Analytica Chimica Acta 1278 (2023) 341716

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Untargeted cannabinomics reveals the chemical differentiation of industrial hemp based on the cultivar and the geographical field location

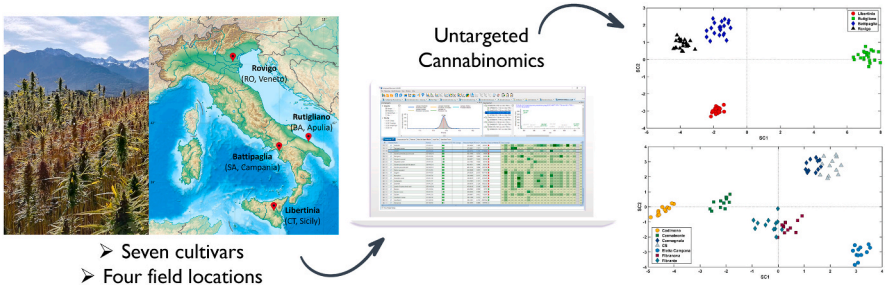
Andrea Cerrato [a], Alessandra Biancolillo [b], Giuseppe Cannazza [c,d], Chiara Cavaliere [a], Cinzia Citti [c,d], Aldo Laganà [a], Federico Marini [a], Massimo Montanari [e], Carmela Maria Montone [a], Roberta Paris [e], Nino Virzì [f], Anna Laura Capriotti [a,*]

[a] Department of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185, Rome, Italy
[b] Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, 67100, L'Aquila, Coppito, Italy
[c] Department of Life Sciences, University of Modena and Reggio Emilia, Via Giuseppe Campi 287, 41125, Modena, Italy
[d] CNR NANOTEC, Campus Ecotekne, University of Salento, Via Monteroni, 73100, Lecce, Italy
[e] CREA-Research Center for Cereal and Industrial Crops, Via di Corticella 133, 40128, Bologna, Italy
[f] CREA-Research Center for Cereal and Industrial Crops, C.so Savoia 190, 95024, Acireale, CT, Italy

HIGHLIGHTS

- Untargeted HRMS-based cannabinomics of 84 samples of industrial hemp.
- Seven cultivars of industrial hemp from four experimental fields were analyzed.
- 54 phytocannabinoids, 134 flavonoids, and 77 phenolic compounds were annotated.
- ASCA models for the individual compound classes as well as a low-level fusel model.
- Phenolic compounds play a decisive role in discriminating industrial hemp samples.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling editor: L. Liang

Keywords:
Cannabis sativa
High-resolution mass spectrometry
Metabolomics
ANOVA-Simultaneous component analysis
Phytocannabinoids

ABSTRACT

Cannabis sativa has long been harvested for industrial applications related to its fibers. Industrial hemp cultivars, a botanical class of *Cannabis sativa* with a low expression of intoxicating Δ^9-tetrahydrocannabinol (Δ^9-THC) have been selected for these purposes and scarcely investigated in terms of their content in bioactive compounds. Following the global relaxation in the market of industrial hemp-derived products, research in industrial hemp for pharmaceutical and nutraceutical purposes has surged. In this context, metabolomics-based approaches have proven to fulfill the aim of obtaining comprehensive information on the phytocompound profile of cannabis samples, going beyond the targeted evaluation of the major phytocannabinoids. In the present paper, an HRMS-based metabolomics study was addressed to seven distinct industrial hemp cultivars grown in four experimental fields in Northern, Southern, and Insular Italy. Since the role of minor phytocannabinoids as well as other phytocompounds was found to be critical in discriminating cannabis chemovars and in determining its biological activities, a comprehensive characterization of phytocannabinoids, flavonoids, and phenolic acids was carried out by LC-HRMS and a dedicated data processing workflow following the guidelines of the metabolomics Quality Assurance and Quality Control Consortium. A total of 54 phytocannabinoids, 134 flavonoids, and 77 phenolic

* Corresponding author.
E-mail address: annalaura.capriotti@uniroma1.it (A.L. Capriotti).

https://doi.org/10.1016/j.aca.2023.341716
Received 20 April 2023; Received in revised form 13 July 2023; Accepted 13 August 2023
Available online 22 August 2023

acids were annotated, and their role in distinguishing hemp samples based on the geographical field location and cultivar was evaluated by ANOVA-simultaneous component analysis. Finally, a low-level fused model demonstrated the key role of untargeted cannabinomics extended to lesser-studied phytocompound classes for the discrimination of hemp samples.

1. Introduction

Cannabis sativa L. is a multi-purpose crop that has found several applications, despite over a century of global prohibition, for its generally sustainable and low-input cultivation [1]. Other than being employed as a source of fiber, seeds, and oil, *C. sativa* has the peculiar ability to synthesize phytocannabinoids, terpenophenolic compounds that probably serve as a defense against predators [2]. Among these, the most notorious compound is Δ^9-tetrahydrocannabinolic acid (Δ^9-THCA), which has the capability, after decarboxylation, to bind cannabinoid receptors 1 and 2 (CB_1 and CB_2, respectively), causing the well-known intoxicating psychoactive effects but also several pharmacological activities [3,4]. Among the major constituents of *C. sativa*, non-psychoactive cannabidiolic acid (CBDA) and cannabidiol (CBD) have also gained significant scientific interest for their anti-inflammatory, anti-epileptic, and pain-relief properties [5,6]. Δ^9-THCA and CBDA are both enzymatically obtained from a common phytocannabinoid progenitor, the cannabigerolic acid (CBGA), that is produced by a type III polyketide synthase [7] (see Table 1).

For its remarkably genetic plasticity, *C. sativa* has been bred all over the world to create new strains and varieties, e.g., genotypes enriched in a single phytocannabinoid or more resilient to specific climates [8]. Historically, *C. sativa* plants have been classified into three chemotypes (sometimes also referred to as chemovars) based on the relative amounts of Δ^9-THC and CBD: chemotype I (Δ^9-THC > CBD), chemotype II (Δ^9-THC and CBD at similar concentrations, Δ^9-THC >0.3%), also indicated with the vernacular name "drug-type", and chemotype III (Δ^9-THC < CBD, Δ^9-THC <0.3%), also called "fiber-type" based on their main utilization [9]. Later, other chemotypes have been described, including a cannabigerol (CBG) predominant one (chemotype IV) [10] and chemotype V with no cannabinoids [11].

Despite most studies being still focused on the pharmacologically active Δ^9-THC and CBD and a few other major cannabinoids, the literature on *C. sativa* is rapidly expanding. In the last few years, in fact, research has enlarged its focus on minor phytocannabinoids [12,13] and other constituents, such as terpenes [14–16], phenolic compounds [14,17], and lipids [18,19]. As a result, the common approaches targeted to a few major compounds might not be sufficient for characterizing *C. sativa* varieties, especially for the evaluation of their medicinal properties [20]. In this context, metabolomics has represented a powerful approach to investigating cannabis constituents as a whole. In 2020, Aliferis et al. introduced the term cannabinomics (cannabis metabolomics) for indicating the approaches that go beyond the analysis of the sole major constituents for cannabis research and development [21]. Hazekamp et al., in the early 2010s, pioneered the use of metabolomics-based workflows for discriminating cannabis cultivars and chemovars by focusing on both major and minor phytocannabinoids as well as terpenes by gas chromatography coupled to flame ionization detector (GC-FID) and mass spectrometry (GC-MS) [22–24]. Similarly, a GC-FID approach targeted to phytocannabinoids and terpenes was employed by Al Bakain et al. for discriminating the geographical field location of cannabis samples from 23 states of the United States [25]. Moreover, a recent paper by Fischedick described for the first time terpene chemotypes by monitoring the terpenoid content of 233 drug-type cannabis samples by GC-FID and GC-MS [26].

Untargeted metabolomics approaches based on liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS), allowing the possibility of identifying numerous as well as unknown compounds, have become pivotal for deciphering the biological role of plant metabolites [27]. Vásquez-Ocmín et al. compared targeted and untargeted LC-HRMS cannabinomics approaches in discriminating cannabis chemovars I-III, highlighting the role of minor constituents [28], and, similarly, Monti et al. combined a targeted method for 15 phytocannabinoids to an untargeted cannabinomics workflow for distinguishing cannabis varieties [29]. A recent study by our research group employed untargeted LC-HRMS coupled with a suspect screening approach (named phytocannabinomics) for the annotation of 135 phytocannabinoid in 50 *C. sativa* accessions and the description of sub-chemotypes that were characterized by different amounts of minor phytocannabinoids [30].

At present, industrial hemp (chemotypes III and IV) has been scarcely investigated compared to drug-type cannabis, despite its significant content in bioactive compounds and nutraceuticals and the progressive relaxation of the prohibition laws worldwide [31]. In the present paper, an untargeted LC-HRMS cannabinomics workflow followed by suspect screening data processing was employed for assessing the role of phytocannabinoids and phenolic compounds in discriminating industrial hemp samples of different varieties and geographical field locations. ANOVA-simultaneous component analysis (ASCA) [32] was then applied to evaluate whether the field location and the cultivar had a significant effect on the composition of the different samples, and, eventually, which phytocompounds are responsible for the main differences.

