Analytica Chimica Acta 1184 (2021) 339020

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Cannabinoids vs. whole metabolome: Relevance of cannabinomics in analyzing *Cannabis* varieties

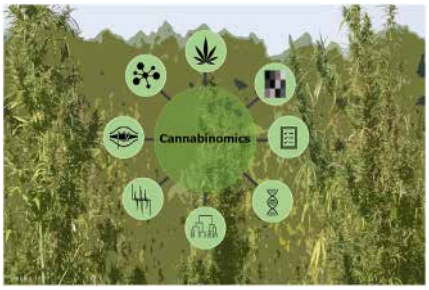
Pedro G. Vásquez-Ocmín [a, b, c, *], Guillaume Marti [d], Maxime Bonhomme [e], Fabienne Mathis [b], Sylvie Fournier [d], Stéphane Bertani [c], Alexandre Maciuk [a]

[a] BioCIS, CNRS, Université Paris-Saclay, 92290, Châtenay-Malabry, France
[b] HEMP-it ADN, 49250, Beaufort-en-Anjou, France
[c] UMR152 PHARMADEV, IRD, UPS, Université de Toulouse, Toulouse, France
[d] Laboratoire de Recherche en Sciences Végétales UMR 5546 UPS/CNRS, Plateforme MetaboHUB - MetaToul - Agromix, Auzeville-Tolosane, France
[e] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse INP, Auzeville-Tolosane, France

HIGHLIGHTS

- 20 varieties of different chemotypes were genotyped and profiled by LC-HRMS.
- Dereplication by MSCleanR workflow and molecular networking resulted in the annotation of 396 compounds.
- Different strategies to differentiate chemotypes were compared by applying statistical analyses either on genotyping data, on the whole metabolome dataset or on the results of the quantification of 11 phytocannabinoids.
- This approach named cannabinomics is fit-to-purpose for many applications, *e.g.* selection of new varieties, quality control, etc.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 May 2021
Received in revised form
30 July 2021
Accepted 30 August 2021
Available online 2 September 2021

Keywords:
Cannabis
Cannabinomics
Chemotypes
Metabolome
Phytocannabinoids

ABSTRACT

Cannabis sativa has a long history of domestication both for its bioactive compounds and its fibers. This has produced hundreds of varieties, usually characterized in the literature by chemotypes, with Δ^9-THC and CBD content as the main markers. However, chemotyping could also be done based on minor compounds (phytocannabinoids and others). In this work, a workflow, which we propose to name cannabinomics, combines mass spectrometry of the whole metabolome and statistical analysis to help differentiate *C. sativa* varieties and deciphering their characteristic markers. By applying this cannabinomics approach to the data obtained from 20 varieties of *C. sativa* (classically classified as chemotype I, II, or III), we compared the results with those obtained by a targeted quantification of 11 phytocannabinoids. Cannabinomics can be considered as a complementary tool for phenotyping and genotyping, allowing the identification of minor compounds playing a key role as markers of differentiation.

* Corresponding author. BioCIS, CNRS, Université Paris-Saclay, 92290, Châtenay-Malabry, France.
E-mail addresses: vasco2224@gmail.com (P.G. Vásquez-Ocmín), alexandre.maciuk@universite-paris-saclay.fr (A. Maciuk).

https://doi.org/10.1016/j.aca.2021.339020
0003-2670/

1. Introduction

Cannabis sativa L. (Cannabaceae) is an herbaceous annual plant cultivated since ancient times mainly in Central Asia and South-East Asia [1]. It has been used as a source of fibers, foods, and its specific class of specialized metabolites called cannabinoids, which are the basis for religious, recreational, and medicinal purposes [2]. Recent analysis suggests that *Cannabis* producing high levels of psychoactive compounds was smoked as part of ritual activities in western China at least 2500 years ago [3]. So far, over 600 constituents have been reported in *Cannabis* plants, including ≈200 terpenes (mono-, di-, sesqui-and triterpenes), ≈25 flavonoids, ≈ 150 cannabinoids, and other compounds like stilbenes, lignans, phytosterols, alkaloids, and amides [4,5]. Cannabinoids from *C. sativa*, localized in glandular trichomes, are also called phytocannabinoids to distinguish them from the mammalian endogenous (endocannabinoids) or synthetic compounds able to bind to cannabinoid receptors. These terpenophenolic compounds (i.e., isoprenylated polyketides) are considered as the active constituents of *Cannabis* plants and may contain 19 to 23 carbons depending on the length of the alkyl side chain [6]. They originate from the convergence of two distinct biosynthetic pathways: the polyketide pathway, giving rise to olivetolic acid (OLA) or its C1, C5, or C7 analogs, and the plastidial methyl-D-erythritol 4-phosphate (MEP) pathway, leading to the prenylation of OLA with geranyl diphosphate (GPP) that undergoes oxidative cyclization reactions [7,8]. Δ^9-Tetrahydrocannabinolic acid (Δ^9-THCA), cannabidiolic acid (CBDA), cannabichromenic acid (CBCA), and their common precursor cannabigerolic acid (CBGA) are the major phytocannabinoids found natively in most of the varieties. These acidic cannabinoids are rather unstable and may spontaneously lose the carboxyl group with exposure to light, air, or heat. Their decarboxylated analogs like cannabidiol (CBD) or Δ^9-tetrahydrocannabinol (Δ^9-THC) show a different, and often higher, bioactivity [9,10]. Many phytocannabinoids were characterized between the 1960s and 1980s [11–13], although new cannabinoids, sometimes showing cannabimimetic action, are continuously identified [6,14–16]. *Cannabis* chemotypes (also called phenotypes or chemovars) are commonly split into three groups based on their Δ^9-THC/CBD ratio: Δ^9-THC/CBD $\gg$ 1 (chemotype I, "drug-type"), Δ^9-THC/CBD $\approx$ 1 (chemotype II, "intermediate-type"), Δ^9-THC/CBD << 1 (chemotype III, "fiber-type" or "hemp") [17–19]. Such a differentiation was historically driven by several reasons, including regulatory, safety and agronomical purposes. Hemp must be differentiated from recreative *Cannabis* and thus the phytocannabinoids content has been the major focus for decades. Additional chemotyping can also be based on the terpenoids, with at least five chemotypes described [20–22]. Hemp usually contains significant amounts of non-intoxicating cannabinoids, *e.g.* cannabidiolic acid (CBDA) or cannabigerolic acid (CBGA). Although environmental factors play a key role in determining the amount of phytocannabinoids present in the plant at different growth stages, the Δ^9-THC/CBD ratio is under genetic control and is quite stable for one given chemotype [23].

To date, medicinal use of *Cannabis* is authorized in more than 50 countries, while its recreational use is authorized in around 20 countries. In December 2020, The UNODC Commission on Narcotic Drugs removed *Cannabis* from Schedule IV, which lists drugs being particularly liable to abuse and to produce ill effects, to incorporate it into Schedule I, which allows its therapeutic use [24]. *Cannabis* and isolated phytocannabinoids have shown beneficial effects for chronic pain, multiple sclerosis, neurodegenerative disorders, treatment-resistant epilepsy, nausea and vomiting associated with chemotherapy diseases, cancer, cardiovascular disorders, metabolic disorders, and more [25–29]. Besides its current move towards medicinal *Cannabis* through an experimental status for 2 years (2021–2023), France has been for decades a leading producer of hemp in Europe. The commercial hemp seed producer (HEMP-it, formerly known as *Fédération Nationale des Producteurs de Chanvre*) holds a large number of *Cannabis* fiber-type varieties [30–34].

The scientific literature on *Cannabis* has grown at a staggering pace, with 10 articles published per day. Yet, a majority of them pertain to Δ^9-THC and CBD and to far less extent the varin type Δ^9-THCV, leaving the remaining metabolites somewhat in the shadows. Chemical differentiation of *Cannabis* chemotypes is currently done by using 2 to 15 major cannabinoids as principal markers. Meanwhile, comprehensive metabolomic approaches allowing the detection and annotation of numerous metabolites are currently becoming mainstream in plant science and have become a key to decipher their biological roles. Techniques used for detection are hyphenated liquid or gas chromatography-mass spectrometry (LC- or GC-MS) or nuclear magnetic resonance spectroscopy (NMR), both techniques being assisted by bioinformatics and statistical analysis. We applied a LC-MS-based metabolomics workflow on medium polarity extracts of *C. sativa* leaves and flowers to classify chemotypes as well as map and annotate metabolites of the phytocannabinoids family and other phytochemical classes. In a collection of 20 genetically diverse chemotypes, we compared the ability of this approach to decipher chemotypes markers with an approach based only on major cannabinoids quantification.

2. Material and methods

All solvents used for chromatography were LC-MS grade (HiPerSolv Chromanorm, VWR, France, Fontenay-Sous-Bois). Milli-Q RG system (Millipore, France) was used to produce high purity water of 18.2 MΩ cm resistivity. Formic acid was >98% for LC-MS (Fluka, Buchs, Switzerland). Phytocannabinoids standards (CBDA, CBGA, CBCA, Δ^9-THCA, CBD, CBG, CBC, Δ^9-THC, Δ^8-THC, THCV, and CBN) were purchased from Sigma-Aldrich (St. Louis, MO, USA) and immediately stored at −30 °C.

2.1. Plant material

Thirteen fiber-type (chemotype III) and three intermediate-type (chemotype II) *Cannabis sativa* varieties were obtained from the HEMPT-it ADN company (Beaufort-en-Anjou, France) and four drug-type varieties (chemotype I) were obtained from CBDIS company (Paris, France) (see the list and description of varieties in **supp info 1**). Chemotype III and II samples were dried at 40 °C, while chemotype I was dried at 80 °C; all samples were finely ground, then stored at room temperature.

