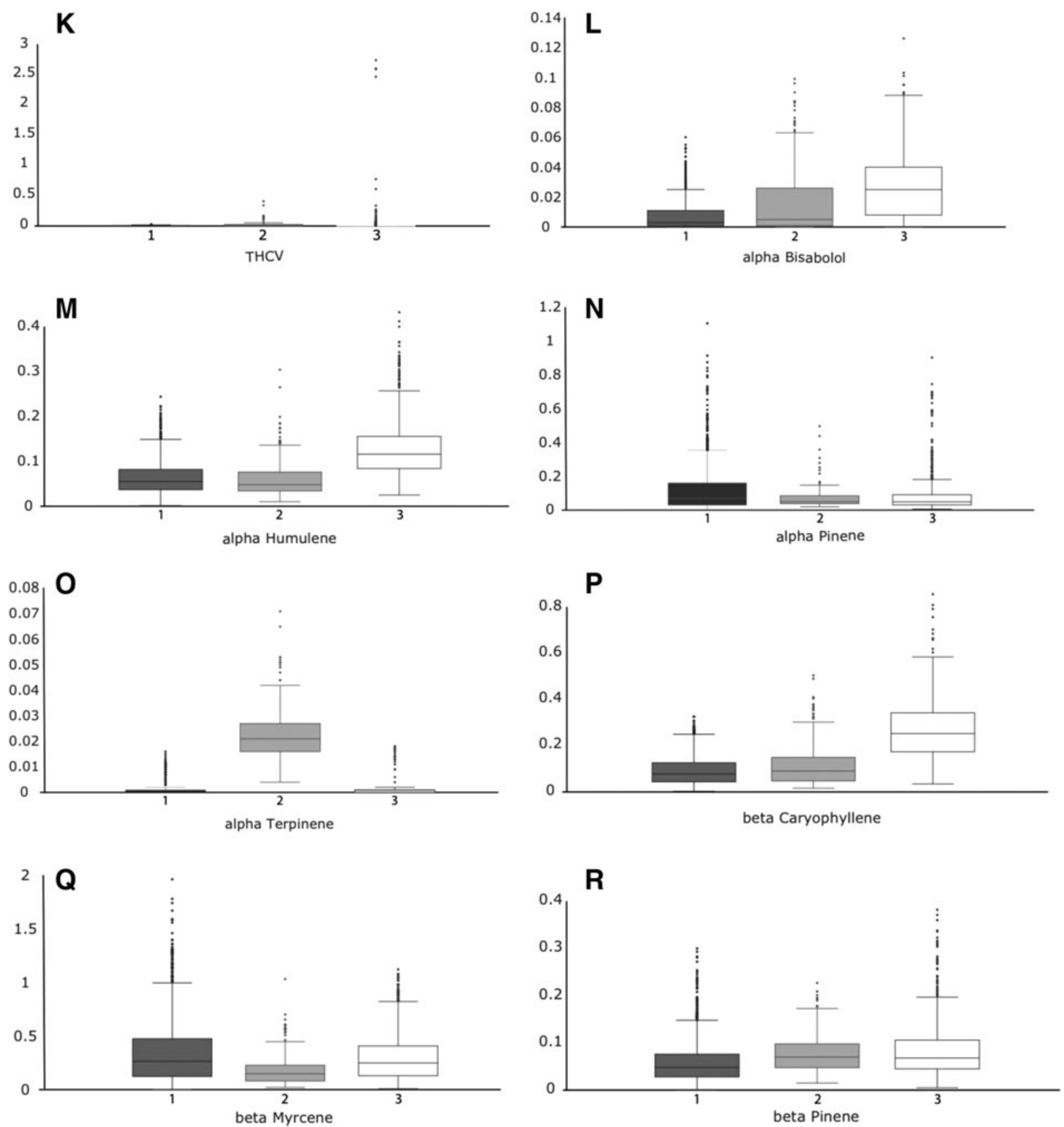

FIG. 3. Continued.

that separates them from each other. Thus, both are two changes removed from the "consensus" sequence; one change is shared, and the other is different for each sample.