2. Experimental section

2.1. Chemical and reagents

Analytical grade ethanol 96% and LC-MS grade acetone, acetonitrile, water, formic acid, and acetic acid (AcOH) were purchased from VWR (Radnor, PA, USA). Δ^9-Tetrahydrocannabivarin (Δ^9-THCV), Δ^9-THC, cannabidivarin (CBDV), CBD, CBG, cannabinol (CBN), cannabichromene (CBC), cannabigerolic acid (CBGA), Δ^9-THCA, and CBDA were purchased from Sigma-Aldrich as Cerilliant certified analytical

Table 1
Description of the analyzed hemp varieties (denomination, flower type, time of flowering, and collection dates).

Variety	Flower type	Time of flowering	Collection date			
			Rovigo (RO)	Rutigliano (BA)	Battipaglia (SA)	Libertinia (CT)
Carmagnola	Dioecious	Late	29–09	9–09	8/9-09	16–09
CS	Dioecious	Late	29–09	8–09	8/9-09	14–09
Fibranova	Dioecious	Late	29–09	15–09	8/9-09	14–09
Fibrante	Dioecious	Late	29–09	16–09	8/9-09	14–09
Eletta Campana	Dioecious	Late	29–09	24–08	8/9-09	16–09
Carmaleonte	Monoecious	Medium	02–09	11–08	11–08	20–07
Codimono	Monoecious	Medium	09–09	4–08	17–08	29–07

standards (Sigma-Aldrich, Milan, Italy). Phytocannabinoid stock solutions were prepared at 100 μg mL^{-1} in H_2O/MeOH 90:10 (*v/v*). A phytocannabinoid working mix solution was prepared at 10 μg mL^{-1} in H_2O/MeOH 90:10 (*v/v*), aliquoted, and stored at −20 °C for further use.

Apigenin, apigenin 7-*O*-glucoside, biochanin A, 3-caffeoylquinic acid, caffeic acid, diosmetin, epicatechin, eriodyctiol, ferulic acid, hesperetin, kaempferol, luteolin, myricetin, naringenin, *p*-coumaric acid, procyanidin B1, procyanidin B2, rutin, quercetin, quercetin 3-*O*-glucoside, and taxifolin analytical standards were purchased from Merck. Phenolic compounds stock solutions were prepared at 10 μg mL^{-1} in H_2O/MeOH 90:10 (*v/v*). A phenolic compound working mix solution was prepared at 0.5 μg mL^{-1} in H_2O/MeOH 90:10 (*v/v*), aliquoted, and stored at −20 °C for further use.

2.2. Plant material and experimental design

Seven hemp Italian varieties, five dioecious (Carmagnola, CS, Fibranova, Fibrante Eletta Campana) and two monoecious (Carmaleonte and Codimono) were used, all registered in the Italian/EU register of plant varieties and characterized by a prevalence of CBDA (chemotype III) and a THCA level below 0.2%. The field experiments were carried out between spring and autumn 2021 in four farms belonging to the Council for Agricultural Research and Economics – Research Center for Cereal and Industrial Crops and Research Center Agriculture and Environment. One experimental farm was located in northern Italy (Rovigo, Lat 45° 04′45,4″ N; Long 11°45′57,3″ E), two were located in southern Italy (Battipaglia, Salerno; Lat. 40°34′56.604″N; Long. 14°58′50.232″E; Rutigliano, Bari; Lat.40°59′32.4″N; Long 17°02′03.2″E), and one in insular Italy (Libertinia, Catania; Lat. 37°32′25″N; Long. 14°34′41″E) (Fig. 1).

The experimental plot consisted of three replicated parcels of 20–25 m^2 each for every variety with a sowing density of 100 plants m^2. Irrigation was performed only if necessary, during the early vegetative stage, therefore depending on the different climatic conditions of each site, except for the Libertinia site where no additional water was supplied to the growing plants. Meteorological information (temperature and rainfall) was collected during the whole period of cultivation in the four sites (Table S1). Sowing dates differ from Libertinia (April 2nd) to the other sites (from April 26th in Rovigo to April 30th in Battipaglia and Rutigliano). Flower type, flowering time, and collection dates are reported in Table 1.

Sample collection was done according to the procedure described in CE Regulation n.809/2014 as it is described hereafter. About 20–30 days from the full flowering, the main apical 30 cm female or monoecious inflorescence (comprising also floral bracts, leaves, stems, and seeds) was taken from at least 9 plants/repetition (n = 3) for each variety and dried at 30–35 °C, at dark. Dried inflorescences and floral bracts were separated manually from stems and seeds, and 20 g material was finally shredded with a 1 mm diameter sieve and kept in the dark and at a temperature lower than 20 °C before processing.

Fig. 1. Location of the four experimental fields.

2.3. Sample preparation

For the extraction of phytocannabinoids from hemp biomass samples, the protocol form the German Pharmacopoeia was employed [33]. Extraction with 45 mL EtOH was carried out on 500 mg of finely grounded freeze-dried hemp biomass in three cycles and the combined extracts were brought to 50 mL final volume with fresh EtOH in a volumetric flask. The solution was filtered through a 0.45 μm regenerated cellulose filter and a 100 μL aliquot was diluted with 900 μL of mobile phase (acetonitrile/H_2O, 60:40, v/v, both with 0.1% formic acid).

Phenolic compounds were extracted following the protocol of Pellati et al. [34] with some modifications. Briefly, 500 mg of each freeze-dried hemp sample was extracted with 10 mL acetone/H_2O/acetic acid (70:29.5:0.5, *v/v/v*), sonicated for 15 min in an ice bath, and then centrifuged for 10 min at 2000×*g*. The supernatant was collected, and the procedure was repeated once. The supernatants were mixed and concentrated to 4.5 mL using a Speed-Vac SC 250 Express (Thermo 164 Avant, Holbrook, NY, USA). Then, 500 μL of MeOH was added to the sample, and the final extract solution (H_2O/MeOH, 90:10, *v/v*) was filtered through a 13-mm Acrodisc Syringe filter with a 0.2 μm GH Polypro membrane (Pall, Ann Arbor, MI, USA). Finally, the extract was aliquoted and stored at −20 °C for further analysis. For both extractions, process blank samples were obtained following the described extraction procedures on a solvent sample.

2.4. UHPLC-HRMS untargeted analysis

Untargeted data acquisition was performed following the recommendations of the metabolomics Quality Assurance and Quality Control Consortium (mQACC) [35]. Intra-study quality control (QC) samples were obtained by pooling 20 μL of each of the 84 hemp extracts. Distinct pooled QC samples were prepared for phytocannabinoids and phenolic compounds.

A Vanquish Core System equipped with a binary pump, a vacuum degasser, thermostated autosampler (4 °C), a column compartment (30 °C), and a diode array detector was interfaced to an Exploris 120 Orbitrap mass analyzer (Thermo Fisher Scientific, Bremen, Germany) with a heated electrospray ionization (HESI) source for phytocannabinoid analysis. The chromatographic parameters were adapted from a previously validated method with slight modifications [36,37]. The separation was achieved on a Poroshell 120 EC C18 column (100 × 3.0 mm I.D., 2.7 μm particle size, Agilent Technologies, Santa Clara, USA) with a gradient elution of acetonitrile from 60% to 95% in 15 min and an isocratic step at 95% acetonitrile held for 3 min, then a washing step of 4 min at 98% acetonitrile and a re-equilibration step at 60% acetonitrile for further 4 min. The flow rate was maintained at 0.5 mL min^{-1} throughout the entire run (26 min). The HESI parameters were set as follows: capillary temperature, 390 °C; vaporizer temperature, 150 °C; electrospray voltage, 3.8 kV; sheath gas, 70 arbitrary units; auxiliary gas, 5 arbitrary units; S lens RF level, 70. In full-scan mode the parameters were set as follows: resolution, 60,000 FWHM (full width at half maximum) at *m/z* 200; scan range, *m/z* 50–750; maximum injection time, 54 ms; isolation window, *m/z* 0.7. In DDA mode the following parameters were employed: resolution, 15,000 FWHM; maximum injection time, 22 ms; isolation window, *m/z* 1.2; stepped NCE (normalized collision energy), 20-40-100 [36,37]. The injection volume was 5 μL. The analyses were acquired with Xcalibur 3.0 (Thermo Fisher Scientific) and processed using TraceFinder 5.0 (Thermo Fisher Scientific).

