Comprehensive chromatographic profiling of cannabis from 23 USA States marketed for medical purposes

RAMIA Z. AL BAKAIN\(^1\)*, YAHYA S. AL-DEGS\(^2\), JAMES V. CIZDZIEL\(^3\) and MAHMOUD A. ELSOHLY\(^4,5\)

\(^1\) Department of Chemistry, School of Science, The University of Jordan, 11942, Amman, Jordan
\(^2\) Chemistry Department, The Hashemite University, 150459, Zarqa, Jordan
\(^3\) Department of Chemistry and Biochemistry, University of Mississippi, University, MS 38677, USA
\(^4\) National Center for Natural Products Research, University, MS 38677-1848, USA
\(^5\) Department of Pharmaceutics and Drug Delivery, School of Pharmacy, University of Mississippi University, MS 38677, USA

Received: January 24, 2020 \(\overset{\circ}{\text{●}}\) Accepted: February 29, 2020
Published online: June 5, 2020

ABSTRACT

In this research, cannabis varieties represent 23 USA States were assayed by GC-FID to generate their complex chemical profiles informative for plants clustering. Results showed that 45 cannabinoids and terpenoids were quantified in all plant samples, where 8 cannabinoids and 18 terpenoids were identified. Among organics, \(\Delta^9\)-THC, CBN (cannabinoids) and Fenchol (terpenoid) not only showed the highest levels overall contents, but also were the most important compounds for cannabis clustering. Among States, Washington, Oregon, California and Hawaii have the highest cannabis content. GC-FID data were subjected to PCA and HCA to find (1) the variations among cannabis chemical profiles as a result of growing environment, (2) to reveal the compounds that were responsible for grouping cultivars between clusters and (3) finally, to facilitate the future profile prediction and States clustering of unknown cannabis based on the chemical profile. The 23 cannabis USA States were grouped into three clusters based on only \(\Delta^9\)-THC, CBN, C1 and Fenchol content. Cannabis classification based on GC-profile will meet the practical needs of cannabis applications in clinical research, industrial production, patients’ self-production, and contribute to the standardization of commercially-available cannabis cultivars in USA.

KEYWORDS
cannabinoids, terpenoids, chemical profiles, GC-FID, USA states, PCA, HCA Clustering, \(\Delta^9\)-THC

INTRODUCTION

The plant cannabis (\textit{Cannabis sativa L.}) is the most widely consumed and popular medicinal botanical drug product in the world due to its high usage and its diverse pharmacological properties [1]. Chemically, cannabis is a complex species containing large number of active constituents [1–3]. Herbal cannabis (known as marijuana), cannabis resin (hashish), and extracts of cannabis resin (hashish oil) are still the most illicit drugs in the world. About 8,000 tons of cannabis are intake in USA per year [1]. In many countries, cannabis is popular including Canada and North America [1, 2, 4]. In 2017, many USA states have legalized the medical use of cannabis, where, 38 licensed producers in Canada are authorized to produce and sell dried marijuana [4]. There has been a major increase of domestic production worldwide not only in USA and Canada, but also in Colombia, Mexico, Jamaica, and Thailand [1].
Cannabinoids and terpenoids are the active ingredients in cannabis [4, 5]. Both classes of compounds are known of their variables biological activities [6]. Terpenoids are of great interest because of their production by the plants reflecting the immediate environment and they are responsible for cannabis’ distinctive odor [7], whereas, cannabinoids tend to reveal genetic relationships [8].

Today, most nations worldwide regard cannabis as an illegal drug of abuse. Despite the abuse potential of cannabis and its illegal status at the federal level in the USA, research into its chemistry and pharmacology has demonstrated that it also has medicinal properties. Cannabis has a long history of human use as a medicinal plant, intoxicant, and ritual drug [9, 10]. Clinical trials into cannabis, pure cannabinoids, and synthetic analogs have demonstrated some effectiveness as analgesics for chronic neuropathic pain, appetite stimulants for anorexia associated with cancer and AIDS patients, multiple sclerosis, pain, inflammation, depression, anticancer, palliative, epilepsy and infection [11–16]. The increased medical interest in these substances has prompted the development of various cannabis based medicines such as the oral Δ2-THC (delta-9-tetrahydrocannabinol) preparation Marinol®, a synthetic analog of Δ8-THC (delta-8-tetrahydrocannabinol) and an oral mucosal spray containing 1:1 ratio of Δ8-THC and CBD (cannabidiol) [17, 18].

There are three classification systems for cannabis. The first, is by species based on physical appearance, THC (tetrahydrocannabinol) content, and geographical origins since environmental factors and marijuana cultivated sources can induce different cannabis profiles [1, 8, 19–22]. The second classification is based on the ratio of two major cannabinoids THC and CBD which is decided by their corresponding allelic loci [23, 24]. The third is based on both cannabinoids and terpenoids for drug standardization and clinical research purposes [24]. Novotny et al. reported that data relative to the use of GC analysis of marijuana samples of different origin indicated that the chromatograms appeared to be different, so correlation between chromatographic data and geographical origin of marijuana might be possible [25]. Hazekamp et al. [26] reported the impact of changing the environmental conditions on the chemical composition and variability of terpenoids and cannabinoids in different cannabis varieties.

A wide variety of analytical techniques have been used for chemical profiling (i.e., fingerprinting) of cannabis. Thin Layer Chromatography [22], fingerprinting with HPLC [26–28], GC coupled with mass spectrometry [1] and 1HNMR have been used to fingerprint cannabis aqueous extracts and tinctures [29] as well as to chemically differentiate cannabis cultivars [30]. SFC also has been used to analyze cannabis [13, 31–34]. However, GC is the most commonly used instrument for analyzing cannabinoids and terpenoids [1, 13, 19, 20, 35]. GC has been used to differentiate cannabis from different countries, including Mexico, Colombia, Jamaica, Thailand, and the USA [1].