2.2. Metabolome analysis

For each sample, 100 mg were extracted in triplicate using 5 mL of acetonitrile/methanol (8:2 v/v) under sonication for 30 min. Extracts were then centrifuged at 4000 g for 15 min. Supernatants were filtered using PTFE membrane syringe filter, (50 mm × 0.2 μm) and stored in vials at −30 °C. A quality control (QC) sample was created by pooling aliquots from each extract. Extracts were diluted 10 times and standards were prepared at 10 μg/mL in the same solvent. Ultra-High Performance Liquid Chromatography – High-Resolution MS (UHPLC–HRMS) analyses were performed on a Q Exactive Plus orbitrap mass spectrometer, equipped with a heated electrospray probe (HESI II) coupled to a U-

HPLC Ultimate 3000 RSLC system (Thermo Fisher Scientific, France). Separation was done on a Luna Omega Polar C18 column (150 mm × 2.1 mm i.d., 1.6 μm, Phenomenex, Sartrouville, France) equipped with a guard column. The mobile phase gradient used water containing 0.05% formic acid (A) and acetonitrile containing 0.05% of formic acid (B). A gradient method at a constant flow rate of 0.3 mL/min was applied under the following conditions: 60–100% B in 20 min, 100% B over 2 min, equilibration over 4 min. The column oven temperature was set to 30 °C, the autosampler temperature was set to 5 °C, and the injection volume was fixed to 2 μL. Mass detection was performed in positive ionization (PI) mode at resolution 30,000 [fullwidth at half-maximum (fwhm) at 400 m/z] for MS1 and resolution 17,500 for MS2 with automatic gain control (AGC) target of 1×10^6 for full scan MS1 and 1×10^5 for MS2. Ionization spray voltage was set to 3.5 kV for PI, and the capillary temperature was set to 256 °C. The mass scanning range was *m/z* 100–1500. Each full MS scan was followed by a data-dependent acquisition of MS/MS spectra for the six most intense ions using stepped normalized collision energy of 20, 40, and 60 eV.

2.3. Data mining process

LC-MS data were processed following the MSCleanR workflow [35]. Briefly, samples were processed with MS-DIAL version 4.38 [36]. MS1 and MS2 tolerances were set to 0.01 and 0.05 Da, respectively, in centroid mode for each dataset. Peaks were aligned on QC reference with a RT tolerance of 0.2 min, a mass tolerance of 0.015 Da, and minimum peak height detection at 4.5×10^6. MS-DIAL data were deconvoluted with MS-CleanR by selecting all filters with a minimum blank ratio set to 0.8, and a maximum relative standard deviation (RSD) set to 40. The maximum mass difference for feature relationships detection was set to 0.005 Da, and the maximum RT difference was set to 0.025 min. The Pearson correlation links were applied for correlation ≥0.8 and *p*-value significance threshold = 0.05. Two peaks were kept in each cluster for further request and the kept features were annotated with MS-FINDER version 3.50 [37]. The MS1 and MS2 tolerances were set to 5 and 15 ppm, respectively. Formula finder was exclusively processed with C, H, O, N, atoms. For compounds annotation, several sets of data were used: 1) experimental LC-MS/MS data of 11 phytocannabinoid standards, analyzed on the same instrument and with the same method, were used as references using MSDial based on accurate and exact mass, fragmentation, retention time; 2) compounds identified in the *Cannabis* genus and the Cannabaceae family, generated from literature data (Dictionary of Natural Products, DNP on DVD v. 28.2, CRC press) were mined by MS Finder based on exact mass and *in-silico* fragmentation; 3) a generic database encompassing exact mass and MS/MS data of compounds from natural products databases included in MS-FINDER (KNApSAcK, PlantCyc, ChEBI, NANPDB, UNPD, and COCONUT). Statistical analysis and mass spectra similarity network (molecular network) was carried out using the open-source software MetGem [38] on final.msp and.csv files obtained with MSCleanR. Values used were MS2 *m/z* tolerance = 0.05 Da, minimum matched peaks = 4 and minimal cosine score value = 0.7. Visualization of the network was performed on Cytoscape version 3.8.2 [39].

2.4. Genotyping

Fifteen cultivars from chemotypes II and III (excepted Epsilon variety) were studied with 24 plants per cultivar to perform genetic analyses. DNA extraction was performed with the Macherey nagel NucleoSpin Plant II kit from foliar discs taken from young plant leaves. Sixteen SSR markers (see **supp info 2**) were chosen for their polymorphism and repeatability [40,41]. The PCR MIX contains 10 ng of genomic DNA, 250 μM dNTP, 0.2 μM of primer, 2.0 mM of $MgCl_2$, 2.0 μg μl^{-1} of BSA, 10 mM of PCR buffer, and 0.3 units Taq Gold polymerase (Perkin Elmer). PCRs consisted of a 5 min preheat at 95 °C followed by two rounds of cycles. Touchdown PCR consisted of 10 cycles of 95 °C denaturation for 15 s, 63 °C touchdown to 53 °C annealing for 30 s, and 72 °C extension for 30 s. This was followed by 19 cycles of 95 °C denaturation for 15 s, 52 °C annealing for 30 s, 72 °C extension for 1 min, and one 72 °C final extension for 10 min.

2.5. Quantification of main phytocannabinoids

Mixes of 11 phytocannabinoids at six final concentrations (10000, 1000, 100, 10, 1, 0.1, and 0.001 ng/mL) were performed to obtain calibration curves for quantitative analysis. All extracts were profiled using the LC-MS method described above and were processed using MZmine 2.53 [42] Briefly, after mass detection, the chromatogram was built and deconvolutioned using ADAP chromato-builder and wavelets (ADAP) algorithms [43]; a grouping of isotope patterns (peak grouper algorithm) and a unique pick list aligned was created; finally, a gap-filling (peak finder algorithm) was performed. A final list of the peak height for each standard at each concentration was carried out and a calibration curve was created for each standard. Validation parameters were determined following the International Conference on Harmonization (ICH) Guidelines [44], and the following characteristics were evaluated: linearity, limit of detection (LOD), and limit of quantification (LOQ).

2.6. Statistical and genetic analyses

The final annotated metabolome dataset generated by the MS-CleanR workflow was uploaded on the MetaboAnalyst 5.0 online platform [45]. The data were normalized by sum and scaled by unit variance before statistical analysis. First, principal component analysis (PCA) was applied as an exploratory data analysis to provide an overview of LC-MS fingerprints. Then, a hierarchical cluster analysis (HCA) was performed to obtain a dendrogram of varieties using the metabolome dataset and the phytocannabinoids quantification dataset. Briefly, for each variable the average value across replicates was calculated, then distance matrices were obtained between the 20 varieties using either 1-*r* (the Pearson correlation coefficient) for the metabolome dataset or the Euclidean distance for the phytocannabinoids quantification dataset, followed by HCA on each matrix using Ward's minimum variance criterion. On the clusters obtained, a partial least squares discriminant analysis (PLS-DA) was conducted using clusters as Y value and the top 50 features were plotted on a heatmap using ANOVA. Additionally, to look for specific biomarkers of each variety the same statistical approach was applied within each cluster.

Genetic differentiation among fifteen *Cannabis* varieties based on 16 microsatellite markers was measured by F_{ST} using the R package *adegenet*, and visualized by HCA using Ward's minimum variance criterion. To evaluate the degree of genetic structuring of cannabis varieties without *a priori*, we used the software STRUCTURE v2.3.4 [46], which implements a Bayesian algorithm to identify *K* user-defined clusters of genetically homogeneous individuals. We analyzed the data with $K = 2$ to 20. The analysis produced for each individual a proportion of membership to each cluster. We used both the "admixture" and "no admixture" models. Allelic frequencies were set to correlate among populations. All analyses were replicated 10 times to ensure proper convergence of the MCMC with a burn-in of 100 000 steps and a MCMC length of 1 000 000 after burn-in. Posterior inference of the optimal number of clusters was performed using the *ad hoc* statistic ΔK based on the likelihood [47].

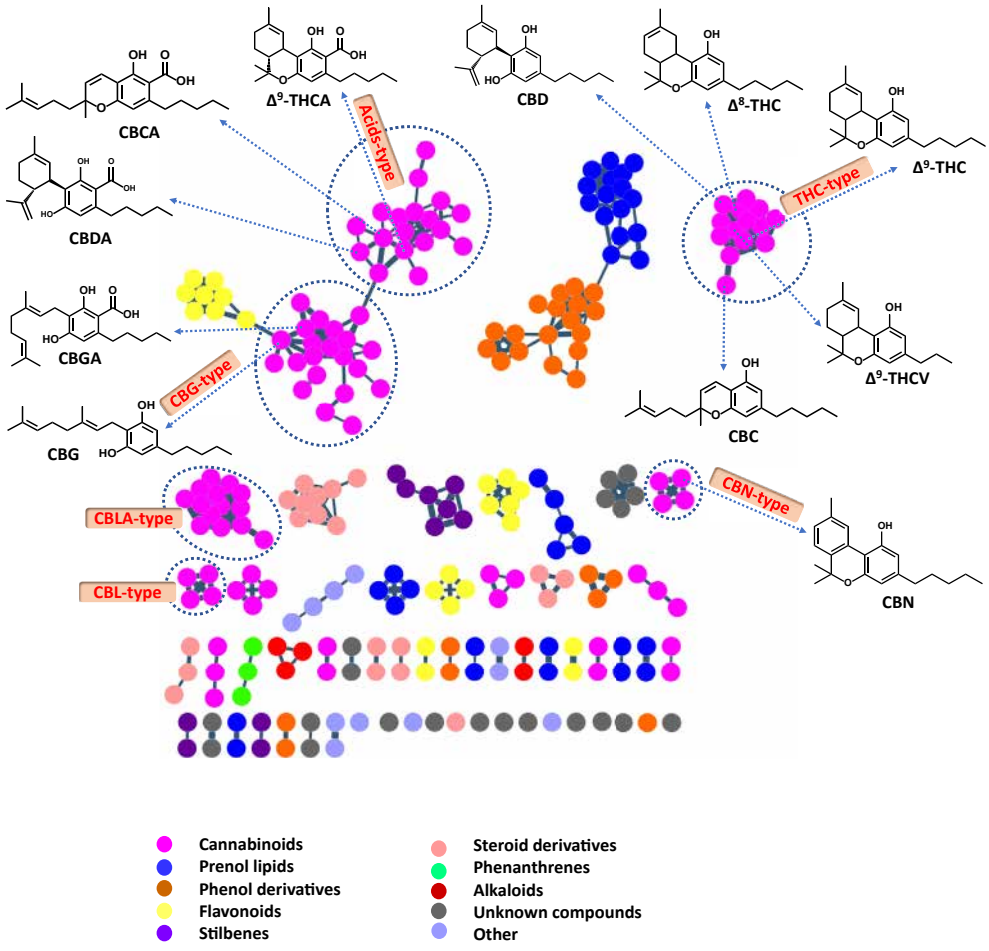


Fig. 1. Mass spectral similarity network (molecular network) of annotated compounds based on MetGem software and visualized by Cytoscape. Colors show the major class of compounds in the metabolome for all *Cannabis* varieties, based on Classyfire attribution [55] and MetGem clustering. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3. Results and discussion