For phenolic compound analysis, a Vanquish binary pump H (Thermo Fisher Scientific, Bremen, Germany), equipped with an autosampler and controlled temperature column compartment, was used for chromatographic separation on a Kinetex XB C18 column (100 × 2.1 mm, 2.6 μm particle size, Phenomenex, Torrance, USA). The mobile phases were H_2O/HCOOH (99.9:0.1, *v/v*; phase A) and ACN/HCOOH (99.9:0.1, *v/v*; phase B) and were mixed with the following gradient: 5–15% phase B in 10 min, 15–35% phase B in 15 min, 35–50% phase B in 5 min; at the end of the gradient, a washing step at 95% phase B for 5 min and a re-equilibration step at 5% phase B were performed. The column was maintained at 40 °C with a constant flow of 600 μL min^{-1}. The chromatographic system was coupled to a hybrid quadrupole-Orbitrap mass spectrometer Q Exactive (Thermo Fisher Scientific) with a heated ESI source, operating in negative ion mode under the following conditions: the capillary temperature was set at 275 °C, spray voltage at 2500 V (−), auxiliary gas heater temperature at 300 °C, sheath gas at 50 a.u. (arbitrary units), auxiliary gas at 15 a.u., sweep gas was 3 a.u., and S-Lens RF level was 50 (%). MS data were acquired in the range 150–1000 *m/z* with the following parameters: resolution (full width at half maximum, FWHM, at *m/z* 200) of 70,000, automatic gain control (AGC) target value was 200,000, the maximum ion injection time was 100 ms, and the isolation window width was 2 *m/z*. MS2 fragmentation was performed was conducted in the Top 5 data-dependent acquisition (DDA) mode with a resolution (FWHM, at *m/z* 200) of 35,000, with AGC target value at 100,000 and dynamic exclusion set to 3 s. Stepped collision energy (CE) fragmentation was achieved in the HCD cell at three-stepped 20-40-60 normalized collision energy (NCE). Raw MS/MS data files were acquired by Xcalibur software (version 3.1, Thermo Fisher Scientific).

Raw MS/MS data files were acquired by Xcalibur software (version 3.1, Thermo Fisher Scientific).

Samples were injected in a randomized order and the chromatographic worklist is schematized in Supplementary Material Table S2. For system suitability testing, the column stability and performance were tested before and after each analytical section using solvent blank samples (H_2O/MeOH, 90:10, *v/v*) and working mix standard solutions. System conditioning, consisting of ten consecutive pooled QCs sample injections, preceded the process blank sample injection for background subtraction, which allowed to discard of both the contaminants present in extraction solvents, mobile phases, and the HPLC-MS system and the compounds subjected to high carry-over effects (more than 10%), which may alter peak areas, possibly resulting in biased statistical analysis. After further system reconditioning with ten more QCs samples, randomized samples were run in groups of five, followed by a QC injection.

2.5. Data preprocessing and compound identification

The.raw data obtained by the analysis of samples, QCs, and the process blank were preprocessed using the software Compound Discoverer version 3.1 (Thermo Fisher Scientific). The phytocannabinoid data were preprocessed using a modified version of a customized data processing workflow that was previously set up in a previous paper [13]. Similarly, the phenolic compound runs were preprocessed using a homemade workflow on Compound Discoverer [17]. However, two distinct preprocessing were set up for flavonoids and phenolic acids, resulting in three distinct data matrices obtained for the following chemometric analysis. Data preprocessing allowed feature alignment, background removal, adduct definition and grouping, molecular formula annotation, and QC-based normalization.

Phytocompounds (phytocannabinoids, flavonoids, and phenolic acids) were identified using a suspect screening approach previously described that is based on the use of homemade databases and software-assisted spectral annotation [13,17]. Details on the data preprocessing on Compound Discoverer are reported in the Supplementary Material as well as the nomenclature of the identified phytocannabinoids that was given according to the literature. Data for the tentatively identified compounds are summarized in Tables S3–S5 with the related confidence level according to Schymanski et al. [38].

2.6. Chemometric strategies for data processing

When multivariate data are acquired according to an underlying experimental design, the effect of the controlled factors and their

interaction can be evaluated by multivariate generalizations of the classical Analysis of Variance (ANOVA). In particular, when dealing with a high number of correlated variables, as in metabolomic studies, the most popular of such techniques is ANOVA-simultaneous component analysis (ASCA) [39]. ASCA works by decomposing the mean-centered data matrix $\boldsymbol{X}_c$ into the additive contributions of individual effect matrices accounting for each term of the experimental design, according to the ANOVA scheme. For instance, in the present study where the experimental design involves two factors, site, and variety, the ANOVA decomposition takes the following form:

$$\boldsymbol{X}_c = \boldsymbol{X}_{var} + \boldsymbol{X}_{site} + \boldsymbol{X}_{int} + \boldsymbol{X}_{res}$$

where $\boldsymbol{X}_{var}$ and $\boldsymbol{X}_{site}$ are the matrices accounting for the effect of site and variety, respectively, $\boldsymbol{X}_{int}$ the effect matrix corresponding to their interaction, and $\boldsymbol{X}_{res}$ collects the residuals, i.e., the variability not explained by the additive ANOVA model.

The entity of the effect of the individually designed terms, i.e., to what extent the experimental profiles are affected by the different levels of that particular factor, can be quantified by calculating the sum of squares of the corresponding effect matrix:

$$Eff_j = \left\|\boldsymbol{X}_j\right\|^2 \quad j = \{site, var, int\}$$

where $\left\|\boldsymbol{X}_j\right\|^2$ indicates the Frobenius' norm of the corresponding matrix.

The significance of the corresponding effect is evaluated by comparing the value of Eff_j to its null distribution which is non-parametrically estimated by permutation tests [40,41]. If a term is deemed to be statistically significant, the corresponding effect can be interpreted by calculating a PCA model of the corresponding effect matrix:

$$\boldsymbol{X}_j = \boldsymbol{T}_j \boldsymbol{P}_j^T$$

where and are the scores and loading matrices calculated for the particular design term. Since effect matrices are made of mean experimental profiles corresponding to the different levels of the factor or interaction term, the scores plot will have as many points as the number of factor levels, so it will account for the between-level variation. Therefore, the significance of the difference among the levels of that factor or interaction can then be graphically appreciated by projecting the residual matrix onto the factor subspace, which will result in a new scores matrix $\boldsymbol{T}_j^{'}$, which will show as well the within-level variation [42]:

$$\boldsymbol{T}_j^{'} = \left(\boldsymbol{X}_j + \boldsymbol{X}_{res}\right)\boldsymbol{P}_j$$

Bootstrapping [43,44] can instead be used to calculate the confidence intervals of the loadings to identify significantly contributing metabolites.

3. Results and discussion

3.1. Untargeted phytocompound identification

In this work, eighty-four different cannabis samples (four experimental fields, seven different varieties, and three distinct samples per each field/variety combination) were analyzed by an untargeted metabolomics approach based on UHPLC-HRMS and suspect screening data analysis. For the two main classes of compounds (phytocannabinoids and phenolic compounds), two distinct extraction procedures were employed. As such, a more selective extraction mixture for phenolic compounds, comprising 30% of water, allowed to minimize the co-extraction of phytocannabinoids, which would have likely resulted in separation and ion suppression issues. This analytical platform permitted the comprehensive characterization of compound classes of well-known structural complexity [17,45]. Phytocannabinoids, for instance, exist in several different structures, that arise from the ten main phytocannabinoid base structures in combination with numerous possible modifications on all three sections of a phytocannabinoid structure (i.e., the resorcinyl core, the terpenoid moiety, and the linear alkyl side chain) [12,13]. For both phytocannabinoids and phenolic compound data acquisition, negative ion mode (ESI(−)) was preferred to positive mode (ESI(+)), being the sole that permits the ionization of phenolic acids and the preferred method for untargeted phytocannabinoid identification [12,13]. Moreover, as samples were analyzed in their native form (i.e., before decarboxylation processes), the presence of carboxyl groups on most phytocannabinoids was also expected to boost their ionization efficiency in the ESI(−), which has also the advantage of generally lower background noise [46]. Our previous results on phenolic compounds in industrial hemp [17] demonstrated that the use of both polarities did not significantly raise the number of identified compounds, whereas it would have doubled the time for data acquisition and processing. Moreover, the obtention of two datasets per class of compounds with almost the same compounds but different peak areas would have caused several redundancies following the statistical analysis. For these reasons, data acquisition in the sole negative ion mode was preferred to a double acquisition.