The current approaches for cannabis classification may be inadequate because they analyze cannabis from botanical perspectives based on only two cannabinoids; THC and CBD. Moreover, there is currently no available comprehensive chemical profiling for all USA states medical-type cannabis samples which is necessary to explore the similarities/differences if any among plants samples of different States. Therefore, this study was carried out.

In this study, a comprehensive work was carried out to identify the compounds most important in distinguishing cannabis varieties and to find the variation on cannabis chemical profiles as a result of growing plants in different environments and in growth time from 23 USA States that have enacted Medical Marijuana laws, including: Alaska, Arizona, California, Colorado, Delaware, Florida, Hawaii, Illinois, Maine, Maryland, Massachusetts, Michigan, Montana, Nevada, New York, Ohio, Oregon, Pennsylvania, Vermont, Washington, West Virginia, Wisconsin and Mississippi. GC-FID was applied for the chemical analysis. Cannabis plants samples obtained from each of the 23 USA States were collected, extracted and analyzed using GC-FID. The plant samples were analyzed to detect all possible cannabinoids and terpenoids from different cannabis seeds and origins which are necessary for cannabis fingerprinting. The method was validated and evaluated for selectivity and precision (i.e., repeatability). The advanced multivariate tools including Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are efficient towards sample clustering [36–41]. Hence, these tools were performed in this study to: (1) identify the compounds most important in distinguishing cannabis varieties, (2) find the variation on cannabis chemical profiles as a result of growing plants in different States and with different in growth times, (3) confirm whether the cultivars (i.e., States) in the cluster analysis would also be grouped together, (4) reveal the compounds that were responsible for grouping cultivars between clusters and (5) develop a database that can predict the origins and type of unknown cannabis grown in the USA. To our knowledge, this study had never been carried out before.

**MATERIAL AND METHOD**

**Cannabis plants**

Cannabis samples (leaves and inflorescences) were collected from the supply of materials provided from seized samples by The Drug Enforcement Administration (DEA). The samples were obtained in tightly closed plastic bags and stored in a dry cool storage facility in the Coy Waller Complex at the University of Mississippi prior to analysis. The samples were selected from 23 States that have enacted Medical Marijuana laws, including: Alaska, Arizona, California, Colorado, Delaware, Florida, Hawaii, Illinois, Maine, Maryland, Massachusetts, Michigan, Montana, Nevada, New York, Ohio, Oregon, Pennsylvania, Vermont, Washington, West Virginia, Wisconsin and Mississippi.

**Reagents and solutions**

Twenty six standards of cannabinoids and terpenoids were provided from Sigma-Aldrich® (St. Louis, USA). Structural formula of organics are summarized in Table 1.
Table 1. Structural formula of assayed cannabinoids and terpenoids

| cannabinoids          | terpenoids         |
|-----------------------|--------------------|
| Δ⁴-THC/Δ⁴-Tetrahydrocannabinol | α-bisabolol |
| Δ⁹-THC/Δ⁹-Tetrahydrocannabinol | α-humulene |
| THCV/ Tetrahydrocannabivarin | α-pinene |
| CBC/Cannabichromene   | β-caryophyllene   |
| CBG/Cannabigerol      | β-pinene          |
| CBN/Cannabinol        | Caryophylleneoxide |
| CBG/Cannabicyclol     | Fenchol           |
| CBPV/ Cannabidivarin  | Carveol           |
| CBG/Cannabicyclol     | Cinol             |
| CBG/Cannabicyclol     | Sabinene          |
| CBG/Cannabicyclol     | Guaiol            |
| CBG/Cannabicyclol     | Linalool          |
| CBG/Cannabicyclol     | Limonene          |
| CBG/Cannabicyclol     | Myrcene           |
| CBG/Cannabicyclol     | Terpinolene       |
| CBG/Cannabicyclol     | Thujone           |
| CBG/Cannabicyclol     | Azulene           |
Phenanthrene (99% purity) used as the internal standard was supplied from Sigma-Aldrich®. All chemicals and solvents used for extraction and other preparations were of HPLC ultra-grade: acetone and ethyl acetate (≥99.7%), hexane (≥98.5%), ethanol (≥98%), and methanol (≥99.8%) were purchased from Sigma-Aldrich®. Chloroform (≥99.8%) was provided from Fischer Scientific (New Jersey, USA). Ultra-pure water (18 MΩ cm−1) generated by Milli-Q Plus water purification system (Millipore, Billerica, MA) was used to prepare aqueous solutions and dilutions.

**Extraction of cannabinoids and terpenoids prior to GC analysis**

Dried cannabis samples were grinded to get a homogenous mixture of leaf particles. A 100 mg portion was transferred to a test tube and 3.0 mL of extraction solution (methanol:chloroform 9:1 v/v spiked with 0.2 mg mL−1 phenanthrene) was added. Phenanthrene was used as both a retention time marker (Rt) between the terpenoids and cannabinoids and as a reference to calculate peak ratio of solute. The extraction tube was then placed in an ultrasonic water bath for 15 min to allow soluble cannabinoids and terpenes to dissolve in the extraction solution. The samples were then centrifuged for 30 s at 2,000 rpm. Finally, the extract was filtered using Acrodisc syringe filter (PAU-Gelman Lab, 0.45 μm, 25 mm diameter) and collected in a screw-capped amber vial. Samples were stored in a freezer (−10 °C) prior to analysis time. Duplicate extractions and injections were made for each cannabis sample.