The sequence data showed that the samples are at least in one and at most in four locations different from the "consensus" sequence. A large group of 39 (54% of sequenced samples) is removed by one change from the consensus sequence. All three terpenoid clusters are represented in this group. The terpenoid clusters appear to be randomly distributed throughout the tree without a pattern suggesting linkage between the sequence polymorphisms and the chemovar clusters (Fig. 4A). Figure 4B categorizes the chemovars by

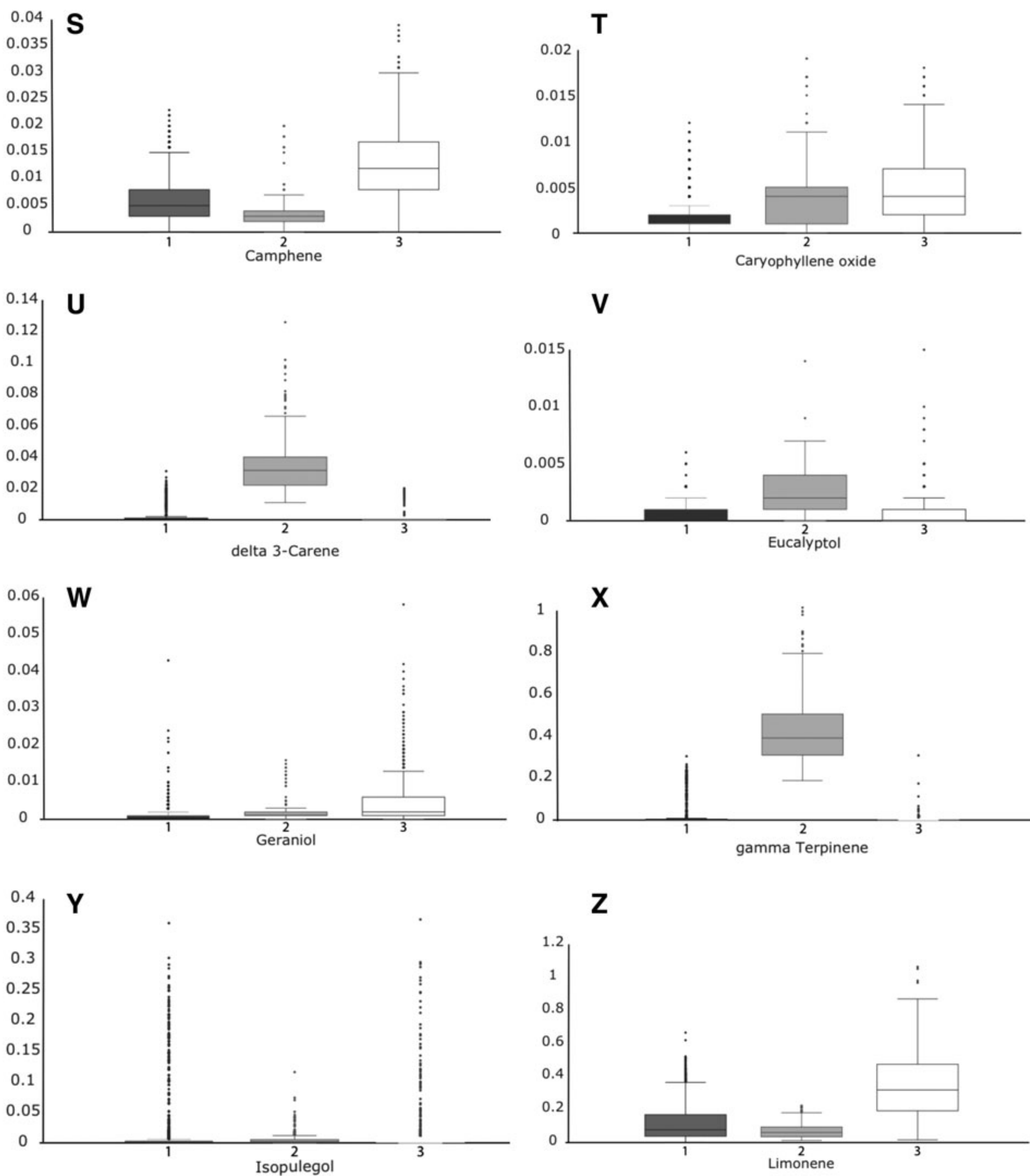

FIG. 3. Continued.

their belonging to 1 of 12 clades that were assigned during the analysis. Finally, the probabilities of these clades being formed as determined by Bayesian inference are printed as percentages on the base branches in Figure 4C.