Following the suspect screening data processing, a total of 265 phytocompounds were tentatively identified, including 54 phytocannabinoids, 134 flavonoids (including polymeric flavonoids), and 77 phenolic acids. Identification data of the annotated compounds are reported in detail in Tables S3–S5, including retention times, accurate mass-to-charge ratios (m/z), mass errors, diagnostic product ions, level of identification [38,47], and IDs (i.e., C1–C54 for cannabinoids, F1–F134 for flavonoids, and P1–P77 for phenolic acids). Unsurprisingly, the number of annotated phytocannabinoids in the analyzed industrial hemp samples was significantly lower than previously reported for drug-type genotypes using the same analytical workflow due to the overall lower content of THC-type phytocannabinoids and their non-enzymatic derivatives (e.g., CBN- and CBT-type phytocannabinoids) [13,30]. CBD-type, CBG-type, and cannabielsoin (CBE)-type phytocannabinoids were the most numerous, with 10, 9, and 9 annotated compounds, respectively. CBE-type compounds are non-enzymatic degradation products of CBD-type compounds, even though recent reports have highlighted the role of microbial enzymes in converting carboxylated CBG-type to carboxylated CBE-type compounds [48]. Other than CBDA (compound C36), which was the most abundant phytocannabinoid in all analyzed samples, several alkyl analogs were annotated, i.e., CBDOA (C1), CBDVA (C11), CBDBA (C18), CBDHA (C42), and CBDPA (C50), as well as three *O*-methylated compounds (C19, C41, C43), which were distinguished by their fragmentation patterns and retention times. Similarly, alkyl chain analogs and O-methylated derivatives of CBG were annotated, as well as SesquiCBGA (C54), a longer prenyl chain analog of CBGA [49].

The flavonoid composition of the analyzed industrial hemp samples was significantly more complex than the phytocannabinoid profile. As previously demonstrated [14,17], the most abundant flavonoids in *C. sativa* inflorescence were flavone derivatives (68 annotate compounds), either hydroxylated (i.e., apigenin, luteolin, tricetin) or methoxylated (i.e., acacetin, chrysoeriol, diosmetin). Flavone derivatives were determined both O-glycosylated and C-glycosylated (i.e., orientin, vitexin) and also O-acyl glycosylated (to acetic and malonic acid). Even cannflavins (F42–F44), the flavonoids unique to *C. sativa* that are known for their significant antioxidant activity [50], are effectively prenylated flavones. Other classes of flavonoids that were annotated are flavonols (i.e., isorhamnetin, kaempferol, quercetin) and flavanones (i.e., eriodictyol, hesperetin, naringenin, pinocembrin), with 25 and 23 compounds, respectively. Furthermore, catechin (F45) and epicatechin (F60) were annotated alongside seven b-type procyanidin dimers and trimers (F36–41, F116).

Most annotated phenolic acids were glycosyl derivatives of hydroxycinnamic acids (i.e., coumaric acid, caffeic acid, ferulic acid, and sinapic acid) and, to a lesser extent, hydroxybenzoic acids. Despite being

less numerous, hydroxycinnamoyl conjugates of quinic and shikimic acid had significant peak areas. The determination of the positional isomers of hydroxycinnamoyl quinic acids was based on the retention trends and fragmentation mechanisms discussed by Clifford [51] and Willems [52]. 3-Caffeoyl quinic acid (neochlorogenic acid, P9), the most abundant of these compounds, was annotated for its lower retention time and its balanced abundance of the product ions deriving from the quinic acid moiety (*m/z* 191.0563) and the caffeic acid moiety (*m/z* 179.0351 and 135.0453). 5-Caffeoyl quinic acid (chlorogenic acid, P24), on the other hand, produced the quinic acid fragment at a significantly higher abundance, whereas 4-caffeoyl quinic acid (cryptochlorogenic acid, P29) was annotated thanks to its diagnostic ion at *m/z* 173.0455. Coumaroyl and feruloyl conjugates as well as the three dicaffeoyl quinic acid positional isomers (P75–P77) were tentatively identified with the same rationale, with each coumaroyl derivative being a pair of coumaric acid isomer derivatives. In Supplementary Material Fig. S1, the MS/MS spectra of the three caffeoyl quinic acids are shown, whereas Fig. S2 shows the extracted chromatograms of the annotated hydroxycinnamoyl quinic acids.

3.2. ASCA analysis of the phytocompounds

The different phytocompounds were analyzed by ASCA, to evaluate the significance of the field location and cultivar effects. In the beginning, ASCA models were calculated separately on phytocannabinoids, flavonoids, and phenolic acids since its application on the entire dataset as it was would have been affected by the variability present from one block of data to the other, resulting in the unwanted situation that a data matrix of higher dimensions would contribute more to the model simply because it contains more variables. To obtain global results, this issue was overcome in a low-level data fusion fashion. Given the relatively numerous groups that were taken into consideration (4 geographical origins and 7 different strains), the statistical analysis was aimed at evaluating the role of minor constituents in discriminating cannabis samples and especially phenolic compounds that were never employed for the purpose rather than at identifying single markers. As such, instead of single marker compounds, trends in the abundance of classes of structurally related compounds were expected, in light of the structural properties of the identified compounds (e.g., several conjugates of the same flavonoid). Regardless of the nature of the data, all ASCA models were calculated to inspect two distinct effects: the location and the variety. Significance has been evaluated by a permutation test, using 10^4 randomizations. Before the analysis, data were preprocessed by $\log_{10}$ and mean-centering to remove off-sets.

3.2.1. ASCA analysis of phytocannabinoids

The ASCA model calculated on the phytocannabinoids data revealed that both the location and the cultivar effects (and their interaction) are significant ($p < 0.001$ for all). The model of the geographical field location effect required 3 Simultaneous Components (SCs), which explain 93.0%, 4.9%, and 2.1% of the total variance, respectively, whereas the one dedicated to the cultivar needed 6 latent variables, explaining the 65.2%, 16.8%, 10.3% 5.6%, 1.3%, and 0.80% of the total variance.

The scores plots associated with these models can be observed in Fig. 2, where samples have been projected onto the space spanned by the first SCs when modeling the field location (Fig. 2A) and the cultivar (Fig. 2B).

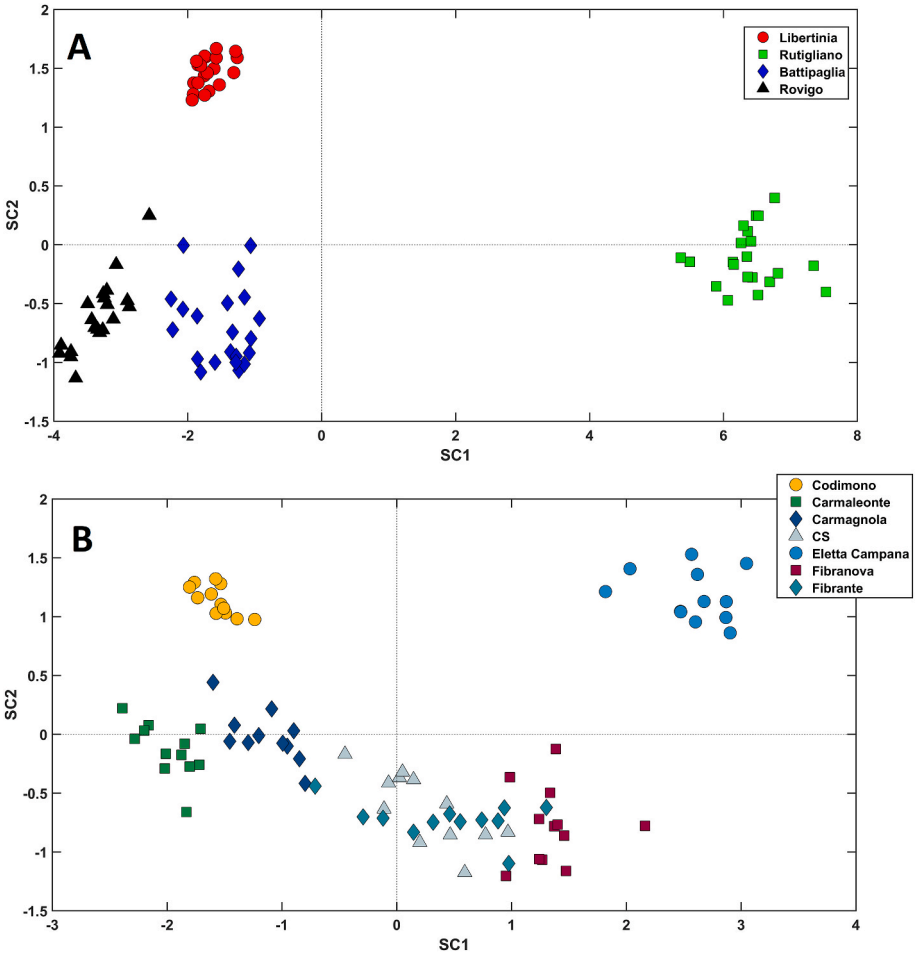


Fig. 2. Scores plot of the ASCA models of the geographical location (A) and cultivar (B) effects for the phytocannabinoid dataset.

The investigation of Fig. 2A makes it apparent that the samples collected in Rutigliano, falling at positive values of SC1, are clearly distinguishable from all the other presenting negative values of this component. Similarly, the second SC pointed out a peculiar composition of materials sampled from Libertinia compared to all the others. All variables contribute significantly to the first SC except compound C28 (CBT isomer 3). The inspection of the loadings (Fig. S3) revealed that samples harvested in Rutigliano (SC1 > 0) presented the least amount of most phytocannabinoids (including all the major ones), whereas the samples collected in the Rovigo site showed generally the highest amounts. On the other hand, SC2 allowed the discrimination of samples harvested in Libertinia based on the higher content of several non-enzymatic deriving cannabinoids, such as CBNVA, CBNDA, CBNA, and several compounds of the CBE and CBT classes.