3.1. Cannabinomics reveals major discriminant compounds between Cannabis varieties

Identification of phytocannabinoids in *Cannabis* plants is difficult for several reasons: only a few of them are commercially available, those that are available ones are expensive to purchase, access to some standards is restricted by national regulations, and the carboxylic analogs are highly labile. Therefore, most studies encompass only a handful of standards. An increasing number of studies use mass spectral libraries (*i.e.* NIST20 library) [48] or in-house databases [49] for the annotation of the *Cannabis* metabolome [48], in a metabolomic approach variously referred to as Cannabinomics, Cannabinoidomics, or Phytocannabinomics [50–52]. We used several mass spectral libraries and three levels of annotations (see above in **2.3**), allowing in most cases an annotation of level 2 according to the metabolomic standard initiative [53]. Application of the MSCleanR workflow to the entire LC-MS dataset in positive mode resulted in 396 compounds. Annotation prioritization was done by ranking from highest to lowest confidence as follows: 1) standards, 2) genus and family DB, and 3) generic DB. Among the compounds detected, 193 had a match in the *Cannabis* genus (48.7%), 41 in the Cannabaceae family (10.4%), 84 in the generic database (21.2%) while 78 remained unannotated (19.7%) (Table in Supp info 3). Extraction with ACN/MeOH 8:2 v/v is expected to yield the medium and low polarity parts of the metabolome, including phytocannabinoids and non-polar compounds. The set of extracts allowed identification or annotation of 105 phytocannabinoids under acidic or neutral form (26.5% of compounds annotated). Other metabolites annotated mostly included flavonoids and terpenoids (mono-, sesqui-, triterpenoids, prenol lipids) (Fig. 1). Unknown compounds were distributed evenly through the chromatogram and their masses spaned in the *m/z* 300–1200 range. Structural similarity between compounds, assessed *via* fragmentation pattern similarity, is reflected by the molecular network (MN) (Fig. 1 and **supp info 4**). Phytocannabinoid standards were dispatched into 4 clusters, following pathways of fragmentations consistent with the previous report of Berman et al. [54]. One cluster gathers neutral forms of THC and CBD, with several compounds being annotated as isomers of THC (Δ^8-, Δ^7-, Δ^4-, $\Delta^{4\text{-}8}$-, Δ^4-, Δ^4-THC), as well as C3-derivatives (varin-type). Another cluster gathers acidic derivatives of THC, CBD, and CBC. Both neutral and acidic derivatives of CBG-type cannabinoids are grouped in another cluster. Four compounds are grouped with CBN in a cluster that contains degradation products, annotated as cannabinodiol isomers and cannabinol hydroxylated on the side chain. It is noteworthy that CBE derivatives (degradation products of CBD)

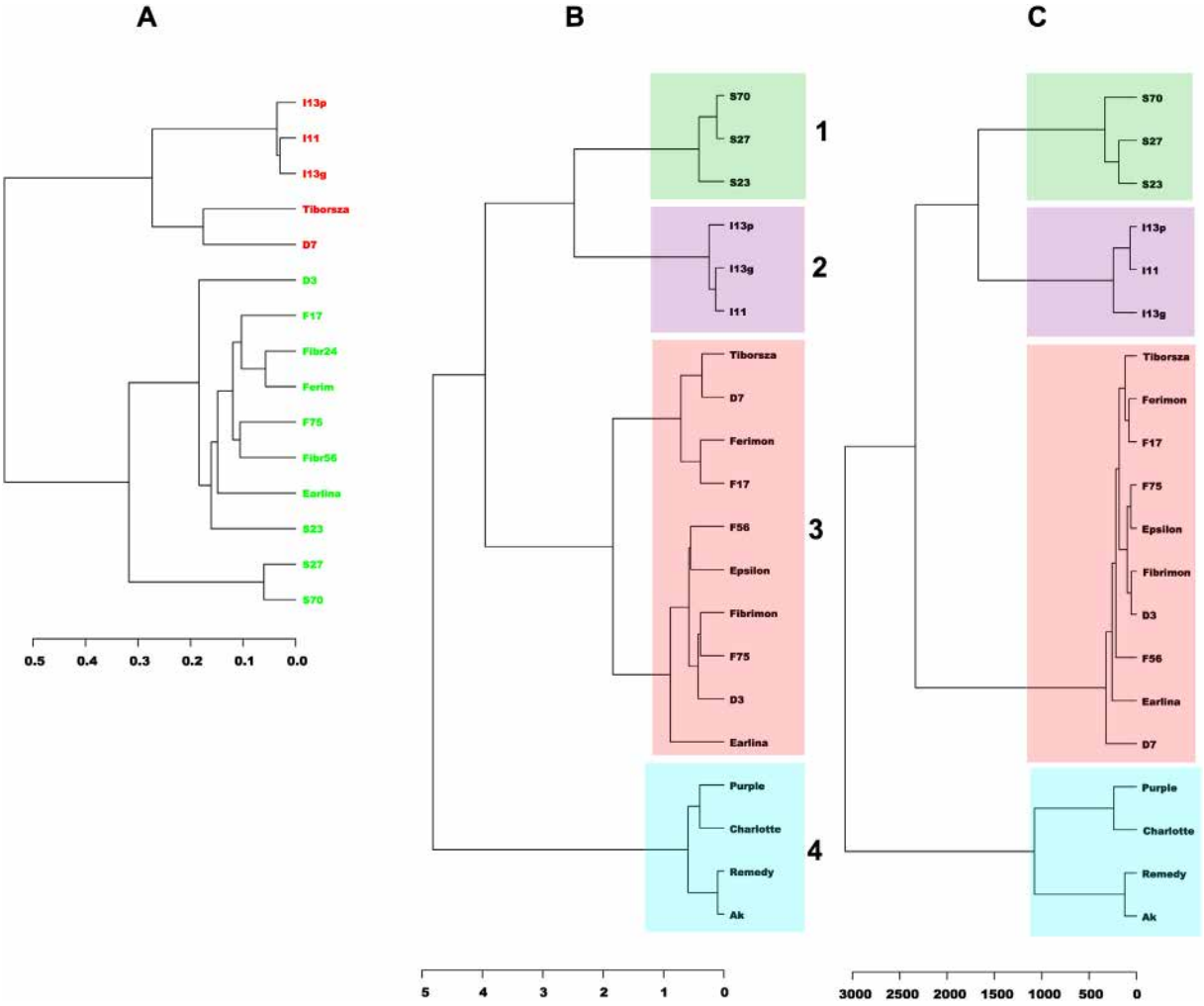


Fig. 2. Hierarchical cluster analyses (HCA) following genetic differentiation (Fst) (**A**), the global metabolome (**B**), and main phytocannabinoids quantified (**C**). Varieties of *Cannabis*: **S27** = *Santhica 27*; **S23** = *Santhica 23*; **S70** = *Santhica 70*; **Tiborsza** = *Tiborszallasi*; **D3** = *Dioïca 3*; **D7** = *Dioïca 7*; **Ferimon** = *Ferimon*; **F17** = *Fedora 17*; **F56** = *Fibrimon F56*; **F75** = *Futura 75*; **Epsilon** = *Epsilon 68*; **Fibrimon** = *Fibrimon 24*; **Earlina** = *Earlina* 8FC; **I11** = *I11*; **I13g** = *I13g*; **I13p** = *I13p*, **AK** = *AK Silver*; **Charlotte** = *Charlotte*; **Purple** = *Purple*; **Remedy** = *Remedy*. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are clustered with CBG derivatives. Cannabicyclolic acid (CBL) derivatives, which are degradation products of cannabichromene (CBC), are clustered in a group containing 14 compounds.

Currently, more than 600 *Cannabis* varieties are commercially available whose genetics, for many of them, is only partially known [50]. In this work, we characterize twenty varieties of *Cannabis* (sixteen hemp varieties and four recreative varieties) using metabolomic tools. To visualize relationships between varieties based on their genotypes and phenotypes, hierarchical cluster analyses (HCA) were performed on each set of data (Fig. 2-A to C). HCA based on genetic differentiation among varieties (F_{ST}) suggests a clustering into two main genetic groups (showed in red and green, Fig. 2-A) as confirmed by the STRUCTURE analyses (**supp info 5**). Most of the varieties grouped together (in green) belong to a pool of varieties developed by Hemp-It by cross-breeding and massal selection, still showing a substantial amount of genetic diversity. Clustering based on LCMS profiles on a wider panel of varieties (Fig. 2-B and **2-C**) yields a very similar clustering for both global metabolome data and quantified main cannabinoids data. These first analyses validate the long-standing literature classifying *C. sativa* varieties based on the THC/CBD ratio, leading to 3 chemotypes: hemp-type (chemotype III), drug-type (chemotype I), and intermediate type (chemotype II) [17–19]. Among the varieties we studied, chemotype III encompasses **cluster 3** (CBD-majority varieties) and **1** (CBG-majority varieties), whereas chemotype II is illustrated by varieties used for seed production (**cluster 2**). Drug-type varieties (chemotype I) are clustered out (**cluster 4**). Samples of chemotype III were developed by HEMP-it (French hemp seed producer) in a dedicated work based on GC and TLC analyses to obtain non-monoclonal varieties with THC content below 0.2% [34]. While the selection criterion produced a homogeneous chemotype III, minor metabolites may nevertheless be used to distinguish sub-varieties, as we show below.