**Chromatographic analysis of samples**

The chromatographic profiles of cannabis were all recorded in the splitless mode using an Agilent GC 6890 series system equipped with a 7683B autosampler. The GC column was an Agilent, DB-5, 30 m length, 0.25 mm internal diameter, film thickness 0.25 μm, (J&W Scientific Inc., Folsom, California, USA). The detailed experimental conditions applied in this study are available in our previous study [40].

Common standard stock solutions of cannabinoids and terpenoids were prepared at the concentration of 100 μg mL−1 in pure MeOH. Each solution was injected individually in two identical injections to determine the retention time of each component, and then the average was registered.

**Chromatographic profiles of extracted samples**

Selectivity was determined by injecting solvent blank to confirm that there were no false signal peaks at the targeted retention time. Intraday reproducibility was determined by injecting an aliquot of reference sample of β-pinene (100 μg mL−1) five times from the same vial in a single day (n = 5). The peak area ratio was calculated for each solute (Peak ratio = Peak area of solute/Peak area of IS) and further used to fingerprint the cannabinoids and terpenoids. The results reported as the average of two identical injections.

In the current investigation, 45 solutes were directly assayed by GC-FID compared to only 23 in our published work [40]. Hence, more comprehensive profile was adopted in this investigative. The identity of the new detected 22 solutes was identified from their retention times compared to the other common cannabinoids/terpenoids [1]. Hence, solutes detected after 30 min are related to cannabinoids-family [1].

**Numerical analysis**

Numerical analysis by PCA and HCA was carried out using Chemoface 1.61 software [42] which had been run under Matlab® (Mathworks, 8.6, USA). The size of chromatographic data used numerical analysis is 23 × 45 (23 samples × 45 organic solutes). Initially, the chromatographic data was loaded in Chemoface prior to PCA and HCA analysis.

**RESULTS AND DISCUSSION**

**Quantitative analysis of cannabis varieties**

For the comparison and the acquisition of a large number of complex chemical profiles, it is necessary to extract the maximum amount of plant contents. Therefore, a mixture of methanol: chloroform (9:1 v/v) was used to extract the highest amount of cannabinoids and terpenoids. Over the 23 USA cannabis samples, 45 active ingredients were detected. Among these ingredients, the identity of 26 compounds (eight cannabinoids and eighteen terpenoids) was established by comparing their retention times with authentic standards. The sesquiterpenoids symbolized as (Tn) and cannabinoids (Cn) have not been identified since no reference compounds were available for confirmation. The detected cannabinoids and terpenoids have variable proportions but comparable to those reported for Canadian cannabis [4]. All solutes were separated in 63 min. The sharp and intense GC peak positioned at ≈31.30 min was for phenanthrene (IS), and used for quantitative evaluation of the chemical profile contents (peak area ratio for each solute = Peak area of solute/Peak area of IS). The relative amount of separated solutes was reported as peak ratio rather concentration as reported in our previous work [40]. Moreover, numerical analysis by PCA and HCA was also based on peak ratios of solutes which is a common procedure. The position of phenanthrene encountered very small variations in both intensity and retention time during all injections. The proposed GC method was stable and convenient to quantify all the 45 organics.

In general, cannabinoid solutes such as CBN, Δ9-THC, CBG, and CBC have higher peak intensity and eluted at longer retention times compared to terpenoids, which would be attributed to the higher polarity of cannabinoids. Interday reproducibility was determined by injecting the same reference sample 12 times using fresh aliquots on each day (n = 12). The intra and inter-day precisions (%RSD) were 0.37% and 0.32%, respectively. These low RSD values indicate that the method was precise in terms of repeatability,
reproducibility, and intermediate precision. Instrumental precision (RSD), defined as the variation in peak area of the IS to all solutes was found to be 1.22%. The retention time (min) and chemical profiles of cannabis samples (expressed as the solute/internal standard peak area ratios) were obtained from duplicate measurements by GC-FID as provided in Table 2.

As indicated in Table 2, forty-five common compounds were detected in each sample, where, 8 cannabinoids and 18 terpenoids were identified based on standard comparison. Usually, the detected cannabinoids were related to six different classes: Δ⁹-Tetrahydrocannabinol (Δ⁹-THC), Δ⁹-

Tetrahydrocannabinol (Δ⁹-THC and THCV), Cannabichromene (CBC), Cannabigerol (CBG), Cannabicyclol (CBL), and Cannabinol (CBN) [7]. The other unidentified cannabinoids and terpenoids were symbolized as C_n and T_n, respectively, where n represents a number. In the current investigation, 45 solutes were directly assayed by GC-FID compared to only 23 in our published work [40]. Hence, more comprehensive profile was adopted in this investigative. The identity of the new 22 detected solutes were identified from their retention times compared to the other common cannabinoids/terpenoids, hence, solutes detected after 30 min are related to cannabinoids-family [1].

The unidentified C_n and T_n have been determined as terpenoids and cannabinoids based on comparison studies [1, 43]. It is known that cannabinoids would be available in neutral and acidic forms and quantification of both forms will require silylation/methylation of the acidic ones before GC analysis [12]. Hence, the provided data in this work gave the total contents of neutral and acidic forms of cannabinoids as no silylation of the acidic groups was carried out. Moreover, most cannabinoids are available in their neutral form, for example, ten isolated forms are known for Δ⁹-

Tetrahydrocannabinol and only two of these are in acidic form.