Discussion

When analyzing the chemical testing data using a variety of statistical methods, we found that there was surprisingly little variability in the chemical profiles among the 2662 dried flower samples. When only the

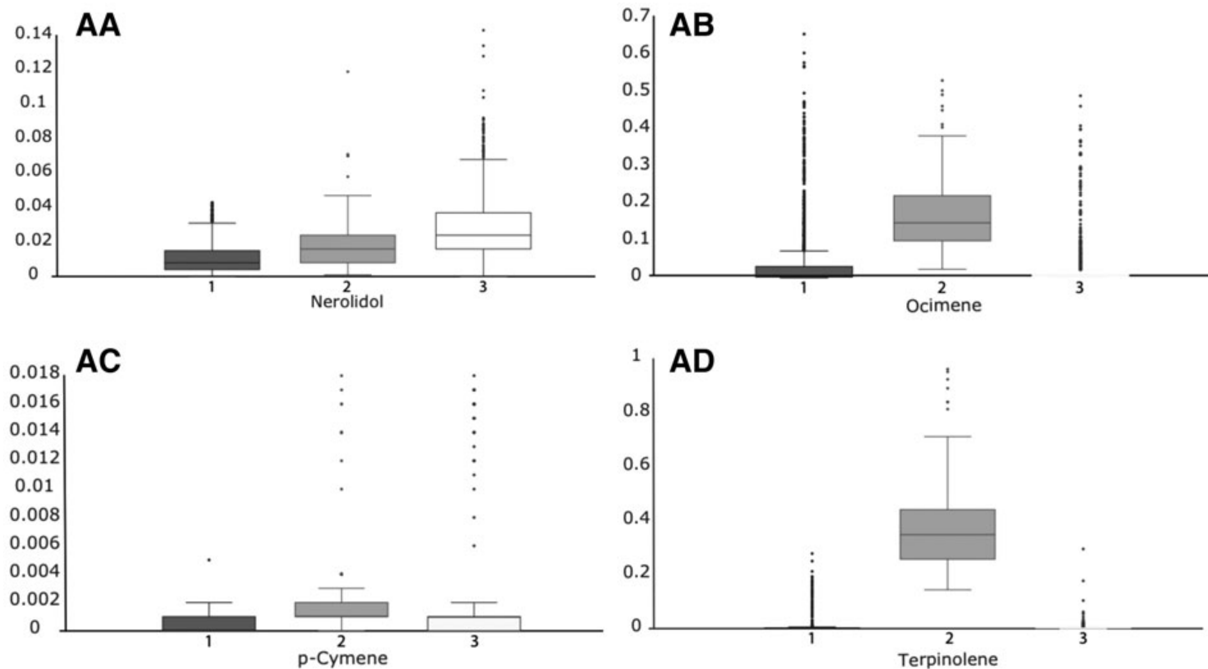

FIG. 3. Continued.

cannabinoid composition was considered, except for the few high-CBD samples, all samples contained a very high amount of THC (>22% on average) and very small amounts of other cannabinoids.

The terpenoid profiles were more variable, revealing three clusters that could be distinguished by individual terpenes, with β-myrcene, γ-terpinene, and limonene being the most abundant terpene in each cluster. Thus, we assume that the terpenoid composition in a sample is much more indicative of the sample's origin and genetic background than the cannabinoid profile. The link between the terpenoids and the three clusters makes sense because the terpenes provide the colors and smells that differentiate the many varieties of cannabis. Other researchers using genetics and chemical profiles of more divergent cannabis samples also ended up with three clusters[3,21]; namely European hemp, narrow leaflet drug-types, and broad leaflet drug-types.

While 93% of samples were within one cluster (which emphasizes the limited variation in the samples) there are still considerable number of samples that fall outside each cluster, with some being rather distinct. In general, the Cluster 3 chemotypes had relatively higher amounts of tetrahydrocannabinolic acid (THCA) and CBGA compared with the all samples *en totum.* The outliers, defined here as the points outside the ellipse drawn around the cluster centroid, in this cluster had higher than the median delta-9THC, CBGA, or CBG pushing them toward the bottom. Most of the outlying samples also tended to have detectable amounts residues from all 11 tested cannabinoids. The few Cluster 2 outliers were similar in that they had detectable amounts of each cannabinoid residue tested. Cluster 2 was defined by lower THCA values on average compared to Cluster 3, but higher relative delta-9THC values. Interestingly, several of the Cluster 3 outliers were from the Durbin Poison chemovar, and two of the four Cluster 2 outliers were from the Super Lemon Haze chemovar. As these are chemotypes, the cluster results are not mapping the genetic expression of all the relevant pathways for these compounds. Further work is required to explore the connection or lack of connection to the genetic identity of these chemovars to their potential chemical outputs.