PCA analysis is a preliminary step in a multivariate analysis to provide an unsupervised overview of LC-MS fingerprints. An unsupervised PCA analysis on MetaboAnalyst 5.0 [45] was carried out to determine how chemotype metabolomes differ from each other, and which metabolites contribute the most to this difference (Fig. 3-A). In our analysis, 51.1% of the total variance was projected in two principal component axes. As expected, replicated injections of the quality control sample (QC, made by mixing an aliquot of each extract) were grouped near the center of the plot. Three chemotype groups were observed. On the PC1, all varieties of chemotypes III and II (hemp type) are separated from chemotype I

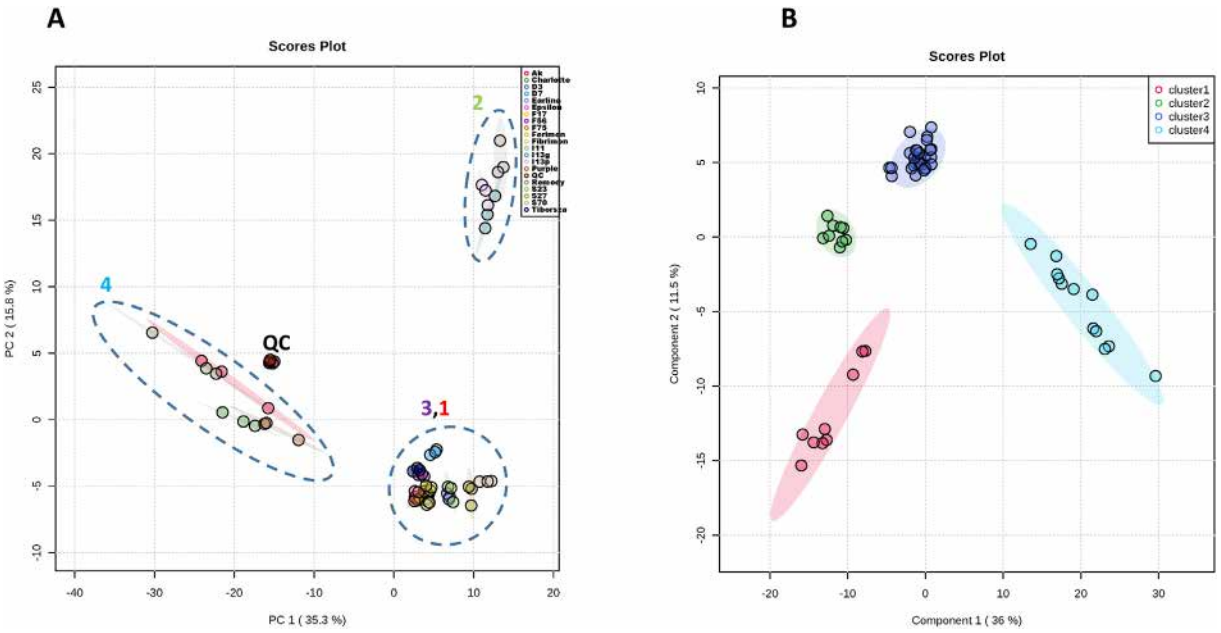


Fig. 3. Global PCA (**A**), PLSDA analyses for clusters (**B**) for *Cannabis* chemotypes. In **A**, numbers relate to the clusters obtained in Fig. 2-B.

(drug type). The differentiation between chemotypes III and II can be observed along PC2. According to the loading plot (data not shown), the separation in Factor 1 and 2 is not attributed to specific metabolite, but is a combined effect of a variety of metabolites. To validate the HCA model, we performed a supervised PLSDA analysis (Partial Least Squares Discriminant Analysis) based on the VIP (variable importance in projection) values. Overall, 47.5% of the total variance was displayed on the first two principal component axes of the PLSDA score plot (Fig. 3-B). Samples can be separated into four subgroups/clusters: samples of **cluster 4** were found distant of other chemotypes on the component 1, and the samples of **clusters 1**, **2**, and **3** as three independent clusters on component 2.

Following, we performed a hierarchical clustering analysis using T-test/Anova to find other characteristic markers differentiating these clusters. A heat map was generated from the top 50 metabolites present in varieties separated into clusters (Fig. 4). Based on the top 50 discriminant metabolites, this analysis results in the same four major clusters, as mentioned above. In these clusters, 24 metabolites out of 50 (48%) are phytocannabinoids; however, this type of compound is unexpectedly not discriminant for **cluster 4.** Samples of **cluster 4** (recreative varieties) have been genetically selected to present a characteristic aroma due mostly to mono- and sesquiterpenes (*e.g.* germacrene D-type, as observed in MN, **supp info 4**). In **cluster 1**, cannabigerol (CBG) and cannabigerolic acid (CBGA) derivatives are the characteristic markers, which is consistent with the fact that this cluster groups the *Santhica* varieties, characterized by a very low concentration of THC and CBD, and significant concentrations of CBG and CBGA [33,56]. However, in **cluster 1**, other distinctive compounds are put forward, *e.g.* CBE-type (produced by photo-oxidation of CBD and CBDA) and CBCA-type compounds. Contrary to prior knowledge stating that the fiber-type hemps are characterized by CBDA [4], **cluster 3**, which brings together ten fiber-type varieties, shows that two other cannabinoids were identified as discriminant biomarkers (Δ^{4-8}-THC and a CBLA derivative). In **cluster 2**, compounds with a 3-carbon resorcinyl side-chain (called propyl phytocannabinoids or *C-3* phytocannabinoids) namely CBDVA, CBCVA, Δ^{9}-THCV, and Δ^{1}-THCVA are presented as the main markers. However, in this cluster, a wide diversity of phytocannabinoids is present, as we can also observe *C-5* phytocannabinoids (namely CBL, CBC, CBR, CBG, CBN, and CBM types).

3.2. Unraveling characteristic markers of Cannabis varieties

To obtain specific markers for each variety within the clusters, we performed independent heatmap analyses and Anova tests. The top 25 most significant compounds are listed in Table 1 (heatmaps in **supp info 6-9**). Among the three varieties of **cluster 1** (*Santhica* varieties with CBG as major cannabinoid), S23 appears to be very different from the two other varieties (S27 and S70) (**supp info 6**). Of the over 25 compounds discriminant of S23, 14 are phytocannabinoids including Δ^{9}-THCA, Δ^{8}-THC, CBN, and their artifacts (CBN-1′S-hydroxy, CBNA and CBND). The CBN-type compounds result from the spontaneous oxidative aromatization of the THC-type derivatives [5]. S27 and S70 are mostly similar and share 3 markers: cannabispirenol B and two stilbenoid carexane derivatives. Discriminant markers of **cluster 2** (hemp varieties for seed production, characterized by C3 varin-type phytocannabinoids) suggest that these varieties are from different origins (**supp info 7**). Interestingly, a strigolactone derivative is identified as characteristic of two varieties. This can be related to their phenotypic aspect, which is short and much ramified. Strigolactones are phytohormones inhibiting shoot branching [57], and a change in their profile has an impact on ramification. The spiran cannabispirol derivatives, like cannabispirenone A previously isolated from a Thailand drug-type *Cannabis* variety [58,59] are characteristic of I11. I13g variety showed a dozen of discriminant compounds, including CBD, which h is a decarboxylated artifact of CBDA, and a cannabisol isomer, a dimer of Δ^{9}-THC with a methylene bride produced naturally in *Cannabis* plants [60]. In **cluster 3** (hemp-type varieties with CBD as major cannabinoid), 6 groups of varieties are distinguishable (**supp info 8**). As shown above by the genetic clustering (Fig. 2-A), D7 and Tiborszallasi varieties are very close, sharing seven discriminative compounds, among them CBDVA, cannabispirenone A, and a CBL derivative. A set of 6 compounds (3 of them being annotated as flavonoids, one as a fatty amide) is highly characteristic of Earlina. Three compounds are characteristic

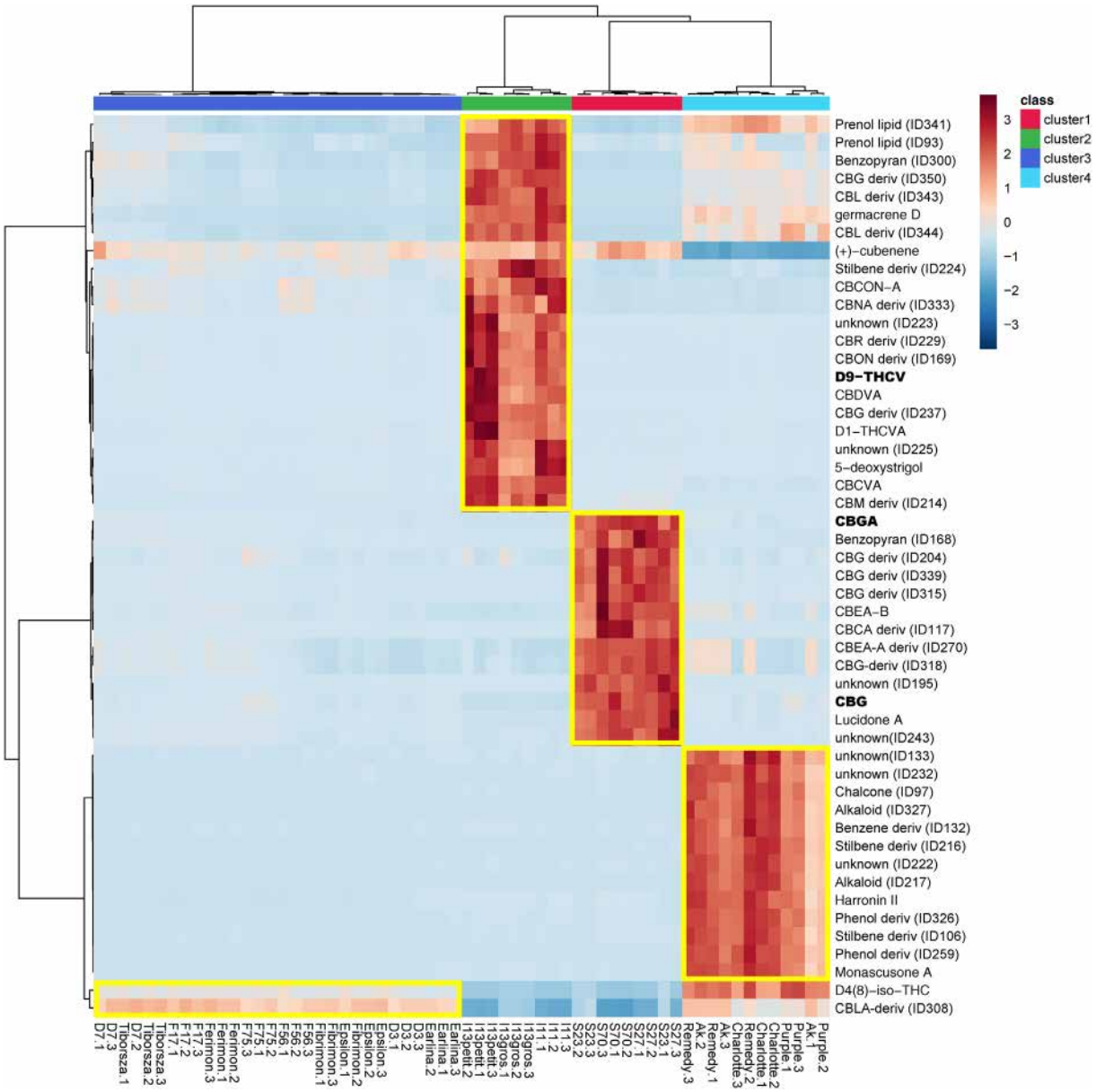


Fig. 4. Hierarchical clustering with the heat map generated from the top 50 metabolites present in all varieties of *Cannabis*. Clusters were grouped based on the HCA analysis showed in Fig. 2. Class of compounds were annotated using MSCleanR workflow. Phytocannabinoid standards are displayed in bold.

of Ferimon and F17 (including cannabistilbene I). Hildgardtol B and a phenanthrene derivative are characteristic of Epsilon. F75 is barely distinguished from other varieties but may be considered as having some similarities with Epsilon. Four compounds (one of them being annotated as the pulchelstyrene B, a cinnamylphenol) are characteristic of F56. Two stilbenoid derivatives and iso-cannabispiradienone appear as being discriminant for D3 and Fibrimon. These results can hardly be interpreted because chemical markers of these hemp varieties result from a massal selection applied to empirically selected populations. Phytocannabinoids can be discriminant for some varieties but not for all of them, excluding the acidic or neutral THC and CBD derivatives, which have been used as selection markers for a low content of these phytocannabinoids. In **cluster 4**, AK Silver and Remedy share the same distinctive compounds (among which Δ^9-THCA and CBGA) (**supp info 9**). Even if Charlotte and Purple variety share Δ^9-THC and CBD as characteristic markers, these compounds are more concentrated in Charlotte, principally Δ^9-THC (see Table 2-quantification).