Among all cannabinoids, the total content in all states of Δ⁹-THC and CBN were remarkably higher than the rest of all other ingredients; 78,520 and 16,450, respectively. It is known that the psychoactive nature of cannabis is highly related to the level of Δ⁹-THC. In the same time, the high content of CBN indicated the long storage time of samples. The explanation is that the quantities of Fenchol formed in the plant, was also available in acceptable amounts of 4,323. CBN content (5,774) found to be as intermediate level between Δ⁹-THC and CBN. It occurs mainly as cannabichromenic acid (CBCA, 2-COOH–CBC, CBC–COOH). Geranylpyrophosphate and olivetolic acid combine to produce cannabigerolic acid (CBGA; the sole intermediate for all other phytocannabinoids), which is cyclized by the enzyme CBCA synthase to form CBGA. Over time, or when heated above 200°C, CBGA is decarboxylated, producing CBC. Beside cannabinoids, terpenoids impart the scent of cannabis plants where most of the peaks are present in the terpenes region (retention time ≈ 30 min).

The marked compound/s in the 23 states is/are the following: limonene, fenchol, linalool, Guaiol, and CBC are typical fingerprints for cannabis Oregon State. Regarding Nevada, it was marked for its high content of Fenchol, CBL, T3 and T7. Washington State can be distinguished easily due to its high content of CBL, C1 and Fenochol in comparison to other states, where it gives notable highest level overall states for C1 with a peak ratio of 6,386. T2 compound was characteristics for cannabis from Vermont State. THCV and CBC contents showed the highest level in Illinois State. Considering the contents obtained for Δ⁹-

THC, all states detected it with a very large content, but Hawaii (6,352), Oregon (5,459) and California (8,035), respectively, have the highest peak ratios among all States. Concerning CBN rates as the second highest content after Δ⁹-THC, where its content considered high in all states with comparable values. The highest content among all 45 organics over all states was for California State (11,192), Oregon (11,511) and Washington (11,095) that has the highest notable content of cannabinoids, especially Δ⁹-

THC and CBN.

For the rest of organics, comparable ratios for cannabis in the States have been observed. Finally, cincel, T11 and α-terpinol were not detected in all samples, and even if detected, their levels were very low. As shown, specific cannabinoids such as Δ⁹-THC, CBN, CBG, C1 and CBC are dominants in all States. This is referred to the strong influence of geographical position, maturity, age and storage conditions and the fluctuation of cannabinoids content between states with age that is numerous [1, 44, 45].

Regarding the terpenoids family, Fenchol was the most abundant one. Terpenoids were grouped into: mono-terpenoids, sesquiterpenoids, and triterpenoids. Although the number of identified terpenoids was relatively high (18 solutes), the content of cannabinoids was higher (Table 2). For terpenoids, two main classes were identified: a) Sesquiterpene including (α-bisabolol/α-humulene/β-caryophyllene/caryophylleneoxide/guaiol), and b) Monoterpene including (α-pinene/β-pinene/α-terpineol/fenchol/linalool/myrcene/ terpinolene/limonene/sabinene/carveol/cineol). The content of Fenchol was relatively high in the majority of collected samples. The explanation is that the quantities of Fenchol within cannabis plants, as well as other plants, vary significantly refers to: the growing conditions, including ground-water mineral content, soil/growing medium mineral content, soil condition, light, temperature, age of the plant, maturity of the plant, storage conditions and air pollution. Therefore, terpene’s levels vary dramatically not just from one growing region to another, but from plant to plant within the same growing area [8, 46].

Based on the content of Δ⁹-THC, cannabis can be divided into three chemical phenotypes [47]: (i) drug type, in which the major compound Δ⁹-THC is about 1–20%; (ii) intermediate type, in which Δ⁹-THC are the leading cannabinoids and their concentration range is 0.3–1.0%; and (iii) fiber type mainly contains Δ⁹-THC is in the concentration ≤0.3%. Another method to distinguish between drug-type and fiber type cannabis has been defined by the UNODC [48] with a simple mathematic equation. According to this criterion, about 86% of the cannabis samples analyzed containing detectable amount of Δ⁹-THC belongs
to the drug type. However, those classification methods are not accurate because an assumption was made that the acidic cannabinoids were completely converted to neutral cannabinoids during the decarboxylation process, for example, THCA acid decarboxylates as a result of high temperature during gas chromatography analysis to produce THC.

As a promising field combining computer science and analytical chemistry, chemometrics has increasingly found application in natural products chemistry, and has been used extensively for analytical data mining, graphical visualization, and class discrimination and prediction [37–41]. An important step in this research is data analysis where mathematical algorithms were used to extract useful information from complex datasets. For instance, Table 2 presents the chemical profiles of cannabis samples obtained from the 23 USA States (Cn and Tn are symbols for unidentified cannabinoids and terpenoids, respectively). The value represents the Peak ratio.
information from huge data sets obtained from GC-FID as will be shown in the following section.

Data analysis and classification of states based on cannabinoids and terpenoids contents

Although Table 2 and Fig. 1 give comprehensive profile of all cannabinoids and terpenoids levels in each state sample, some compounds are more important to the clustering. In order to clearly differentiate among cannabis States and to specify the compounds responsible for clustering the groups, the GC-FID scan data were subjected to HCA and PCA analysis. PCA and HCA are unsupervised clustering techniques commonly employed to reduce the complexity of multivariate data sets without losing important information, observe variance in data sets, and visualize data clustering. In our study, 45 cannabinoids and terpenoids are the original variables (45 dimensions) in PCA. By calculating the
covariance matrix between these 45 dimensions, PCA can generate 45 PCs that are orthogonal to each other and can explain 100% of the total variance of the orthogonal data. In this work, the first two PCs explain 89.73% of the total variance. Each PC is correlated with the original 45 variables. The chromatographic data was preprocessed using mean-center methodology for better interpretably of PCA and HCA outputs [49, 50]. All detected organics were rather necessary for states clustering. Accordingly, the number of variables used in clustering was 45 (detected compounds) \( \times 23 \) (USA States cannabis samples). The resulted HCA clustering of states is provided in Fig. 2.