The results of the Nevada-based analyses suggest that when creating new chemovars, plants were predominantly selected for high-THC content and to a much lesser degree for CBD content, across different terpene backgrounds. All the other cannabinoids seem to have been mostly ignored in the breeding efforts, even though

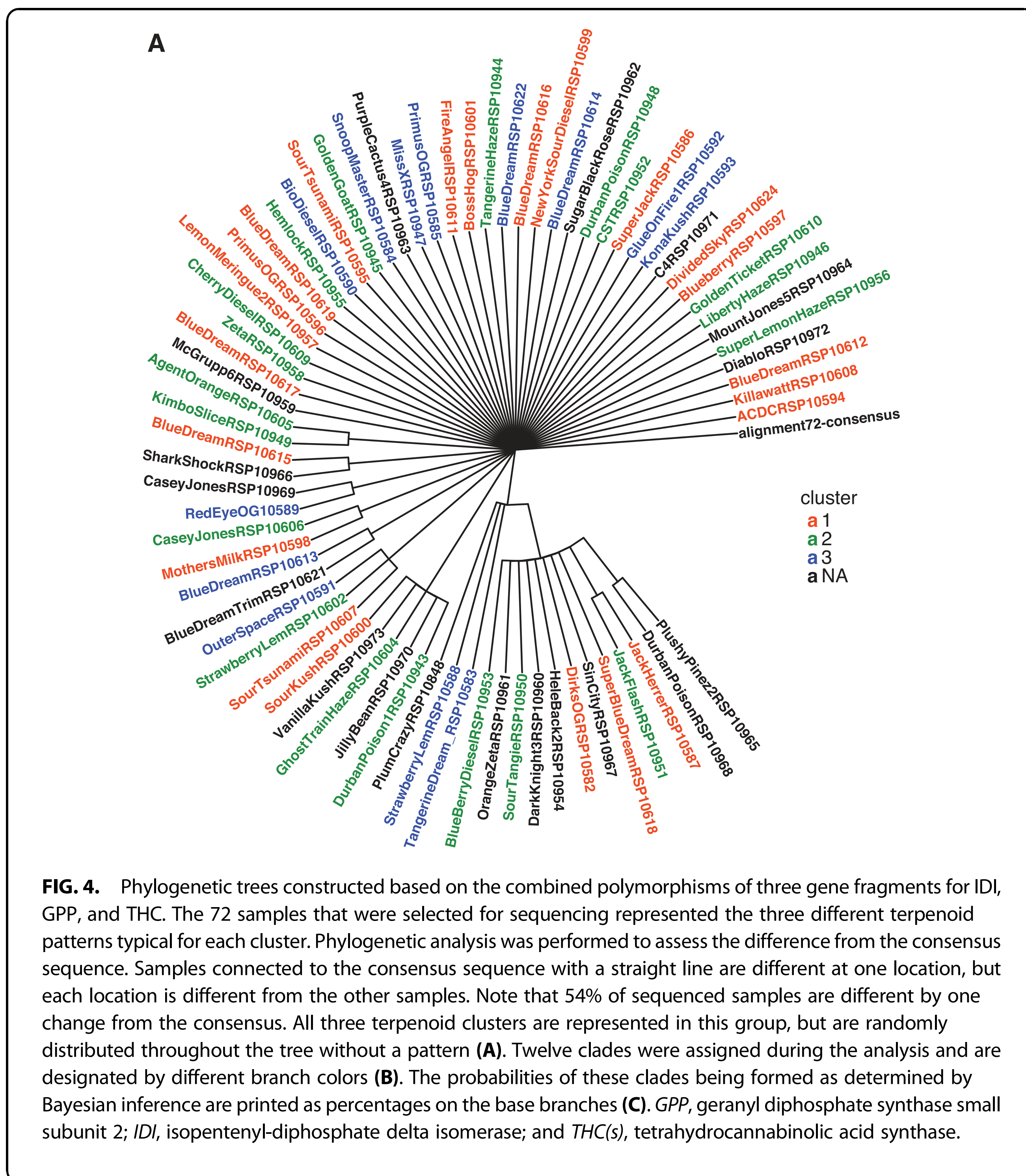

FIG. 4. Phylogenetic trees constructed based on the combined polymorphisms of three gene fragments for IDI, GPP, and THC. The 72 samples that were selected for sequencing represented the three different terpenoid patterns typical for each cluster. Phylogenetic analysis was performed to assess the difference from the consensus sequence. Samples connected to the consensus sequence with a straight line are different at one location, but each location is different from the other samples. Note that 54% of sequenced samples are different by one change from the consensus. All three terpenoid clusters are represented in this group, but are randomly distributed throughout the tree without a pattern **(A)**. Twelve clades were assigned during the analysis and are designated by different branch colors **(B)**. The probabilities of these clades being formed as determined by Bayesian inference are printed as percentages on the base branches **(C)**. *GPP*, geranyl diphosphate synthase small subunit 2; *IDI*, isopentenyl-diphosphate delta isomerase; and *THC(s)*, tetrahydrocannabinolic acid synthase.

(*Figure continued*→)

they have known or suspected specific medical values.[22,23] It remains to be seen how these findings translate to other locales, since, in this study, we are by definition examining a highly selective population of cannabis that represents a small sample of the total cannabis germplasm potential.

The very limited variability in chemical profiles shows that most of these chemovars, while named differently, are almost the same or at least very similar. There is no standard for naming chemovars, and new names can be created or used randomly. Products