3.3. Classification based on targeted phytocannabinoids quantification

Dozens of methods are used to quantify phytocannabinoids, based on many techniques: LC-HRMS [61], LC-DAD [62], HPTLC [63], GC-FID [31,64], GC-MS [65], Nuclear Magnetic Resonance (NMR) [66], and Triple quadrupole Multiple Reaction Monitoring (QqQ-MRM) [67], among others. Techniques and methods used in *Cannabis* quality control monographs differ between countries. Phytocannabinoids as a class of compounds are characterized by the occurrence of many isomers and the labile nature of the native acidic forms. Such complexity is well depicted in the molecular network of *Cannabis* extracts (Fig. 1). As a consequence, the quantification of phytocannabinoids is usually based only on the quantification of their major compounds. We quantified 11

Table 1
Main markers by cluster from *Cannabis* varieties based on heat map analyses.[a] = same annotation; [b] = $[M+Na]^+$; [c] = $[M + H-H_2O]^+$; [d] = $[M + H-2H_2O]^+$; [e] = $[M+K]^+$; [f] = standard; [g] = only in AK variety; [h] = only in Remedy variety; [i] = only in purple variety; [j] = only in charlotte variety. In bold: phytocannabinoids standards. MN = molecular networking.

ID	*m/z* $[M+H]^+$	RT	Formula	Compound annotated	Level of annotation	Classyfire (class)
Cluster 1						
S23						
183	311.201	11.70	$C_{21}H_{26}O_2$	**CBN** [f]	genus	cannabinoid
198	315.232	13.21	$C_{21}H_{30}O_2$	**D8-THC** [f]	genus	cannabinoid
226	327.196	10.19	$C_{21}H_{26}O_3$	Bauhinol B	genus	stilbene
272	341.211	20.69	$C_{22}H_{28}O_3$	CBLA-derivative	Based on MN	cannabinoid
294	355.191	13.63	$C_{22}H_{26}O_4$	CBNA	genus	cannabinoid
298	355.191	1.86	$C_{22}H_{26}O_4$	CBG-derivative	Based on MN	cannabinoid
317	359.222	11.63	$C_{22}H_{30}O_4$	Unknown cannabinoid	based on MN	cannabinoid
346	309.185[c]	11.81	$C_{21}H_{26}O_3$	CBN; 1′S-Hydroxy	genus	cannabinoid
375	383.222[c]	7.80	$C_{24}H_{32}O_5$	Lucidone B	generic	steroids and steroid derivative
S27/70						
177	311.164	3.61	$C_{20}H_{22}O_3$	Carexane P-isomer 1[a]	genus	stilbene
180	311.165	2.20	$C_{20}H_{22}O_3$	Carexane P-isomer 2[a]	genus	stilbene
206	243.102[d]	1.423	$C_{15}H_{18}O_5$	Cannabispirenol B derivative	genus	cannabinoid
Cluster 2						
I13g						
138	104.107	1.15	–	Unknown compound	–	–
139	315.232	8.953	$C_{21}H_{30}O_2$	**CBD** [f]	genus	cannabinoid
199	315.232	12.55	$C_{21}H_{30}O_2$	D4(8)-iso-THC	genus	cannabinoid
228	327.232	1.53	$C_{22}H_{30}O_2$	Cannabisol derivative	genus	cannabinoid
301	375.217	2.30	$C_{22}H_{30}O_5$	8,8′-Lignan-3,3′,4,4′,5-pentol-3,3′,4,5-Tetra-Me ether	genus	phenol derivative
I13p						
189	313.180[c]	11.12	$C_{20}H_{26}O_4$	CBDVA[a]	genus	cannabinoid
I11						
114	245.118	1.50	$C_{15}H_{16}O_3$	Cannabispirenone A	genus	cannabinoid
244	331.154	2.11	$C_{19}H_{22}O_5$	5-deoxystrigol	family	prenol lipid
Cluster 3						
Earlina						
58	609.271	13.90	–	Unknown compound	–	–
128	268.134	1.51	$C_{17}H_{17}NO_2$	(2E)-3-(4-hydroxyphenyl)-N-(2-phenylethyl)prop-2-enimidic acid	genus	phenol derivative
143	277.216	3.78	$C_{18}H_{28}O_2$	5-dodecylbenzene-1,3-diol-isomer 3[a]	family	phenol derivative
144	277.217	3.96	$C_{18}H_{28}O_2$	5-dodecylbenzene-1,3-diol-isomer 1[a]	family	phenol derivative
164	291.196	1.69	–	Unknown compound	–	–
218	322.275	6.84	$C_{20}H_{35}NO_2$	Alpha-Linolenoyl ethanolamide	generic	prenol lipid
F56						
129	271.097	2.04	$C_{16}H_{14}O_4$	2-(3,5-Dihydroxy-2,4-dimethylphenyl)-4-hydroxybenzofuran	genus	benzene and substituted derivative
208	317.139	3.67	$C_{18}H_{20}O_5$	Pulchelstyrene B	genus	flavonoid
284	345.206	6.50	$C_{21}H_{28}O_4$	Unknown compound	–	–
290	353.100[b]	2.56	$C_{18}H_{18}O_6$	2,3,5,6-tetramethoxy-9,10-dihydrophenanthrene-1,4-dione	genus	phenanthrene
D3/Fibrimon						
113	243.101	1.58	$C_{15}H_{14}O_3$	Isocannabispiradienone	genus	cannabinoid
177	311.164	3.61	$C_{20}H_{22}O_3$	Carexane P-isomer 1[a]	genus	stilbene
180	311.165	2.20	$C_{20}H_{22}O_3$	Carexane P-isomer 2[a]	genus	stilbene
Epsilon/F75						
154	287.128	2.84	$C_{17}H_{18}O_4$	3,4-Dihydro-7-(4-hydroxyphenyl)-2,2-dimethyl-2H-1-benzopyran-3,5-diol	genus	phenanthrene
263	339.160	4.82	$C_{21}H_{22}O_4$	Hildgardtol B	genus	flavonoid
Ferimon/F17						
191	313.180	2.54	$C_{20}H_{24}O_3$	Cannabistilbene I	genus	cannabinoid
Tiborszallasi/D7						
114	245.118	1.50	$C_{15}H_{16}O_3$	Cannabispirenone A	genus	cannabinoid
189	313.180[c]	11.12	$C_{20}H_{26}O_4$	CBDVA[a]	genus	cannabinoid
343	355.227[c]	14.85	$C_{23}H_{32}O_4$	CBL-Me ester	genus	cannabinoid
Cluster 4						
AK/Remedy						
82	198.091	2.97	$C_{13}H_{11}NO$	2-Hydroxy-3-methyl-9H-carbazole[h]	generic	carbazole alkaloid
130	287.092[e]	1.69	–	Tanacetols derivative	Based on MN	Prenol lipids
165	293.248	18.36	$C_{19}H_{32}O_2$	Grevillol [g]	genus	phenol derivative
280	343.227[c]	8.61	$C_{22}H_{32}O_4$	**CBGA** [f]	genus	cannabinoid
391	359.222	15.41	$C_{22}H_{30}O_4$	**D9-THCA** [f]	genus	cannabinoid
Purple/Charlotte						
139	315.232	8.953	$C_{21}H_{30}O_2$	**CBD** [f]	genus	cannabinoid
201	315.232	12.92	$C_{21}H_{30}O_2$	**D9-THC** [f,j]	genus	cannabinoid
227	327.232	2.04	$C_{22}H_{30}O_2$	Unknown compound [j]	–	–
241	329.212	2.91	$C_{21}H_{28}O_3$	CBON isomer [j]	genus	cannabinoid
281	343.227	3.2	$C_{22}H_{30}O_3$	CBG derivative [i]	Based on MN	cannabinoid

[a] = same annotation.