HCA results are shown in Fig. 2. This dendrogram was obtained by calculating the Euclidean distance among samples and grouping them by the complete linkage method. There are three main groups that are clearly discriminated; Group A includes 11 States; Alaska, California, Florida, Hawaii, Illinois, Maryland, Michigan, Montana, Nevada, Ohio, and Oregon. Group B contains only Washington. Where, group C clustered 11 States; Arizona, Colorado, Delaware, Maine, Massachusetts, New York, Pennsylvania, Vermont, West Virginia, Wisconsin and Mississippi. All states are clustered while shown no mixing in different USA States.

The results of PCA projection of the data of GC into the plan of the first two principles components are carrying an accumulative average of 89.73% of the total variance. Hence, loading, score and bi-plots can be viewed using two factors only.

PCA was applied to the matrix of 23 \( \times 45 \) (23 USA States \( \times 45 \) detected cannabis). The results of PCA projection of the data from the first two principle components are carrying 89.73% of the total variance as shown in Fig. 3A.

As indicated in Fig. 3A, the score plot indicated three main different clusters collecting different number of states—this clustering corresponds with cannabinoids and terpenoids content. For example, cluster A has 11 states as mentioned in HCA analysis which is related to the similar
and/or comparable contents of cannabis samples. Cluster B has only Washington; accordingly, cannabis samples obtained from Washington is significantly different from the rest of samples from other states, due to distinct contents of C1 than the other states.

PC1 describes 77.27% of the variance of the data set, and as shown in Fig. 3A, PC1 has high positive loading for states of group C which includes 11 states and also positive loading on all cannabinoids and terpenoids except $\Delta^9$-THC, and negative loading for group A which contains 11 states (Fig. 3A) and one content which is $\Delta^9$-THC (Fig. 3B). On the other hand, PC2, accounting for 12.46% of the original information has a significance contribution from Washington State only and C1 component which makes PC2 a “cannabinoid” distinct item. Again, together, these 2 PCs account for 89.73% of the total variance in data.

Figure 3B, loading plot for PC1 and PC2, gives an intuitive explanation whereby the longer the radial separation of the compound from the center, the more important the compound is in distinguishing states. The mathematical explanation is that the radial equals the square sum of the compound’s correlations with PC1 and PC2. From the loading plot, it can be seen that the $\Delta^8$-THC was responsible for isolating 11 states Hawaii, California, Florida, Michigan, Alaska, Illinois, Montana, Nevada, Oregon, Maryland and Ohio who all have high content of $\Delta^8$-THC > 3000 and (Table 2). The position of Washington highly depends on C1, since it showed the second highest content overall components of cannabis (6,386).

In conclusion, if States are separated along PC1, they contain a distinct amount of $\Delta^8$-THC, CBN and Fenchol. If cultivars are separated along PC2, they contain different amount of cannabinoids C1.

In PCA loading plot Fig. 3B, that has been illustrated to show the most significant solutes for states clustering. $\Delta^2$-THC and C1 were not correlated with other cannabinoids and terpenoids and more significant for samples clustering. While with low distinct for clustering; Fenchol and CBN showed lower impact on states separation. This result is supported by the results obtained above for the marker solutes and indeed with the our recently published outcomes.

Table 3. The average levels of the total 45 compounds in each cluster of Fig. 2

| No. | Compound name | Cluster A | Cluster B | Cluster C |
|-----|---------------|-----------|-----------|-----------|
| 1.  | $\alpha$-Pinene | 844       | 13        | 373       |
| 2.  | Sabinene      | 232       | 48        | 287       |
| 3.  | $\beta$-Pinene | 403       | 23        | 214       |
| 4.  | Myrcene       | 394       | 23        | 193       |
| 5.  | Limonene      | 616       | 10        | 148       |
| 6.  | Cineol        | 79        | 0         | 29        |
| 7.  | Terpinolene   | 304       | 22        | 290       |
| 8.  | Linalool      | 718       | 31        | 360       |
| 9.  | Thujone       | 244       | 29        | 145       |
| 10. | T1            | 119       | 29        | 129       |
| 11. | $\alpha$-Terpinol | 91   | 0         | 43        |
| 12. | Carvel        | 162       | 17        | 277       |
| 13. | Azulene       | 214       | 6         | 158       |
| 14. | $\beta$-Caryophyllene | 1,645 | 49    | 1,261     |
| 15. | T2            | 1,211     | 7         | 1,767     |
| 16. | T3            | 679       | 95        | 205       |
| 17. | $\alpha$-Humulene | 579  | 23        | 575       |
| 18. | T4            | 171       | 27        | 84        |
| 19. | T5            | 320       | 8         | 244       |
| 20. | T6            | 802       | 9         | 713       |
| 21. | T7            | 987       | 11        | 527       |
| 22. | Caryophylleneoxide | 330  | 19        | 255       |
| 23. | Guaiol        | 749       | 23        | 338       |
| 24. | T8            | 590       | 61        | 998       |
| 25. | T9            | 586       | 28        | 632       |
| 26. | $\alpha$-Bisabolol | 669 | 35        | 470       |
| 27. | Fenchol       | 2,007     | 614       | 455       |
| 28. | T10           | 100       | 6         | 65        |
| 29. | T11           | 95        | 8         | 81        |
| 30. | T12           | 94        | 7         | 117       |
| 31. | T13           | 210       | 23        | 168       |
| 32. | CBDV          | 433       | 6         | 556       |
| 33. | THCV          | 876       | 55        | 362       |
| 34. | CBL           | 544       | 216       | 145       |
| 35. | CBC           | 3,578     | 53        | 2,143     |
| 36. | $\Delta^8$-THC | 557   | 14        | 176       |
| 37. | $\Delta^9$-THC | 53,500 | 2,191     | 22,829    |
| 38. | C1            | 768       | 6,386     | 587       |
| 39. | CBG           | 2,634     | 140       | 1,549     |
| 40. | CBN           | 9,575     | 616       | 6,259     |
| 41. | C2            | 495       | 29        | 412       |
| 42. | C3            | 189       | 22        | 131       |
| 43. | C4            | 627       | 21        | 639       |
| 44. | C5            | 783       | 33        | 719       |
| 45. | C6            | 310       | 8         | 224       |
for samples clustering, but when only 26 standards were injected as master solutes for clustering [40]. The other 41 cannabis contents were positioned close together and this indicating their limited usage for cannabis states classification. It was interesting to notice the limited performance of some important cannabinoids and terpenoids (CBC and CBG, pinene, etc.) for cannabis states clustering compared to Fenchol. In summary, Fenchol, CBN, C1 and Δ9-THC seems to be the most significant contents for cannabis clustering with comprehensive chemical profiles provided or/and only common contents included (Fig. 4) (Table 3).