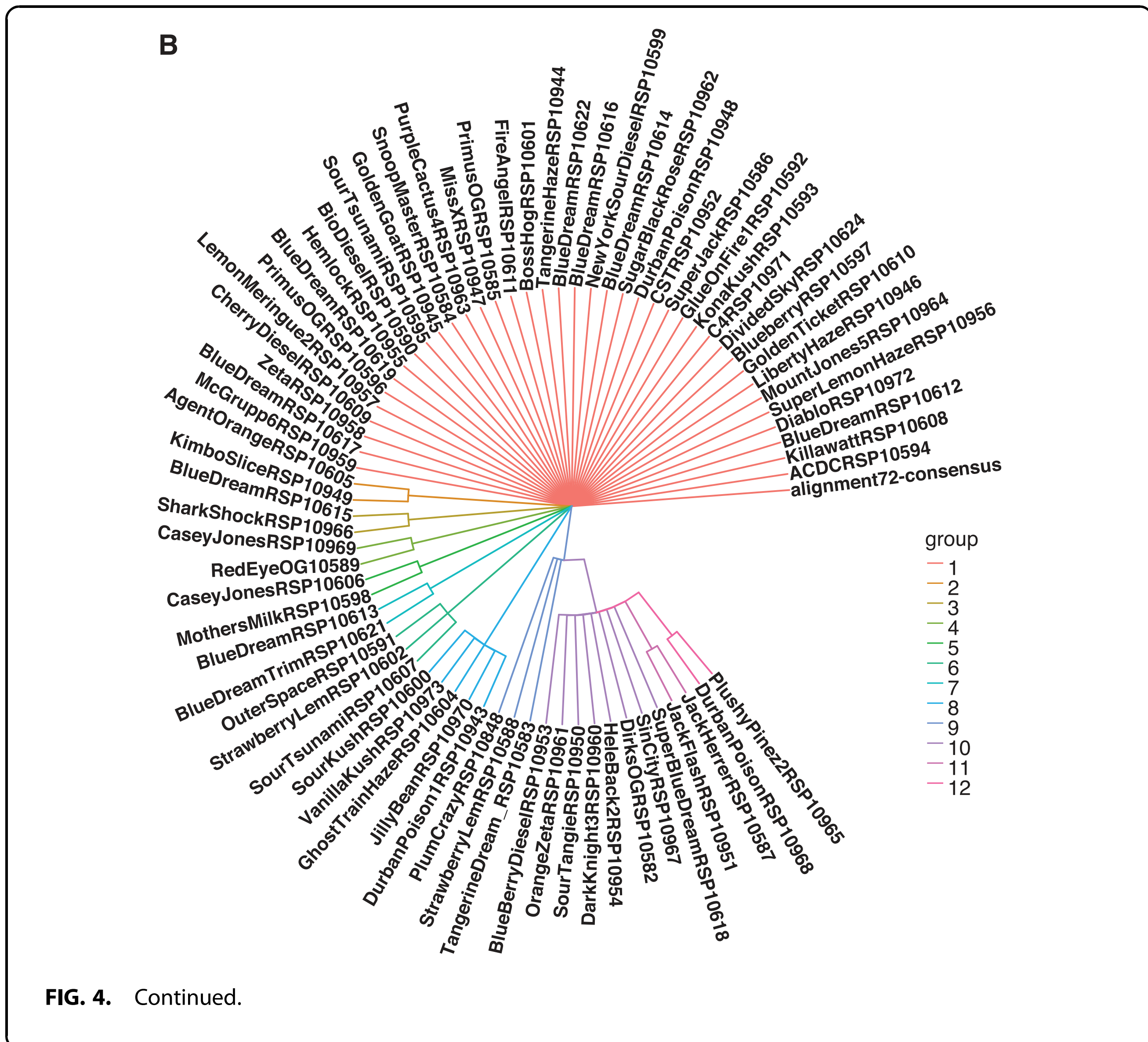

FIG. 4. Continued.

labeled with the same chemovar name can differ significantly in their chemical profiles depending on the origin of the mother plant, where they were grown, and how they were processed. There is no protection for patients from inadvertent or intentional mislabeling. Doctors could recommend products with a higher degree of confidence if products were tied to the chemical signature of active ingredients. Alternatively, lack of patient and physician education on the medical benefits of the minor cannabinoids and terpenes may also contribute to the dominance of THC in this medical market.

The DNA analysis did not show a correlation of the twelve genetic clades, as defined by certain nucleotide change patterns within the terpene clusters revealed by PCA. One interpretation of three terpene clusters distributed randomly across samples is that environmental factors contribute significantly to the chemical profiles of cannabis; perhaps more so than the genetic origin. Alternatively, improved DNA-markers should be developed to assist with the genetic classification of *Cannabis* spp., which may improve the alignment of the genotypes with the chemotypes.

These results also illustrate the complexity of using DNA marker fragments in chemovar identification programs. *Cannabis* varieties—chemovars—are highly conserved at the genetic level and attempts to use