Concerning the cultivar effect, the scores plot shows that the different cultivars are distributed along the first CS. In particular, Eletta Campana and Fibranova fell at positive values of SC1, Fibrante and CS around zero, and all the others (Codimono, Carmaleonte, and Carmagnola) at negative values. The loadings associated with this component can be observed in Fig. 3; from this plot, it is possible to appreciate which compounds contribute to the diversification of the investigated materials. In the figure, red and blue bars represent significant and non-significant variables (the indexes correspond to those in Tables S3–S5), respectively. Among the significant ones, those presenting a positive score are the compounds whose concentration is higher in samples falling at positive values of SC1 (Eletta Campana and Fibranova) and lower in individuals presenting negative SC1-values, on the other hand, the phytocannabinoids having negative loadings on SC1 are highly contained in Camaleonte, Codimono, and Carmagnola, and then are at the poorest concentration in Eletta Campana.

In general, samples of Eletta Campana showed the highest amount of most of the phytocannabinoids, especially O-methylated ones (compounds C19, C30, C41, C43, and C49). Interestingly, samples having SC1 < 0 presented higher amounts of CBG-type phytocannabinoids (e.g., compounds C3, C39, and C54). Since CBG-type cannabinoids are the progenitor of the other classes [2], samples of Codimono, Carmagnola, and Carmaleonte may have a lower expression (or activity) of the enzymes that convert CBG-type phytocannabinoids into other ones. Along the second SC (Fig. S4), samples follow a clear trend: Codimono and Eletta Campana fall at positive values of SC2 differing from all the other cultivars. The inspection of the loadings can be imputed to a higher concentration of varinoids (three carbon-chain phytocannabinoids [45], compounds C9, C11–C14).

3.2.2. *ASCA analysis of flavonoids*

Whether the site of cultivation and the cultivar were significant effects on the composition of hemp flavonoids has been also investigated. Similarly, as before, both effects and their interaction were significant ($p < 0.001$ for all). The model of the geographical field location effect required 3 SCs, explaining the 58.5%, 28.8%, and 12.7% of the total variance. The scores plot is displayed in Fig. 4A, and its inspection confirmed samples harvested in Rutigliano and Libertinia presented a distinctive composition compared to the individuals grown in the other experimental fields.

Indeed, contrary to all the other individuals, samples from Rutigliano presented positive scores for the first component. The inspection of the loadings (Fig. S5) revealed this is driven by a higher content of C-glycosylated compounds, such as in (iso)orientin and several derivatives (compounds F2–F6, F9-10), as well as methyl-luteolin C-glucuronide (F98) and tricetin C-hexoside (F126) and significantly lower content of proanthocyanidins (F110, F129-134). Among the other compounds with higher abundances in samples from Rutigliano, there were hydroxyluteolin derivatives (F17-18) and some flavone glucuronides (F21, F128). On the other hand, samples from Libertinia, Battipaglia, and Rovigo presented higher concentrations in non-conjugated flavonoids (aglycones, i.e., F24, F40, F47, F59, F74, F82, and F111) and cannflavins (F36-38). Inspecting SC2 it appeared that samples from Libertinia were characterized by a higher content of conjugated flavonols (e.g., F71–73, F76–80, F114-117), whereas Battipaglia and Rovigo had the highest concentrations of proanthocyanidins.

The ASCA model of the cultivar effect unveiled a strict link between flavonoid composition and cultivar. In fact, the inspection of the scores plot (Fig. 4B) revealed clear grouping tendencies, with strong overlap between Carmagnola and CS, and among Carmaleonte, Fibranova, and Fibrante is appreciable. Eletta Campana, Carmagnola, and CS fell all at positive values of SC1 (Fig. S6A), whereas all the others have negative values of this component. It is worth mentioning that, unlike the results for the different field locations, the ASCA models of the cultivars based on phytocannabinoids and flavonoids furnished quite different appearances. Codimono has the highest abundance of C-glycosylated compounds, whereas Eletta Campana had a significantly higher amount of O-methylated flavones (i.e., F20, F23, F41, F44, F100), that was in line with the result of O-methylated phytocannabinoids and possibly hints a superior activity of the enzymes involved in O-methylations. The loadings plot of SC2 (Fig. S6B) showed a general trend for most flavonoids to be more abundant for higher SC2 values, with the cluster of Carmagnola and CS showing generally the lowest amounts.

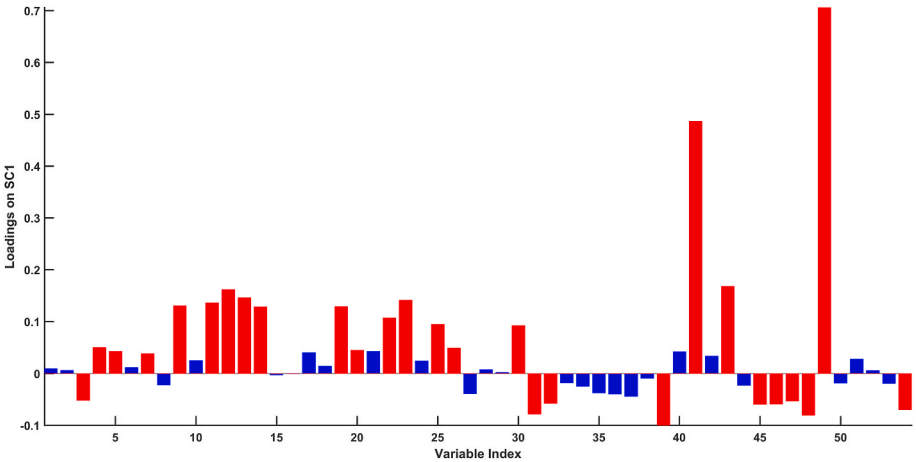


Fig. 3. Graphical representation of the loadings of the variables alongside the first simultaneous component of the ASCA model of the cultivar effect obtained for the phytocannabinoid dataset.

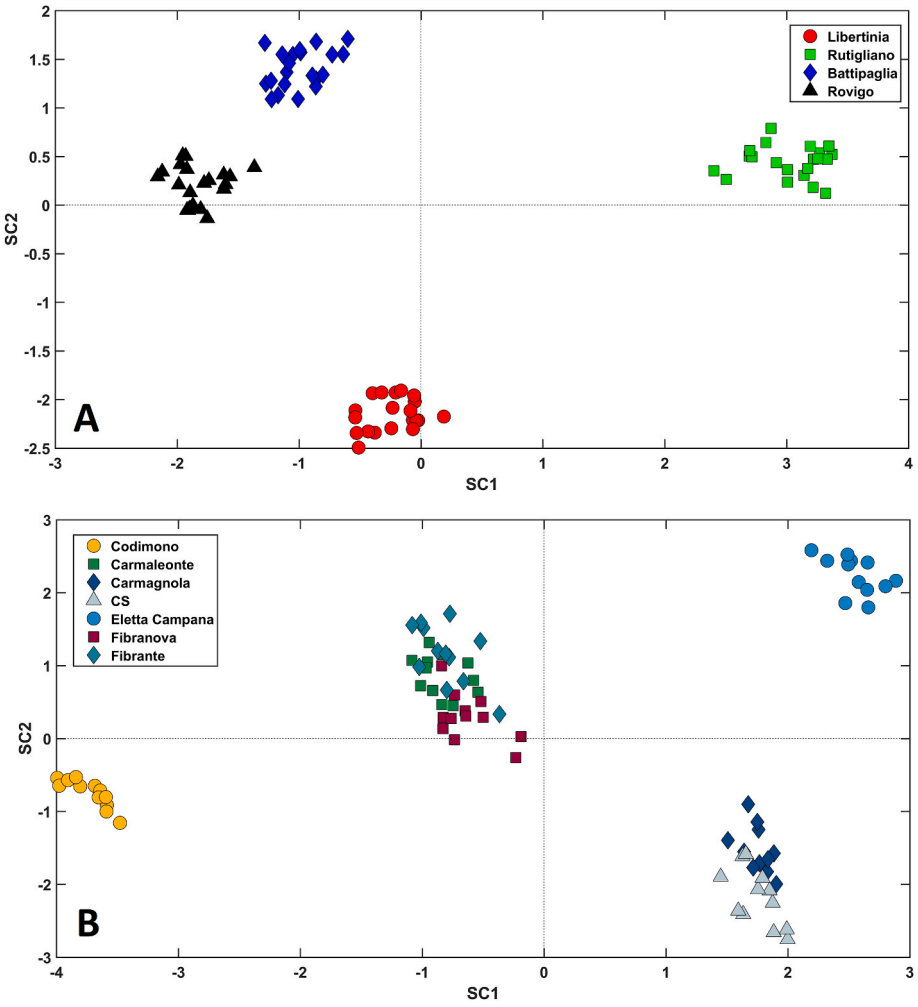


Fig. 4. Scores plot of the ASCA models of the geographical location (A) and cultivar (B) effects for the flavonoid dataset.