As depicted in Fig. 3C, Δ9-THC, CBN and Fenchol were of high efficiency to separate large number of states from the rest. On the other hand, C1 was dominant to separate Washington away from the rest of cannabis samples obtained from other states. Compared to terpenoids, number of cannabinoids for states clustering is more significant due to their: a) therapeutic uses including pain management and neurological disorders [4, 5, 7, 11], and b) large abundance in cannabis [7]. As shown in Table 2, Δ9-THC and CBN were available in large excess compared to the rest of compounds, 78,520, 16,450, respectively. Where, Δ9-THC and CBN have notably large difference in contents than other components. In fact, Δ9-THC is a common constituent with levels varying even within the same sample depending on the composition of the sample (i.e., leaves vs. bud, vs. mixture and the ratio of small leaves to large leaves). Hence, the variation in THC content among samples is expected. In the same time, CBN is a degradation product of THC and reflects the age of the sample, seeds source and storage environments. Therefore, with the obtained separation among the states, this data set could be used as a database set to simple and fast classification future of unknown sample of cannabis from any of the studied states.

Since Δ2-THC, CBN from cannabinoids, and Fenchol from terpenoids group showed the highest content over all components, the impact of excluding these 3 contents on states clustering is studied in the next section.

Examination of chemical profiles for distinguished peaks characteristics at specific states: Impact of Δ9-THC, CBN and Fenchol

To determine if certain distinct chemical “marker” compounds presence in cannabis plants from one State, but absent in plants from another State, will affect the clustering and cannabis classification, data were re-arranged and PCA was run again with excluding the highest three contents from the chemical profiles; Δ2-THC, CBN from cannabinoids, and Fenchol from terpenoids. PCA analysis carried out 95.98% of the total variance. PCA outputs showed that excluding Δ2-THC, CBN and Fenchol: 1) has improved the separation of Nevada, Oregon and Illinois from the rest of states as shown in the score plot in Fig. 5A. Interestingly, this
states-origin cannabis when Δ⁹-THC, CBN and Fenchol were included was clustered in one cluster which is A (Fig. 2) and (Compare Figs 3A and 5A) with Alaska, Montana, Maryland, Ohio, Hawaii, California, Florida and Michigan, and this would indicate the importance of Δ⁹-THC, CBN, and Fenchol for these states. 2). In addition, removing Δ⁹-THC, CBN, and Fenchol merge cluster A and cluster B (Fig. 2) in only one group of 19 states (i.e., all 11 States of group C with 8 States of group A), which means that Δ⁹-THC, CBN and Fenchol are the main responsible for States and cannabis separation. Finally, 3). only Washington State cannabis-origin has not been change in position, which means that this state depends on C1 content and not other compounds since it was not affected upon excluding as shown in PCA (Compare Fig. 3A with Fig. 5A).

As shown in Table 2, Δ⁹-THC, CBN, C1 and CBC cannabinoids and Fenchol as terpenoids were available in large excess compared to the rest of compounds, 78,520, 16,450, 7,741, 5,774, and 3,076, respectively. Figure 5B indicated that after excluding Δ⁹-THC, CBN, and Fenchol, CBC and C1 were the most significant variables needed for samples clustering. This result supports the data obtained in Table 1 which showed that after Δ⁹-THC and CBN, compounds C1 and CBC becomes the second important and largest contents. On the other hand, CBG, CBL and THCV have same influence and both not highly correlated with C1 (angle 90°). The rest of variables (cannabis contents) were accumulated in the center indicating their limited applicability for samples clustering. In fact, this result confirms the reality and the rigidity of the outputs obtained in Table 2 and in Fig. 3 that Δ⁹-THC and CBN which have the highest contents; have the largest impact on cannabis clustering, and if excluded; CBC and C1 will be the responsible for cannabis states clustering (Fig. 5). To validate this result, bi-plot (Fig. 5C), has been demonstrated in 2PCs and 3PCs plots for clarification issue only since 2 PCs only has very overlapped data and difficult to be read. This figure indicated that CBC was necessary to separate Oregon while, C1 was necessary to isolate Washington from the rest of states. At this stage, it is clear that sample clustering is highly sensitive to the included cannabinoids, but with low distinct to terpenoids.

CONCLUSIONS

GC-FID was adopted to record the chemical profiling of 45 terpenoids and cannabinoids in 23 USA-cannabis samples. The obtained profiles were further used to cluster cannabis samples with the aid of PCA and HCA. The clustering results would uncover the geographic origin of grown cannabis specimen. Using HCA and PCA, the 23 USA cannabis plants were classified; group A consists of 11 states and also group C, where group B has only Washington State that showed totally different cannabis contents. Multivariate analysis showed also which contents are critical in discriminating cultivars since samples were grouped into 3 clusters; cluster A is THC dominant, cluster B is C1 dominant, and finally cluster C is Fenchol and CBN dominant. The results were different from cluster analysis using THC, CBN and Fenchol content only, which supports the hypothesis that classifications based exclusively on limited numbers of content may be insufficient when considering all medically relevant compounds in cannabis.