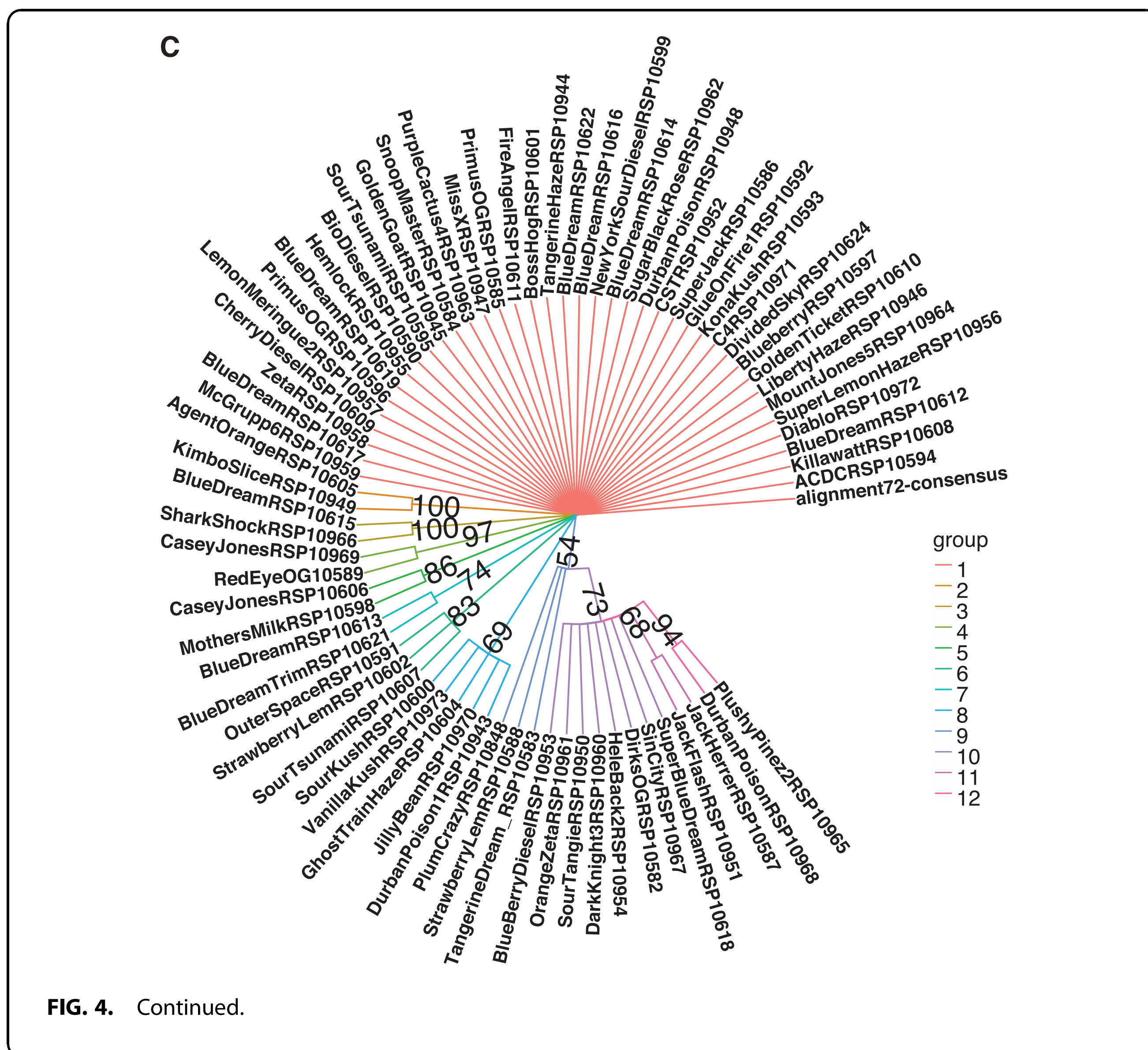

FIG. 4. Continued.

standard molecular clock approaches to phylogeny have not worked. Using DNA markers to differentiate among related samples usually requires 10–15 markers that are carefully chosen and validated to identify a variety with certainty in any target crop.[21,24] It will require an extensive effort involving multiple crosses to establish such markers. However, once these markers are available, breeding of varieties with new properties, such as high content of "minor" cannabinoids and specific terpene profiles, can be accelerated.

Conclusions

There was little variability in the chemical profiles of 2662 samples labeled with 396 chemovar names. High-THC chemovars that lacked significant amounts of other cannabinoids were most abundant. For medical cannabis patients, diversity in chemical profiles is very desirable. Unfortunately, patients had only three chemovars to choose from in the Nevada medical cannabis market from January 2016 to June 2017.

Breeder-reported names have very little value in identifying the medical value of cannabis, yet, stakeholders in state-regulated cannabis programs rely on these names. Many different names are assigned to basically the same products. Because of a lack of validated information, medical patients and caregivers may only understand the information for the THC potency. Patients who would benefit from specialty products,

such as significant levels of "minor" cannabinoids, have not been served during the sampling period.

Acknowledgments

We thank Dr. Adam Famoso (Louisiana State University) for critical review of the article. This research was funded by GB Sciences, Las Vegas, Nevada. Editorial support was provided by The Curry Rockefeller Group.

Author Disclosure Statement

U.R.-P. and A.L.S.-H. are employees of GB Sciences. M.S., C.O., S.J., A.H., and H.T. received research funding from GB Sciences.