[b] $[M+Na]^+$.
[c] $[M + H-H_2O]^+$.
[d] $[M + H-2H_2O]^+$.
[e] $[M+K]^+$.
[f] standard.
[g] only in *AK* variety.
[h] only in *Remedy* variety.
[i] only in *purple* variety.
[j] = only in *charlotte* variety. In bold: phytocannabinoids standards. MN = molecular networking.

phytocannabinoids using LCMS as showed in Table 2 summarized (see also violin plots by clusters in supp info 10-20 and supp info 21 with Table 2 extended). Quantitative results match with markers identified in Fig. 4, notably for Δ^9-THCV (detected only in cluster 2) and CBG-CBGA mainly present in cluster 1 (supp info 14, 15). It is important to highlight that acidic phytocannabinoids are non-enzymatically decarboxylated into their corresponding neutral forms, which occur both within the plant and, to a much larger extent, upon heating after harvesting [68]. Therefore, acidic and neutral forms are related to a certain extent, depending on the temperature of drying. All samples studied were dried at 40°C, except for samples from cluster 4 which were dried at 80°C. Δ^9-THCA, the native form of Δ^9-THC, is elevated in cluster 2 (chemotype II, intermediate type, seed-producing varieties) and surprisingly still present in cluster 3 (chemotype III, fiber-type) (supp info 19). Unsupervised PLSDA based on the results of the quantification of 11 phytocannabinoids showed that clusters 2, 3, and 4 are grouped on component 2, and cluster 1 on component 1 (Fig. 5-A). 62.9% of the total variance was displayed on the first two principal component axes of the PLSDA score plot. To identify the most important metabolites allowing discrimination between the clusters, we performed a supervised PLSDA based on the variable importance in projection (VIP) values. Thus, a metabolite with a VIP >1 is regarded as significantly discriminant. CBD, CBDA, and CBGA are the overall main discriminant quantified phytocannabinoids (Fig. 5-B). Some varieties presented as drug-type can be assimilated to fiber-types, *e.g. AK* and *Remedy* samples (Fig. 5 A and **C**), although their phytocannabinoids content is higher (Table 2). Such drug-type varieties have been selected to produce a significant amount of CBD but only moderate amounts of THC in order to be considered as "light cannabis" in some countries (close to 0,2% Δ^9-THC), thus these recreative varieties are barely distinguished from hemp varieties when clustering is based only on main phytocannabinoids. Hemp varieties have been specifically selected to have phytocannabinoids biosynthetic machinery, which is significantly inhibited. Samples of **cluster 2** stand out because of their specific expression of Δ^9-THCV, and also because of a significant concentration of Δ^9-THCA. As observed above, such varieties are used for seed production outside EU (mostly in China) but have not been subject to an efficient selection to reduce Δ^9-THC content. As expected, **cluster 1** (hemp varieties with CBG as major cannabinoid) are characterized by CBG and CBGA.

An overall analysis of the data shows that *Cannabis* varieties are more efficiently discriminated and characterized by untargeted cannabinomics analysis (metabotyping) rather than quantitation of major phytocannabinoids. In this sense, the performance measure (Q2 and R2) were compared between metabolomic and quantification models on PLSDA analysis, suggesting that the metabotyping represents a most robust model (0.39 vs. 0.77, 0.26 vs. 0.75 on comp1 and 0.50 vs. 0.96, 0.45 vs. 0.94 on comp2 respectively).

4. Conclusions

Literature on *Cannabis* is expanding quickly but is still mostly focused on their pharmacologically active compounds Δ^9-THC and CBD. The metabolomics approach is a powerful tool to unravel the complexities of the actual entire metabolome. Cannabinomics must therefore be understood as an approach including the analyses not only of phytocannabinoids but also other compounds including terpenes, stilbenes, flavonoids, etc. In this work, cannabinomics contributed to deciphering the main discriminant metabolites in 20 *Cannabis* varieties. Among the 13 fiber-type varieties, we identified characteristic markers allowing an efficient differentiation beyond phytocannabinoids only. Varieties of the intermediate chemotype used for seed production ("I" varieties) are identified as mostly producing *C-3* phytocannabinoids. This workflow can be used as a complementary tool to be used along with genotyping with the aim to differentiate *C. sativa* varieties and to allow the identification of

Table 2
UHPLC-MS quantification of 11 phytocannabinoids from 20 *Cannabis* varieties. Results were reported as % in powder, n = 3. For details and validation data, see supp info 21.

	Δ^9-THC	Δ^9-THCA	CBD	CBDA	Δ^8-THC	CBG	CBGA	CBC	CBCA	Δ^9-THCV	CBN
AK Silver	0.006	1.603	4.270	14.535	0.519	0.169	1.116	0.344	1.964	<LOD	0.050
Charlotte	0.047	0.106	7.248	7.380	0.470	0.050	0.085	0.594	0.616	<LOQ	0.099
Purple	0.036	0.367	7.046	9.387	0.854	0.260	0.289	0.781	1.283	<LOD	0.067
Remedy	0.007	2.015	4.848	15.100	0.659	0.256	1.313	0.444	2.237	<LOD	0.053
D3	0.002	0.165	1.875	5.104	0.130	0.052	0.103	0.144	0.415	<LOD	0.028
D7	0.002	0.908	1.186	7.926	0.254	0.088	0.656	0.106	0.812	<LOD	<LOQ
Earlina	0.003	0.089	1.458	3.487	0.089	0.064	0.146	0.151	0.298	<LOD	<LOQ
Epsilon	0.003	0.240	2.090	6.282	0.174	0.071	0.231	0.172	0.562	<LOD	0.027
Ferimon	0.002	0.297	1.414	7.203	0.136	0.089	0.456	0.130	0.678	<LOD	<LOQ
Fibrimon	0.003	0.212	1.607	5.524	0.128	0.050	0.151	0.195	0.595	<LOD	0.024
F17	0.002	0.224	1.482	6.106	0.142	0.092	0.441	0.165	0.659	<LOD	<LOQ
F56	0.003	0.793	1.863	5.334	0.406	0.069	0.280	0.155	0.485	<LOD	0.078
F75	0.004	0.197	2.188	5.643	0.148	0.147	0.448	0.202	0.502	<LOD	0.030
I11	<LOQ	2.686	0.040	0.411	0.445	<LOQ	0.056	0.035	0.419	0.242	<LOQ
I13gros	<LOQ	1.366	0.091	1.126	0.300	<LOQ	0.040	0.034	0.364	0.165	<LOQ
I13petit	<LOQ	2.408	0.033	0.458	0.365	<LOQ	0.057	<LOQ	0.411	0.339	<LOQ
S23	0.002	0.070	0.421	2.221	0.054	0.825	6.599	0.126	0.287	<LOD	<LOD
S27	0.003	0.016	0.085	0.568	0.008	0.601	6.185	0.202	0.208	<LOD	<LOD
S70	0.002	<LOQ	0.039	0.155	0.003	0.382	3.945	0.094	0.091	<LOD	<LOD
Tiborszallasi	0.001	0.628	1.121	5.834	0.285	0.077	0.490	0.094	0.520	<LOD	<LOQ

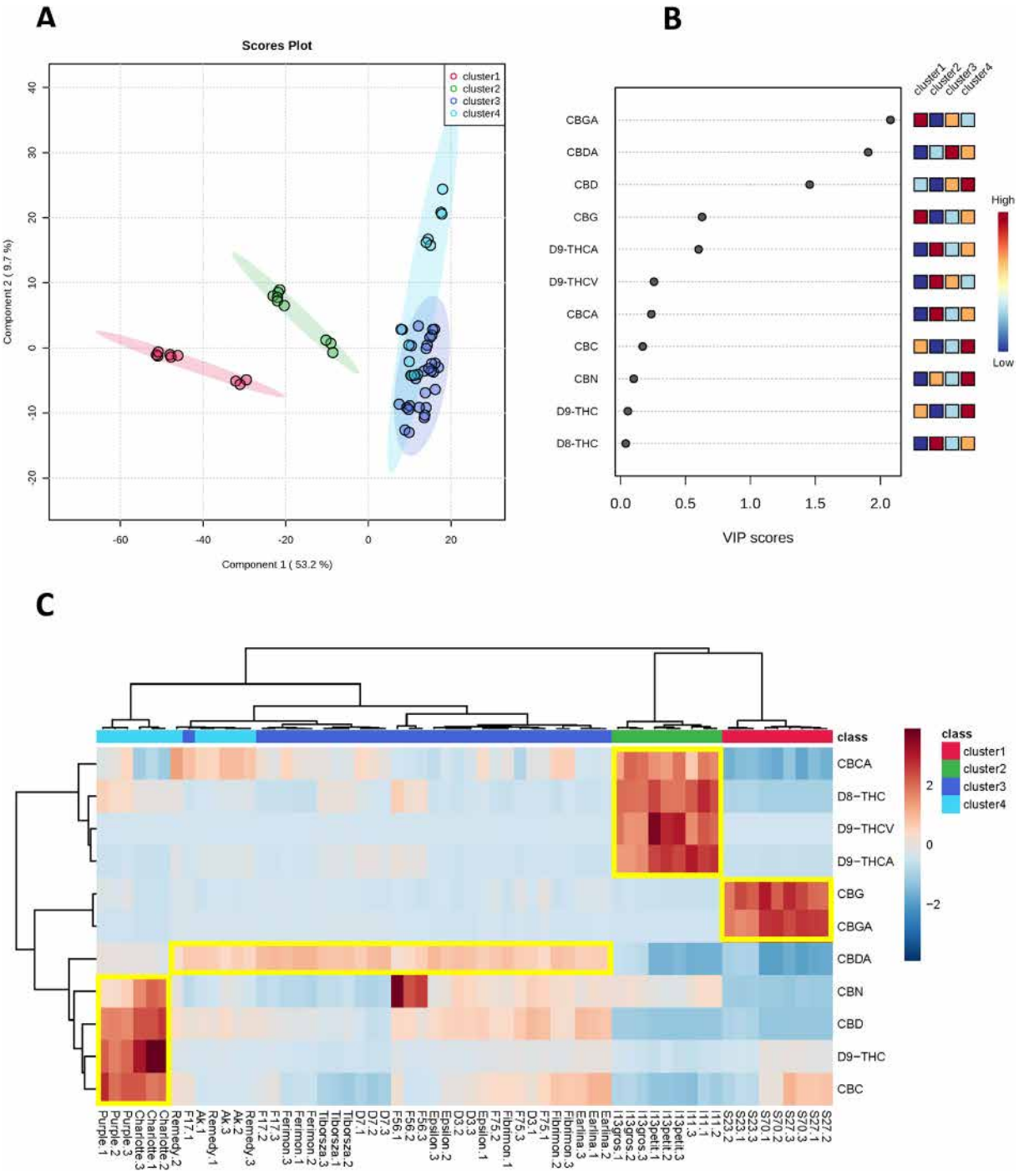


Fig. 5. Statistical analyses based of quantitative data on 11 phytocannabinoids: PLSDA based on average concentration (ng/mL) for each phytocannabinoid by cluster (obtained in Fig. 2-C) (**A**); PLSDA-VIP projection by cluster (**B**); hierarchical clustering analysis with the heat-map for phytocannabinoid standards by clusters (**C**).

minor compounds playing a key role as markers of differentiation. This workflow could offer a shorter road to creating chemotypes and phenotypes that meet the demand of production needs for material, food, and medicinal purposes. In the medicinal field, cannabinomics could facilitate improved breeding to create *Cannabis* varieties with greater expression of minor phytocannabinoids of medicinal value. Cannabinomics should also be used as a systematic basis for studies correlating chemical profile and therapeutical effects, as it allows to properly address the entourage effect.