Conflicts of interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

ACKNOWLEDGMENT

Dr. Ramia Al Bakain gratefully acknowledges the financial support for her Fulbright post doctorate year 2017/2018 at The University of Mississippi, USA, provided by (1) The Binational Fulbright Commission in Jordan, and (2) the University of Jordan. Indeed, great appreciation goes to Dr. Omar Marzouk for his experimental help and technical support. Many thanks also go to Dr. Mohamed Radwan, Elsayed Ibrahim, Avery Claire Jones, and Chandrani Gon Majumdar for their assistance.

REFERENCES

1. Brenneisen, R.; Elsohly, M. Chromatographic and spectroscopic profiles cannabis of different origins: part 1. J. Forensic. Sci. 1988, 33, 1385–404.
2. Elsohly, M. E. Constituents of Cannabis Sativa; Handbook of Cannabis, 2014.
3. Lee, D. C.; Schliez, N. J.; Peters, E. N.; Dworkin, R. H.; Turk, D. C.; Strain, E. C.; Vandrey, R. Systematic review of outcome domains and measures used in psychosocial and pharmacological treatment trials for cannabis use disorder. Drug. Alcol. Depen. 2019, 194, 500–17.
4. Jin, D.; Jin, S.; Yu, Y.; Lee, C.; Chen, J. Classification of cannabis cultivars marketed in canada for medical purposes by quantification of cannabinoids and terpenes using HPLC-DAD and GC-MS. J. Anal. Bioanal. Tech. 2017, 8, 1–9.
5. Sohly, M. E. Constituents of Cannabis Sativa; Handbook of Cannabis, 2014.
6. Pertwee, R. G. Handbook of Cannabis; Oxford University Press, 2007, pp 33–49.
12. Wang, M.; Wang, Y.-H.; Avula, B.; Radwan, M. M.; Wasas, A. S.; Mehmmedic, Z.; van Antwerp, J.; Elsohly, M. A.; Khan, I. A. Quantitative determination of cannabinoids in cannabis and cannabis products using ultra-high-performance supercritical fluid chromatography and diode array/mass spectrometric detection. *J. Forensic. Sci.* 2017, 62, 602–11.

13. Baron, E. Comprehensive Review of Medicinal Marijuana, Cannabinoids, and Therapeutic Implications in Medicine and Headache: What a Long Strange Trip It’s Been, 2015.

14. Schier, A. R.; Ribeiro, N. P.; Silva, A. C.; Hallak, J. E.; Crippa, J. A.; Nardi, A. E.; Zuardi, A. W. Cannabidiol, a Cannabis sativa constituent, as an anxiolytic drug. *Rev. Bras. Psiquiatr.* 2012, 34, S104–17.

15. Porter, B.; Jacobson, C. Report of a parent survey of cannabidiol-enriched cannabis use in pediatric treatment-resistant epilepsy, 2013.

16. http://www.hc-sc.gc.ca/dhp-mps/marihuana/med/infoprof-eng.php.

17. Amar, M. B. Cannabinoids in medicine: A review of their therapeutic potential. *J. Ethnopharmacol.* 2016, 165, 1–25.

18. Hazekamp, A.; Grothenhermen, F. Review on clinical studies with cannabis and cannabinoids 2005–2009. *Cannabinoids* 2010, 5, 1–21.

19. Hillig, K. W.; Mahlberg, P. G. A chemotaxonomic analysis of cannabinoid variation in cannabis (cannabaceae). *Am. J. Bot* 2004, 91, 966–75.

20. Hillig, K. W. A chemotaxonomic analysis of terpenoid variation in Cannabis. *Biochem. Syst. Ecol.* 2004, 32, 875–91.

21. Hillig, K. Genetic evidence for speciation in Cannabis (Cannabaceae). *Genet. Resour. Crop Evol.* 2005, 52, 161–80.

22. Szabady, B.; Hidvégi, E.; Niryredy, S. Determination of neutral cannabinoids in hemp samples by overpressed-layer chromatography. *Chromatographia* 2002, 65, S165–8.

23. Zuardi, A. W.; Hallak, J. E. C.; Crippa, J. A. S. Interaction between cannabidiol (CBD) and Δ9-tetrahydrocannabinol (THC): influence of administration interval and dose ratio between the canna-binoids. *Psychopharmacology* 2012, 219, 247–9.

24. Stromberg, L. Minor components of Cannabis Resin III: Comparative Gas chromatographic analysis of hashish. *J. Chro.* 1972, 68, 253–8.

25. Novotny, M.; Lee, M.L.; Low, C.-E.; Raymond, A. Analysis of marijuana samples from different origins by high-resolution gas-liquid chromatography for forensic application. *Anal. Chem.* 1976, 48, 24–9.

26. Hazekamp, A.; Fischedickm, J. T. Cannabis—from cultivator to chemomar, Drug Test. Analysis; Wiley Online Library, 2012.

27. Ndjoko, K.; Wolfender, J.-L.; Hostettmann, K. Analysis of cannabino-ids by liquid chromatography—Thermospray mass spectrometry and liquid chromatography—Tandem mass spectrometry. *Chromatographia* 1998, 47, 72–6.

28. Van der Kooy, F.; Maltese, F.; Choi, Y. H.; Kim, H. K.; Verpoorte, R. Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. *Planta Med.* 2009, 75, 763–75.