References

1. Americans for Safe Access. Medical marijuana access in the United States: a patient-focused analysis of the patchwork of state laws. Americans for Safe Access: Washington, DC, 2017.
2. Fischedick JT, Hazekamp A, Erkelens T, et al. Metabolic fingerprinting of *Cannabis sativa* L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. Phytochemistry. 2010;71:2058–2073.
3. Orser C, Johnson S, Speck M, et al. Terpenoid chemoprofiles distinguish drug-type *Cannabis sativa* L. cultivars in Nevada. Nat Prod Chem Res. 2018;6:304.
4. Interagency Taxonomic Information System. Ulmaceae and Cannabaceae of North America Update, database (version 2011). Updated for ITIS by the Flora of North America Expertise Network, in connection with an update for USDA PLANTS (2007–2010); Reston, VA.
5. Farag S, Kayser O. Chapter 1: the cannabis plant: botanical aspects. In Preedy VR (ed.), Handbook of cannabis and related pathologies: biology, pharmacology, diagnosis, and treatment. New York: Elsevier, 2017.
6. Russo EB, Marcu J. Cannabis pharmacology: the usual suspects and a few promising leads. Adv Pharmacol. 2017;80:67–134.
7. Lewis MA, Russo EB, Smith KM. Pharmacological foundations of cannabis chemovars. Planta Med. 2018;84:225–233.
8. Chandra S, Lata H, ElSohly MA. et al. Cannabis cultivation: methodological issues for obtaining medical-grade product. Epilepsy Behav. 2017;70(Pt B):302–312.
9. Orser C. Empowering the cannabis consumer in a rapidly expanding global market. Cannabis Sci Technol. 2018;1:50–55.
10. Hazekamp A, Fischedick JT. Cannabis—from cultivar to chemovar. Drug Test Anal. 2012;4:660–667.
11. Welling MT, Liu L, Shapter T, et al. Characterisation of cannabinoid composition in a diverse *Cannabis sativa* L. *germplasm* collection. Euphytica. 2016;208:463–475.
12. Kojoma M, Seki H, Yoshida S, et al. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in "drug-type" and "fiber-type" *Cannabis sativa* L. Forensic Sci Int. 2006;159:132–140.
13. Pacifico D, Miselli F, Micheler M, et al. Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. Mol Breed. 2006;17: 257–268.
14. Gao C, Xin P, Cheng C, et al. Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. PLoS One. 2014;9:e110638.
15. de Meijer EP, Bagatta M, Carboni A, et al. The inheritance of chemical phenotype in *Cannabis sativa* L. Genetics. 2003;163:335–346.
16. Potter DJ. A review of the cultivation and processing of cannabis (*Cannabis sativa* L.) for production of prescription medicines in the UK. Drug Test Anal. 2014;6:31–38.
17. Eriksson L, Andersson PL, Johansson E, et al. Megavariate analysis of environmental QSAR data. Part I—a basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). Mol Divers. 2006;10:169–186.
18. Altshuler D, Pollara VJ, Cowles CR, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;407:513–516.
19. van Bakel H, Stout JM, Cote AG, et al. The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. 2011;12:R102.
20. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17:754–755.
21. Lynch RC, Vergara D, Tittes S. et al. Genomic and chemical diversity in Cannabis. Crit Rev Plant Sci. 2016;35:349–363.
22. Russo EB. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br J Pharmacol. 2011; 163:1344–1364.
23. Russo EB, McPartland JM. Cannabis is more than simply delta(9)-tetrahydrocannabinol. Psychopharmacology (Berl). 2003;165:431–432; author reply 433–434.
24. Sawler J, Stout JM, Gardner KM, et al. The genetic structure of marijuana and hemp. PLoS One. 2015;10:e0133292.

Cite this article as: Reimann-Philipp U, Speck M, Orser C, Johnson S, Hilyard A, Turner H, Stokes AJ, Small-Howard AL (2019) Cannabis chemovar nomenclature misrepresents chemical and genetic diversity; survey of variations in chemical profiles and genetic markers in Nevada medical cannabis samples, *Cannabis and Cannabinoid Research* 5:3, 215–230, DOI: 10.1089/can.2018.0063.

Abbreviations Used

CBC = cannabichromene
CBD = cannabdiol
CBDA = cannabdiolic acid
CBDV = cannabivarin
CBG = cannabigerol
CBGA = cannabigerolic acid
CBN = cannabinol
delta 8 THC = delta 8 tetrahydrocannabinolic acid
delta 9-THC = delta 9 tetrahydrocannabinolic acid
GPP = geranyl diphosphate synthase small subunit 2
IDI = isopentenyl-diphosphate delta isomerase
NCBI = National Center for Biotechnology Information's
PCA = principal component analysis
THCA = tetrahydrocannabinolic acid
THC = tetrahydrocannabinol
THCV = tetrahydrocannabivarin