3.2.3. *ASCA analysis of phenolic acids*

When ASCA is applied to the Phenolic acids data set, the cultivation site, cultivar, and their interaction resulted in significant effects ($p < 0.001$ for all). The model of the field location required 3 SCs, explaining 43.5%, 37.6%, and 18.9% of the total variance, whereas the one associated with the cultivar needed 6 SCs for the explanation of 64.7%, 18.7%, 7.3%, 4.6%, 2.4% and 2.2% of the variance, respectively. The scores plot associated with the model of the geographical field location effect depicted a different scenario concerning the previously discussed models. In fact, as observed in Fig. 5A, samples from Rutigliano and Battipaglia overlap, discriminating themselves from those harvested in Libertinia and Rovigo. Seemingly, flavonoids and phenolic acids were differently affected by the pedoclimatic conditions of hemp growth, with the latter much more apparently affected by the latitude (which is similar between Rutigliano and Battipaglia).

The inspection of the loadings (Fig. S7) revealed the role of hydroxycinnamic acids conjugated with quinic acid. Samples from Rutigliano, in fact, were differentiated by lower levels of hydroxycinnamoyl conjugates in positions 3 and 5 and significantly higher amounts of conjugates in position 4 of quinic acid (P29, P36, P47, P61, P77). These results not only indirectly contributed to confirming the MS-based identification that was earlier described, but also enlightened a peculiarity in a set of samples that was not driven by the nature of the biomolecules but by their type of conjugation. On the other hand, the inspection of SC2 revealed that Rovigo has a peculiar composition of phenolic acids driven by a higher content in shikimic acid conjugates (P55, P68-69, P73).

The interpretation of the scores plot associated with the model of the cultivar effect (Fig. 5B and S8) is less straightforward than the previously discussed ones. The main difference which came out concerned Carmaleonte samples that showed significantly higher concentrations in dicaffeoyl quinic acids (Compounds P75-77) compared to all other cultivars.

3.2.4. *ASCA analysis of low-level fused data*

Not surprisingly, the merged data model revealed that the field location and the cultivar were significant. The model of the former effect needed 3 SCs for the explanation of 79.1%, 13.7%, and 7.2% of the total variance, while the model associated with the cultivar required 6 SCs, explaining the 40.71%, 27.8%, 16.2%, 8.2%, 4.5% and 2.5% of the variance, respectively. The merged data model of the geographical field locations was similar to that of the sole phytocannabinoids and flavonoids, even though the clusters were more compact. Based on these results, the pedoclimatic conditions of the experimental fields appeared to play a significant role in the overall phytocompound expression. The significant discrimination of samples from Rutigliano from all others on SC1 (Fig. 6A) can be hardly associated with the meteorological data in Table S2, since plants grew at similar temperatures compared to those from Battipaglia, and even if May and June (when the transition from vegetative to the reproductive stage is ongoing) were rainier than in the other sites, this could have not been affected the plant secondary metabolism at maturity. However, the higher total rainfall in September could have affected the total phytocannabinoid content at maturity which was lower in Rutigliano for all types of phytocannabinoids. This

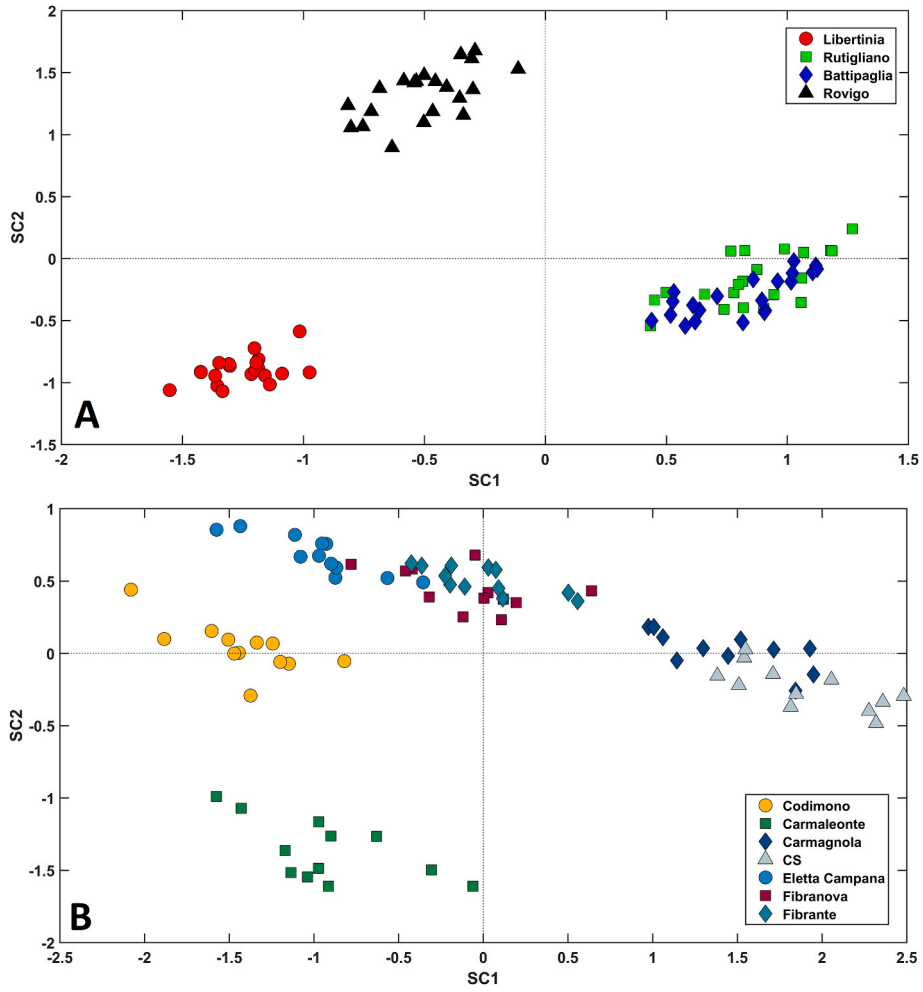


Fig. 5. Scores plot of the ASCA models of the geographical location (A) and cultivar (B) effects for the phenolic acid dataset.

observation is in agreement with the ecological function of cannabinoids in protecting the plant against UV light and desiccation, as well as plant defense in general [2]. On the other hand, the peculiarity of samples from insular Italy (Libertinia) that is visible on SC2 and explained by higher concentrations of several flavonoids could be linked to the hotter and drier climate of that area. The effect of climate on the phenolic compound expression was previously studied by Kumar [53] and Kabtni [54], who both concluded that arid climates favor the biosynthesis of phenolic compounds.

The merged data model of the hemp varieties (Fig. 6B) shows that samples belonging to the different cultivars spread along SC1, with Codimono and Carmaleonte at negative SC1 values, the non-overlapping cluster of Fibranova and Fibrante around zero, and Eletta Campana and the cluster of Carmagnola and CS at positive SC1 values.

In contrast to the three previous models on the single phytocompound classes, the merged model is the sole allowing the discrimination between monoecious and dioecious varieties (SC1 < 0 for medium-flowering monoecious cultivars and SC1 around zero or > 1 for late-flowering dioecious ones), with a significantly higher concentration of O-methylated cannabinoids and flavonoids at more positive SC1 values and larger amounts of several phenolic acids for monoecious varieties.

Codimono and Carmaleonte have different genetic backgrounds and the two clusters are themselves well distinct along SC1. Instead, the genetic pool of the Italian dioecious varieties is partly shared as CS, Fibranova, Fibrante, and Eletta Campana derived from Carmagnola throughout different breeding programs [55]. Carmagnola is a Northern Italian landrace and is the oldest Italian hemp dioecious cultivar used by many breeders to develop hemp varieties. CS or "Carmagnola Selezionata" was selected directly from this landrace in the 1960s and the common background explains the cluster formed by these two varieties. Fibranova was selected from crosses among Carmagnola and Russian hemp strains to increase productivity and fiber quality while Eletta Campana was derived from a German variety and despite the common background is the more different one among the dioecious varieties as clearly appears also from the ASCA analyses of both single and fused data. Fibrante was obtained by mutagenesis of the pollen of Carmagnola and Fibranova and resulted in more similarities for phytocannabinoids and flavonoids to Fibranova.