29. Politi, M.; Peschel, W.; Wilson, N.; Zloh, M.; Prieto, J. M.; Heinrich, M. Direct NMR analysis of cannabis water extracts and tinctures and semiquantitative data on D9-THC and D9-THC-acid. *Phytochemistry* 2008, 69, 562–70.

30. Choi, Y. H.; Kim, H. K.; Hazekamp, A.; Erkelens, C.; Lefeber, A. W. M.; Verpoorte, R. Metabolomic differentiation of Cannabis sativa cultivars using 1H NMR spectroscopy and principal component analysis. *J. Nat. Prod.* 2004, 67, 953–7.

31. Breitenbach, S.; Rowe, W. F.; McCord, B.; Lurie, I. S. Assessment of ultra high performance supercritical fluid chromatography as a separation technique for the analysis of seized drugs: Applicability to synthetic cannabinoids. *J. Chrom. A* 2016, 1440, 201–11.

32. Toyo-o’oka, T.; Kikura-Hanajiri, R. A reliable method for the separation and detection of synthetic cannabinoids by supercritical fluid chromatography with mass spectrometry, and its application to plant products. *Chem. Pharm. Bull.* 2015, 63, 762–9.

33. Bäckström, B.; Cole, M. D.; Carrott, M. J.; Jones, D. C.; Davidson, G.; Coleman, K. A preliminary study of the analysis of Cannabis by supercritical fluid chromatography with atmospheric pressure chemical ionisation mass spectrometric detection. *Sci. Justice* 1997, 37, 91–7.

34. Later, D. W.; Richter, B. E.; Knowles, D. E.; Andersen, M. R. Analysis of various classes of drugs by capillary supercritical fluid chromatography. *J. Chromatogr. Sci.* 1986, 24, 249–53.

35. Lehmann, T.; Brenneisen, R. High performance liquid chromatogra- phic profiling of cannabis products. *J. Liq. Chrom.* 1995, 18, 689–700.

36. Crispino, C.; Fernandes, K.; Kamogawa, M.; Nóbrega, J.; Nogueira, A. R.; Ferreira M. Multivariate classification of cigarettes according to their elemental content determined by inductively coupled plasma optical emission spectrometry. *Anal. Sci.* 2007, 23, 435–8.

37. Al Bakain, R.; Rivals, I.; Sassiat, P.; Thiébaut, D.; Hennion, M.-C.; Euvrard, G.; Vial, J. Comparison of different statistical approaches to evaluate the orthogonality of chromatographic separations: Application to reverse phase systems. *J. Chrom. A* 2011, 1218, 2963–75.

38. Al Bakain, R.; Rivals, I.; Sassiat, P.; Thiébaut, D.; Hennion, M.-C.; Euvrard, G.; Vial, J. Impact of the probe solutes set on orthogonality evaluation in reverse phase chromatographic systems. *J. Chrom. A* 2012, 1232, 231–41.

39. Al Bakain, R. Z.; Al-Deys, Y.; Andri, B.; Thiébaut, D.; Vial, J.; Rivals, I. Supercritical fluid chromatography of drugs: parallel factor analysis for column testing in a wide range of operational conditions. *J. Anal. Methods Chem.* 2017, 2017, 1–13.

40. Al Bakain, R. Z.; Al-Deys, Y. S.; Cizdziel, J. V.; Elsohly, M. A. Comprehensive classification of USA cannabis samples based on chemical profiles of major cannabinoids and terpenoids. *J. Liq. Chrom. Relat Tech.*, 2019, doi:10.1080/10826076.2019.1701015.

41. Andri, B.; Dispas, A.; Marini, R. D. E.; Hubert, P.; Sassiat, P.; Al Bakain, R.; Thiébaut, D.; Vial, J. Combination of partial least squares regression and design of experiments to model the reten- tion of pharmaceutical compounds in supercritical fluid chromato- graphy. *J. Chrom. A* 2017, 1491, 182–94.

42. Nunes, C. A.; Freitas, M. P.; Pinheiro, A. C. M.; Bastos, S. C. Chemoface: a novel free user-friendly interface for chemometrics. *J. the Brazil. Chem. Soc.* 2012, 23, 2003–10.

43. Fischedick, J. T.; Hazekamp, A.; Erkelens, T.; Choi, Y. H.; Verpoorte, R. Metabolic fingerprinting of Cannabis sativa L., cannabino- ids and terpenoids for chemotaxonomic and drug standardization purposes. *Phytochemistry* 2010, 71, 2058–73.

44. Turner, C. E.; Ma, C. Y.; Russell, M. H.; Elsohly, M. A. Analysis of micro-encapsulated d-limonene dimercaptan, a possible herbicide
marker for Cannabis sprayed with paraquat, using gas chromatography. Bull. Narc. 1981, 33, 43–54.
45. Turner, C. E.; Ma, C. Y.; Elsohly, M. A. Constituents of Erythroxylin coca II. Gas-chromatographic analysis of cocaine and other alkaloids in coca leaves. J. Ethnopharmacol. 1981, 3, 293–8.
46. http://www.maximumyield.com/definition/3932/fenchol.
47. Broséus, J.; Anglada, F.; Esseiva, P. The differentiation of fibre- and drug type Cannabis seedlings by gas chromatography/mass spectrometry and chemometric tools. Forensic Sci. Int. 2010, 200, 87–92.
48. United Nations Office on Drug Crime. Recommended method for the identification and analysis of cannabis and cannabis products; Vienna, Austria UNODC 39-41, 2014.
49. Brereton, R. G. Applied Chemometrics for Scientists. John Wiley & Sons: England, 2007.
50. Otto, M. Chemometrics: Statistics and Computer Application in Analytical Chemistry, 3rd ed.; Wiley-VCH, Weinheim, 2016.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1556/1326.2020.00767.

Open Access statement. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.