Author's contribution

Conceptualization: PGV-O, AM and GM; methodology: PGV-O, AM, GM, SF, and SB; chemical data acquisition/curation: PGVO and GM; biological data curation: MB and FM; investigation: PGV-O, GM, FM, MB, SF, SB, and AM. PGV-O wrote the original draft and all the authors contributed to the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to acknowledge HEMP-it ADN for the post-doctoral fellowship of PGV-O (CNRS UMR 8076 BioCIS). PGV-O was also supported by a postdoctoral fellowship (contract number 04077858) from the French National Research Institute for Sustainable Development (IRD), UMR 152 PHARMADEV, IRD-UPS, Toulouse, France. The authors are grateful to Bruno Figadère (CNRS UMR 8076 BioCIS, Châtenay-Malabry); Justine Chervin and Virginie Puech-Pagès (MetaboHUB-MetaToul-Agromix, LRSV, Auzeville-Tolosane); Guillaume Cabanac (IRIT, Toulouse); Claire Thouminot and Christophe Fevrier (HEMP-it ADN, Beaufort-en-Anjou). The authors are also grateful to Alice Gadea (UMR 152 PHARMADEV, IRD-UPS, Toulouse, France), Alice Kan, and Joshua Lee Halford (Department of Biochemistry, UCR, Riverside, USA) for their important critical discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2021.339020.

References

[1] G. Ren, X. Zhang, Y. Li, K. Ridout, M.L. Serrano-Serrano, Y. Yang, A. Liu, G. Ravikanth, M.A. Nawaz, A.S. Mumtaz, N. Salamin, L. Fumagalli, Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*, Sci. Adv. 7 (2021), eabg2286, https://doi.org/10.1126/sciadv.abg2286.

[2] S.A. Bonini, M. Premoli, S. Tambaro, A. Kumar, G. Maccarinelli, M. Memo, A. Mastinu, *Cannabis sativa*: a comprehensive ethnopharmacological review of a medicinal plant with a long history, J. Ethnopharmacol. 227 (2018) 300–315, https://doi.org/10.1016/j.jep.2018.09.004.

[3] M. Ren, Z. Tang, X. Wu, R. Spengler, H. Jiang, Y. Yang, N. Boivin, The origins of cannabis smoking: chemical residue evidence from the first millennium BCE in the Pamirs, Sci. Adv. 5 (2019), https://doi.org/10.1126/sciadv.aaw1391 eaaw1391.

[4] C.M. Andre, J.-F. Hausman, G. Guerriero, *Cannabis sativa*: the plant of the thousand and one molecules, Front. Plant Sci. 7 (2016), https://doi.org/10.3389/fpls.2016.00019.

[5] L.O. Hanuš, S.M. Meyer, E. Muñoz, O. Taglialatela-Scafati, G. Appendino, Phytocannabinoids: a unified critical inventory, Nat. Prod. Rep. 33 (2016) 1357–1392, https://doi.org/10.1039/C6NP00074F.

[6] C. Citti, P. Linciano, F. Russo, L. Luongo, M. Iannotta, S. Maione, A. Laganà, A.L. Capriotti, F. Forni, M.A. Vandelli, G. Gigli, G. Cannazza, A novel phytocannabinoid isolated from *Cannabis sativa* L. with an *in vivo* cannabimimetic activity higher than Δ9-tetrahydrocannabinol: Δ9-Tetrahydrocannabiphorol, Sci. Rep. 9 (2019) 20335, https://doi.org/10.1038/s41598-019-56785-1.

[7] T. Gülck, B.L. Møller, Phytocannabinoids: origins and biosynthesis, Trends Plant Sci. 25 (2020) 985–1004, https://doi.org/10.1016/j.tplants.2020.05.005.

[8] X. Luo, M.A. Reiter, L. d'Espaux, J. Wong, C.M. Denby, A. Lechner, Y. Zhang, A.T. Grzybowski, S. Harth, W. Lin, H. Lee, C. Yu, J. Shin, K. Deng, V.T. Benites, G. Wang, E.E.K. Baidoo, Y. Chen, I. Dev, C.J. Petzold, J.D. Keasling, Complete biosynthesis of cannabinoids and their unnatural analogues in yeast, Nature 567 (2019) 123–126, https://doi.org/10.1038/s41586-019-0978-9.

[9] F.E. Dussy, C. Hamberg, M. Luginbühl, T. Schwerzmann, T.A. Briellmann, Isolation of Δ9-THCA-A from hemp and analytical aspects concerning the determination of Δ9-THC in cannabis products, Forensic Sci. Int. 149 (2005) 3–10, https://doi.org/10.1016/j.forsciint.2004.05.015.

[10] J.M. McPartland, C. MacDonald, M. Young, P.S. Grant, D.P. Furkert, M. Glass, Affinity and efficacy studies of tetrahydrocannabinolic acid A at cannabinoid receptor types one and two, Cannabis Cannabinoid Res. 2 (2017) 87–95, https://doi.org/10.1089/can.2016.0032.

[11] R. Mechoulam, Y. Gaoni, Recent advances in the chemistry of hashish, Fortschritte Chem. Org. Naturstoffe Prog. Chem. Org. Nat. Prod. Progres Dans Chim. Subst. Org. Nat. 25 (1967) 175–213, https://doi.org/10.1007/978-3-7091-8164-5_6.

[12] R. Mechoulam, N.K. McCallum, S. Burstein, Recent advances in the chemistry and biochemistry of *cannabis*, Chem. Rev. 76 (1976) 75–112, https://doi.org/10.1021/cr60299a002.

[13] C.E. Turner, M.A. Elsohly, E.G. Boeren, Constituents of *Cannabis sativa* L. XVII. A review of the natural constituents, J. Nat. Prod. 43 (1980) 169–234, https://doi.org/10.1021/np50008a001.

[14] C. Citti, P. Linciano, F. Forni, M.A. Vandelli, G. Gigli, A. Laganà, G. Cannazza, Analysis of impurities of cannabidiol from hemp. Isolation, characterization and synthesis of cannabidibutol, the novel cannabidiol butyl analog, J. Pharmaceut. Biomed. Anal. 175 (2019) 112752, https://doi.org/10.1016/j.jpba.2019.06.049.

[15] F. Pollastro, O. Taglialatela-Scafati, M. Allarà, E. Muñoz, V. Di Marzo, L. De Petrocellis, G. Appendino, Bioactive prenylogous cannabinoid from fiber hemp (*Cannabis sativa*), J. Nat. Prod. 74 (2011) 2019–2022, https://doi.org/10.1021/np200500p.

[16] M.M. Radwan, M.A. ElSohly, A.T. El-Alfy, S.A. Ahmed, D. Slade, A.S. Husni, S.P. Manly, L. Wilson, S. Seale, S.J. Cutler, S.A. Ross, Isolation and pharmacological evaluation of minor cannabinoids from high-potency *Cannabis sativa*, J. Nat. Prod. 78 (2015) 1271–1276, https://doi.org/10.1021/acs.jnatprod.5b00065.

[17] P.S. Fetterman, E.S. Keith, C.W. Waller, O. Guerrero, N.J. Doorenbos, M.W. Quimby, Mississippi-grown *Cannabis sativa* L.: preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part, J. Pharmaceut. Sci. 60 (1971) 1246–1249, https://doi.org/10.1002/jps.2600600832.

[18] E. Small, H.D. Beckstead, Common cannabinoid phenotypes in 350 stocks of Cannabis, Lloydia 36 (1973) 144–165.

[19] E. Small, H.D. Beckstead, Cannabinoid phenotypes in cannabis sativa, Nature 245 (1973) 147–148, https://doi.org/10.1038/245147a0.

[20] J.T. Fischedick, Identification of terpenoid chemotypes among high (−)-trans-Δ9- tetrahydrocannabinol-producing *Cannabis sativa* L. Cultivars, Cannabis Cannabinoid Res. 2 (2017) 34–47, https://doi.org/10.1089/can.2016.0040.

[21] A. Hazekamp, J.T. Fischedick, *Cannabis* - from cultivar to chemovar, Drug Test. Anal. 4 (2012) 660–667, https://doi.org/10.1002/dta.407.

[22] A. Hazekamp, K. Tejkalová, S. Papadimitriou, Cannabis: from cultivar to chemovar II—a metabolomics approach to *cannabis* classification, Cannabis Cannabinoid Res. 1 (2016) 202–215, https://doi.org/10.1089/can.2016.0017.

[23] A.D. Kinghorn, H. Falk, S. Gibbons, J. Kobayashi (Eds.), Phytocannabinoids: Unraveling the Complex Chemistry and Pharmacology of *Cannabis sativa*, Springer International Publishing, 2017, https://doi.org/10.1007/978-3-319-45541-9.

[24] United Nations, UN Commission Reclassifies Cannabis, yet Still Considered Harmful, UN News, 2020. https://news.un.org/en/story/2020/12/1079132. (Accessed 6 May 2021).

[25] T.P. Freeman, C. Hindocha, S.F. Green, M.A.P. Bloomfield, Medicinal use of cannabis based products and cannabinoids, BMJ 365 (2019) l1141, https://doi.org/10.1136/bmj.l1141.

[26] N.M. Kogan, R. Mechoulam, Cannabinoids in health and disease, Dialogues Clin. Neurosci. 9 (2007) 413–430.

[27] S.T. Lukhele, L.R. Motadi, Cannabidiol rather than *Cannabis sativa* extracts inhibit cell growth and induce apoptosis in cervical cancer cells, BMC Compl. Alternative Med. 16 (2016), https://doi.org/10.1186/s12906-016-1280-0.

[28] E.T. Wargent, M.S. Zaibi, C. Silvestri, D.C. Hislop, C.J. Stocker, C.G. Stott, G.W. Guy, M. Duncan, V. Di Marzo, M.A. Cawthorne, The cannabinoid Δ9-tetrahydrocannabivarin (THCV) ameliorates insulin sensitivity in two mouse models of obesity, Nutr. Diabetes 3 (2013) e68, https://doi.org/10.1038/nutd.2013.9.

[29] G. Watt, T. Karl, In vivo evidence for therapeutic properties of cannabidiol (CBD) for alzheimer's disease, Front. Pharmacol. 8 (2017), https://doi.org/10.3389/fphar.2017.00020.

[30] G. Fournier, Les chimiotypes du chanvre (*Cannabis sativa* L.) Intérêt pour un programme de sélection, Agronomie 1 (1981) 679–688, https://doi.org/10.1051/agro:19810809.