4. Conclusions

The increasing spread of hemp and its derivatives has led to the need for analytical approaches for evaluating its composition beyond the focus on the major phytocannabinoids of *C. sativa*. In this regard, metabolomics has been proposed as a powerful tool for this purpose, leading to the coining of the term cannabinomics. As previously stated, cannabinomics must therefore be understood as an approach that extends the analysis from phytocannabinoids to other classes of phytocompounds [28]. In this work, an untargeted HRMS cannabinomics approach directed to major and minor phytocannabinoids, flavonoids, and phenolic acids was employed for evaluating the contribution of these classes of compounds in discriminating industrial samples based on their cultivar and geographical field location. Our results have

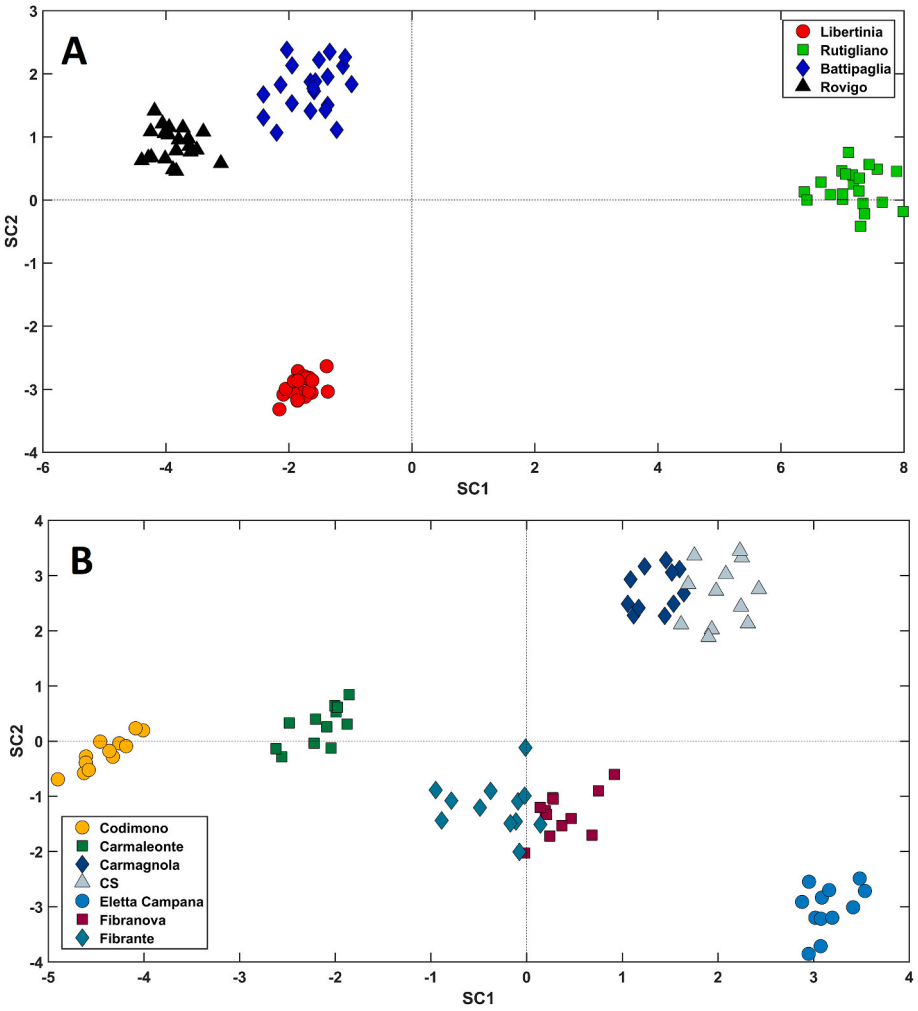


Fig. 6. Scores plot of the ASCA models of the geographical location (A) and cultivar (B) effects for the low-level fused dataset.

demonstrated that phenolic compounds not only can effectively discriminate hemp samples but they do so often in different ways compared to phytocannabinoids (especially in terms of the cultivar effect). Moreover, a low-level fused model was built employing all tentatively identified compounds and demonstrated significantly better results in terms of group classification compared to the sole phytocannabinoids due to the quite similar phytocannabinoid profile of these chemotype III hemp varieties. As such, our study demonstrated the need for comprehensive cannabinomics approaches for evaluating the best conditions to meet the demands for material, food, and medicinal purposes. Further studies are needed for extending the classification models to other classes of compounds of industrial interest, such as terpenes and lipids.

Credit author statement

Andrea Cerrato: Writing – original draft, Methodology; Alessandra Biancolillo: Formal analysis, Visualization; Giuseppe Cannazza: Supervision, Funding acquisition; Chiara Cavaliere: Writing – review & editing; Cinzia Citti: Investigation; Aldo Laganà;: Conceptualization, Project administration; Federico Marini: Formal analysis; Massimo Montanari: Resources; Carmela Maria Montone: Investigation; Roberta Paris: Resources, Funding acquisition, Writing – review & editing; Nino Virzì;: Resources; Anna Laura Capriotti: Conceptualization, Writing – review & editing, Supervision

Fundings

This work was partly supported by UNIHEMP research project "Use of iNdustrIal Hemp biomass for Energy and new biocheMicals Production" (ARS01_00668) funded by Fondo Europeo di Sviluppo Regionale (FESR) (within the PON R&I 2017-2020—Axis 2—Action II—OS 1.b; Grant decree UNIHEMP prot. n. 2016 of July 27, 2018; CUP B76C18000520005)

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank Laura D'Andrea (CREA – Agriculture and Environment, Bari, Italy) and Francesco Raimo (CREA - Research Center for Cereal and Industrial Crops, Caserta, Italy) for the hemp samples and the support during the research activities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2023.341716.