[31] G. Fournier, A propos de la teneur en Δ-9-THC dans les variétés de chanvre à fibres cultivées en France, Ann. Toxicol. Anal. 15 (2003) 30–34, https://doi.org/10.1051/ata/2003032.

[32] G. Fournier, O. Beherec, S. Bertucelli, Intérêt du rapport Δ-9-THC/CBD dans le contrôle des cultures de chanvre industriel, Ann. Toxicol. Anal. 15 (2003) 250–259, https://doi.org/10.1051/ata/2003003.

[33] G. Fournier, O. Beherec, S. Bertucelli, Santhica 23 et 27 : deux variétés de chanvre (*Cannabis sativa* L.) sans Δ-9-THC, Ann. Toxicol. Anal. 16 (2004) 128–132, https://doi.org/10.1051/ata/2004017.

[34] G. Fournier, J. Bausset, O. Beherec, M. Desvals, S. Bertucelli, La simplification du contrôle des cultures de chanvre industriel est possible, Ann. Toxicol. Anal. 19 (2007) 201–209, https://doi.org/10.1051/ata:2007023.

[35] O. Fraisier-Vannier, J. Chervin, G. Cabanac, V. Puech, S. Fournier, V. Durand, A. Amiel, O. André, O.A. Benamar, B. Dumas, H. Tsugawa, G. Marti, MS-CleanR, A feature-filtering workflow for untargeted LC–MS based metabolomics, Anal. Chem. 92 (2020) 9971–9981, https://doi.org/10.1021/acs.analchem.0c01594.

[36] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, Nat. Methods 12 (2015) 523–526, https://doi.org/10.1038/nmeth.3393.

[37] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software, Anal. Chem. 88 (2016) 7946–7958, https://doi.org/10.1021/acs.analchem.6b00770.

[38] F. Olivon, N. Elie, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MetGem

Software for the generation of molecular networks based on the t-SNE algorithm, Anal. Chem. 90 (2018) 13900–13908, https://doi.org/10.1021/acs.analchem.8b03099.
[39] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.
[40] H. Al-Ghanim, Development of microsatellite markers in *Cannabis sativa* for fingerprinting and genetic relatedness analyses, FIU Electron. Theses Diss. https://doi.org/10.25148/etd.FI13101576, 2003.
[41] S. Gilmore, R. Peakall, Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana), Mol. Ecol. Notes 3 (2003) 105–107, https://doi.org/10.1046/j.1471-8286.2003.00367.x.
[42] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinf. 11 (2010) 395, https://doi.org/10.1186/1471-2105-11-395, 395.
[43] Y. Ni, M. Su, Y. Qiu, W. Jia, X. Du, ADAP-GC 3.0: improved peak detection and deconvolution of Co-eluting metabolites from GC/TOF-MS data for metabolomics studies, Anal. Chem. 88 (2016) 8802–8811, https://doi.org/10.1021/acs.analchem.6b02222.
[44] ICH, Validation of Analytical Procedures: Text and Methodology, International Conference on Harmonization, 2005.
[45] J. Chong, D.S. Wishart, J. Xia, Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis, Curr. Protoc. Bioinforma. 68 (2019) e86, https://doi.org/10.1002/cpbi.86.
[46] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[47] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software structure: a simulation study, Mol. Ecol. 14 (2005) 2611–2620, https://doi.org/10.1111/j.1365-294X.2005.02553.x.
[48] A. Rashid, V. Ali, M. Khajuria, S. Faiz, S. Gairola, D. Vyas, GC–MS based metabolomic approach to understand nutraceutical potential of Cannabis seeds from two different environments, Food Chem. 339 (2021) 128076, https://doi.org/10.1016/j.foodchem.2020.128076.
[49] C.M. Montone, A. Cerrato, B. Botta, G. Cannazza, A.L. Capriotti, C. Cavaliere, C. Citti, F. Ghirga, S. Piovesana, A. Laganà, Improved identification of phytocannabinoids using a dedicated structure-based workflow, Talanta 219 (2020) 121310, https://doi.org/10.1016/j.talanta.2020.121310.
[50] K.A. Aliferis, D. Bernard-Perron, Cannabinomics: application of metabolomics in *cannabis* (*Cannabis sativa* L.) Research and development, Front. Plant Sci. 11 (2020), https://doi.org/10.3389/fpls.2020.00554.
[51] P. Berman, L. Sulimani, A. Gelfand, K. Amsalem, G.M. Lewitus, D. Meiri, Cannabinoidomics – an analytical approach to understand the effect of medical *Cannabis* treatment on the endocannabinoid metabolome, Talanta 219 (2020) 121336, https://doi.org/10.1016/j.talanta.2020.121336.
[52] A. Cerrato, C. Citti, G. Cannazza, A.L. Capriotti, C. Cavaliere, G. Grassi, F. Marini, C.M. Montone, R. Paris, S. Piovesana, A. Laganà, Phytocannabinomics: untargeted metabolomics as a tool for *cannabis* chemovar differentiation, Talanta (2021) 122313, https://doi.org/10.1016/j.talanta.2021.122313.
[53] D.J. Creek, W.B. Dunn, O. Fiehn, J.L. Griffin, R.D. Hall, Z. Lei, R. Mistrik, S. Neumann, E.L. Schymanski, L.W. Sumner, R. Trengove, J.-L. Wolfender, Metabolite identification: are you sure? And how do your peers gauge your confidence? Metabolomics 10 (2014) 350–353, https://doi.org/10.1007/s11306-014-0656-8.
[54] P. Berman, K. Futoran, G.M. Lewitus, D. Mukha, M. Benami, T. Shlomi, D. Meiri, A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in *Cannabis*, Sci. Rep. 8 (2018) 14280, https://doi.org/10.1038/s41598-018-32651-4.
[55] Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner, D.S. Wishart, ClassyFire: automated chemical classification with a comprehensive, computable taxonomy, J. Cheminf. 8 (2016) 61, https://doi.org/10.1186/s13321-016-0174-y.
[56] J. Baranauskaite, M. Marksa, L. Ivanauskas, K. Vitkevicius, M. Liaudanskas, V. Skyrius, A. Baranauskas, Development of extraction technique and GC/FID method for the analysis of cannabinoids in *Cannabis sativa* L. spp. santicha (hemp), Phytochem. Anal. 31 (2020) 516–521, https://doi.org/10.1002/pca.2915.
[57] F.-D. Boyer, A. de Saint Germain, J.-B. Pouvreau, G. Clavé, J.-P. Pillot, A. Roux, A. Rasmussen, S. Depuydt, D. Lauressergues, N. Frei dit Frey, T.S.A. Heugebaert, C.V. Stevens, D. Geelen, S. Goormachtig, C. Rameau, New strigolactone analogs as plant hormones with low activities in the rhizosphere, Mol. Plant 7 (2014) 675–690, https://doi.org/10.1093/mp/sst163.
[58] L. Crombie, W.M.L. Crombie, Natural products of Thailand high Δ1-THC-strain *Cannabis*. The bibenzyl-spiran-dihydrophenanthrene group: relations with cannabinoids and canniflavones, J. Chem. Soc. Perkin 1 (1982) 1455–1466, https://doi.org/10.1039/P19820001455.
[59] L. Crombie, W. Mary, L. Crombie, S.V. Jamieson, Isolation of cannabispiradienone and cannabidihydrophenanthrene. biosynthetic relationships between the spirans and dihydrostilbenes of Thailand *Cannabis*, Tetrahedron Lett. 20 (1979) 661–664, https://doi.org/10.1016/S0040-4039(01)86030-5.
[60] F. Zulfiqar, S.A. Ross, D. Slade, S.A. Ahmed, M.M. Radwan, Z. Ali, I.A. Khan, M.A. ElSohly, Cannabisol, a novel Δ9-THC dimer possessing a unique methylene bridge, isolated from *Cannabis sativa*, Tetrahedron Lett. 53 (2012) 3560–3562, https://doi.org/10.1016/j.tetlet.2012.04.139.
[61] C. Citti, G. Ciccarella, D. Braghiroli, C. Parenti, M.A. Vandelli, G. Cannazza, Medicinal *cannabis*: principal cannabinoids concentration and their stability evaluated by a high performance liquid chromatography coupled to diode array and quadrupole time of flight mass spectrometry method, J. Pharmaceut. Biomed. Anal. 128 (2016) 201–209, https://doi.org/10.1016/j.jpba.2016.05.033.
[62] L.A. Ciolino, T.L. Ranieri, A.M. Taylor, Commercial cannabis consumer products part 2: HPLC-DAD quantitative analysis of cannabis cannabinoids, Forensic Sci. Int. 289 (2018) 438–447, https://doi.org/10.1016/j.forsciint.2018.05.033.
[63] J.T. Fischedick, R. Glas, A. Hazekamp, R. Verpoorte, A qualitative and quantitative HPTLC densitometry method for the analysis of cannabinoids in *Cannabis sativa* L, Phytochem. Anal. 20 (2009) 421–426, https://doi.org/10.1002/pca.1143.
[64] D. Wianowska, A.L. Dawidowicz, M. Kowalczyk, Transformations of Tetrahydrocannabinol, tetrahydrocannabinolic acid and cannabinol during their extraction from *Cannabis sativa* L, J. Anal. Chem. 70 (2015) 920–925, https://doi.org/10.1134/S1061934815080183.
[65] M.M. Delgado-Povedano, C. Sánchez-Carnerero Callado, F. Priego-Capote, C. Ferreiro-Vera, Untargeted characterization of extracts from *Cannabis sativa* L. cultivars by gas and liquid chromatography coupled to mass spectrometry in high resolution mode, Talanta 208 (2020) 120384, https://doi.org/10.1016/j.talanta.2019.120384.
[66] A. Hazekamp, Y.H. Choi, R. Verpoorte, Quantitative analysis of cannabinoids from *Cannabis sativa* using 1H-NMR, Chem. Pharm. Bull. (Tokyo) 52 (2004) 718–721, https://doi.org/10.1248/cpb.52.718.
[67] M. Andersson, K.B. Scheidweiler, C. Sempio, A. Barnes, M.A. Huestis, Simultaneous quantification of eleven cannabinoids and metabolites in human urine by liquid chromatography tandem mass spectrometry using WAX-S tips, Anal. Bioanal. Chem. 408 (2016) 6461–6471, https://doi.org/10.1007/s00216-016-9765-8.
[68] I.J. Flores-Sanchez, R. Verpoorte, Secondary metabolism in *cannabis*, Phytochemistry Rev. 7 (2008) 615–639, https://doi.org/10.1007/s11101-008-9094-4.