References

[1] P.-A. Chouvy, Cannabis cultivation in the world: heritages, trends and challenges, EchoGéo (2019), https://doi.org/10.4000/echogeo.17591.
[2] 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.
[3] M. Haney, Cannabis use and the endocannabinoid system: a clinical perspective, Am. J. Psychiatr. 179 (2022) 21–25, https://doi.org/10.1176/appi.ajp.2021.21111138.
[4] E. Groce, The health effects of cannabis and cannabinoids: the current state of evidence and recommendations for research, J. Med. Regul. (2018), https://doi.org/10.30770/2572-1852-104.4.32.
[5] J.A. Crippa, F.S. Guimarães, A.C. Campos, A.W. Zuardi, Translational investigation of the therapeutic potential of cannabidiol (CBD): toward a new age, Front. Immunol. 9 (2018), https://doi.org/10.3389/fimmu.2018.02009.
[6] T.P. Freeman, C. Hindocha, S.F. Green, M.A.P. Bloomfield, Medicinal use of cannabis based products and cannabinoids, BMJ (2019) l1141, https://doi.org/10.1136/bmj.l1141.
[7] M.N. Tahir, F. Shahbazi, S. Rondeau-Gagné, J.F. Trant, The biosynthesis of the cannabinoids, J. Cannabis Res. 3 (2021) 7, https://doi.org/10.1186/s42238-021-00062-4.
[8] E. Small, Evolution and classification of cannabis sativa (marijuana, hemp) in relation to human utilization, Bot. Rev. 81 (2015) 189–294, https://doi.org/10.1007/s12229-015-9157-3.
[9] E. Small, H.D. Beckstead, Cannabinoid phenotypes in cannabis sativa, Nature 245 (1973) 147–148, https://doi.org/10.1038/245147a0.
[10] G. Fournier, C. Richez-Dumanois, J. Duvezin, J.-P. Mathieu, M. Paris, Identification of a new chemotype in cannabis sativa : cannabigerol - dominant plants, biogenetic and agronomic prospects, Planta Med. 53 (1987) 277–280, https://doi.org/10.1055/s-2006-962705.
[11] G. Mandolino, A. Carboni, Potential of marker-assisted selection in hemp genetic improvement, Euphytica 140 (2004) 107–120, https://doi.org/10.1007/s10681-004-4759-6.
[12] 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.
[13] 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.
[14] E. Mazzara, J. Torresi, G. Fico, A. Papini, N. Kulbaka, S. Dall'Acqua, S. Sut, S. Garzoli, A.M. Mustafa, L. Cappellacci, D. Fiorini, F. Maggi, C. Giuliani, R. Petrelli, A comprehensive phytochemical analysis of terpenes, polyphenols and cannabinoids, and micromorphological characterization of 9 commercial varieties of cannabis sativa L, Plants 11 (2022) 891, https://doi.org/10.3390/plants11070891.
[15] G. Micalizzi, F. Alibrando, F. Vento, E. Trovato, M. Zoccali, P. Guarnaccia, P. Dugo, L. Mondello, Development of a novel microwave distillation technique for the isolation of cannabis sativa L. Essential oil and gas chromatography analyses for the comprehensive characterization of terpenes and terpenoids, including their enantio-distribution, Molecules 26 (2021) 1588, https://doi.org/10.3390/molecules26061588.
[16] Y. Pieracci, R. Ascrizzi, V. Terreni, L. Pistelli, G. Flamini, L. Bassolino, F. Fulvio, M. Montanari, R. Paris, Essential oil of cannabis sativa L: comparison of yield and chemical composition of 11 hemp genotypes, Molecules 26 (2021) 4080, https://doi.org/10.3390/molecules26134080.
[17] A. Cerrato, G. Cannazza, A.L. Capriotti, C. Citti, G. La Barbera, A. Laganà, C. M. Montone, S. Piovesana, C. Cavaliere, A new software-assisted analytical workflow based on high-resolution mass spectrometry for the systematic study of phenolic compounds in complex matrices, Talanta 209 (2020), 120573, https://doi.org/10.1016/j.talanta.2019.120573.
[18] M. Antonelli, B. Benedetti, G. Cannazza, A. Cerrato, C. Citti, C.M. Montone, S. Piovesana, A. Laganà, M. Antonelli, B. Benedetti, G. Cannazza, A. Cerrato, C. Citti, C.M. Montone, S. Piovesana, A. Laganà, New insights in hemp chemical composition: a comprehensive polar lipidome characterization by combining solid phase enrichment, high-resolution mass spectrometry, and cheminformatics, Anal. Bioanal. Chem. 412 (2020) 413–423, https://doi.org/10.1007/s00216-019-02247-6.
[19] S. Piovesana, S.E. Aita, G. Cannazza, A.L. Capriotti, C. Cavaliere, A. Cerrato, P. Guarnaccia, C.M. Montone, A. Laganà, In-depth cannabis fatty acid profiling by ultra-high performance liquid chromatography coupled to high resolution mass spectrometry, Talanta 228 (2021), 122249, https://doi.org/10.1016/j.talanta.2021.122249.
[20] K.S. Ladha, P. Ajrawat, Y. Yang, H. Clarke, Understanding the medical chemistry of the cannabis plant is critical to guiding real world clinical evidence, Molecules 25 (2020) 4042, https://doi.org/10.3390/molecules25184042.
[21] 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.
[22] J.T. Fischedick, A. Hazekamp, T. Erkelens, Y.H. Choi, R. Verpoorte, Metabolic fingerprinting of Cannabis sativa L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes, Phytochemistry 71 (2010) 2058–2073, https://doi.org/10.1016/j.phytochem.2010.10.001.
[23] A. Hazekamp, J.T. Fischedick, Cannabis - from cultivar to chemovar, Drug Test. Anal. 4 (2012) 660–667, https://doi.org/10.1002/dta.407.
[24] 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.
[25] R.Z. Al Bakain, Y.S. Al-Degs, J.V. Cizdziel, M.A. Elsohly, Comprehensive classification of USA cannabis samples based on chemical profiles of major cannabinoids and terpenoids, J. Liq. Chromatogr. Relat. Technol. (2020), https://doi.org/10.1080/10826076.2019.1701015.
[26] J.T. Fischedick, Identification of terpenoid chemotypes among high (−)- trans -Δ 9 - tetrahydrocannabinol-producing cannabis sativa L. Cultivars, Cannabis Cannabinoid Res 2 (2017) 34–47, https://doi.org/10.1089/can.2016.0040.
[27] L. Perez de Souza, S. Alseekh, F. Scossa, A.R. Fernie, Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research, Nat. Methods 18 (2021) 733–746, https://doi.org/10.1038/s41592-021-01116-4.
[28] P.G. Vásquez-Ocmín, G. Marti, M. Bonhomme, F. Mathis, S. Fournier, S. Bertani, A. Maciuk, Cannabinoids vs. whole metabolome: relevance of cannabinomics in analyzing Cannabis varieties, Anal. Chim. Acta 1184 (2021), 339020, https://doi.org/10.1016/j.aca.2021.339020.
[29] M.C. Monti, P. Frei, S. Weber, E. Scheurer, K. Mercer-Chalmers-Bender, Beyond Δ9-tetrahydrocannabinol and cannabidiol: chemical differentiation of cannabis varieties applying targeted and untargeted analysis, Anal. Bioanal. Chem. 414 (2022) 3847–3862, https://doi.org/10.1007/s00216-022-04026-2.
[30] 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 230 (2021), 122313, https://doi.org/10.1016/j.talanta.2021.122313.
[31] M. Rehman, S. Fahad, G. Du, X. Cheng, Y. Yang, K. Tang, L. Liu, F.-H. Liu, G. Deng, Evaluation of hemp (Cannabis sativa L.) as an industrial crop: a review, Environ. Sci. Pollut. Res. 28 (2021) 52832–52843, https://doi.org/10.1007/s11356-021-16264-5.
[32] J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: analysis of multivariate data obtained from an experimental design, J. Chemom. 19 (2005) 469–481, https://doi.org/10.1002/cem.952.
[33] Cannabis Flos, New text of the German Pharmacopoeia, Bonn, Ger, 2018.
[34] F. Pellati, V. Brighenti, J. Sperlea, L. Marchetti, D. Bertelli, S. Benvenuti, New methods for the comprehensive analysis of bioactive compounds in cannabis sativa L. (hemp), Molecules 23 (2018) 2639, https://doi.org/10.3390/molecules23102639.
[35] J.A. Kirwan, H. Gika, R.D. Beger, D. Bearden, W.B. Dunn, R. Goodacre, G. Theodoridis, M. Witting, L.-R. Yu, I.D. Wilson, Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management, Metabolomics 18 (2022) 70, https://doi.org/10.1007/s11306-022-01926-3.
[36] F. Tolomeo, F. Russo, D. Kaczorova, M.A. Vandelli, G. Biagini, A. Laganà, A. L. Capriotti, R. Paris, F. Fulvio, L. Carbone, E. Perrone, G. Gigli, G. Cannazza, C. Citti, Cis-Δ9-tetrahydrocannabinolic acid occurrence in Cannabis sativa L, J. Pharm. Biomed. Anal. 219 (2022), 114958, https://doi.org/10.1016/j.jpba.2022.114958.
[37] F. Tolomeo, F. Russo, M.A. Vandelli, G. Biagini, A.L. Capriotti, A. Laganà, L. Carbone, G. Gigli, G. Cannazza, C. Citti, HPLC-UV-HRMS analysis of cannabigerovarin and cannabigerobutol, the two impurities of cannabigerol extracted from hemp, J. Pharm. Biomed. Anal. 203 (2021), 114215, https://doi.org/10.1016/j.jpba.2021.114215.
[38] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ. Sci. Technol. 48 (2014) 2097–2098, https://doi.org/10.1021/es5002105.
[39] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.-J.A.N. Lamers, J. van der Greef, M. E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (2005) 3043–3048, https://doi.org/10.1093/bioinformatics/bti476.
[40] D.J. Vis, J.A. Westerhuis, A.K. Smilde, J. van der Greef, Statistical validation of megavariate effects in ASCA, BMC Bioinf. 8 (2007) 322, https://doi.org/10.1186/1471-2105-8-322.
[41] M. Anderson, C. Ter Braak, Permutation tests for multi-factorial analysis of variance, J. Stat. Comput. Simulat. 73 (2003) 85–113, https://doi.org/10.1080/00949650215733.
[42] G. Zwanenburg, H.C.J. Hoefsloot, J.A. Westerhuis, J.J. Jansen, A.K. Smilde, ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison, J. Chemom. 25 (2011) 561–567, https://doi.org/10.1002/cem.1400.
[43] M.E. Timmerman, H.A.L. Kiers, A.K. Smilde, E. Ceulemans, J. Stouten, Bootstrap confidence intervals in multi-level simultaneous component analysis, Br. J. Math. Stat. Psychol. 62 (2009) 299–318, https://doi.org/10.1348/000711007X265894.
[44] M.E. Timmerman, H.A.L. Kiers, A.K. Smilde, Estimating confidence intervals for principal component loadings: a comparison between the bootstrap and asymptotic

results, Br. J. Math. Stat. Psychol. 60 (2007) 295–314, https://doi.org/10.1348/000711006X109636.
[45] D. Caprioglio, H.I.M. Amin, O. Taglialatela-Scafati, E. Muñoz, G. Appendino, Minor Phytocannabinoids, A misleading name but a promising opportunity for biomedical research, Biomolecules 12 (2022) 1084, https://doi.org/10.3390/biom12081084.
[46] P. Liigand, K. Kaupmees, K. Haav, J. Liigand, I. Leito, M. Girod, R. Antoine, A. Kruve, Think negative: finding the best electrospray ionization/MS mode for your analyte, Anal. Chem. 89 (2017) 5665–5668, https://doi.org/10.1021/acs.analchem.7b00096.
[47] R.M. Salek, C. Steinbeck, M.R. Viant, R. Goodacre, W.B. Dunn, The role of reporting standards for metabolite annotation and identification in metabolomic studies, GigaScience 2 (2013) 13, https://doi.org/10.1186/2047-217X-2-13.
[48] B.M. Lange, J.J. Zager, Comprehensive inventory of cannabinoids in Cannabis sativa L.: can we connect genotype and chemotype? Phytochem. Rev. 21 (2022) 1273–1313, https://doi.org/10.1007/s11101-021-09780-2.
[49] 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.
[50] M.L. Barrett, D. Gordon, F.J. Evans, Isolation from cannabis sativa L. of cannflavin-a novel inhibitor of prostaglandin production, Biochem. Pharmacol. (1985), https://doi.org/10.1016/0006-2952(85)90325-9.
[51] M.N. Clifford, S. Knight, N. Kuhnert, Discriminating between the six isomers of dicaffeoylquinic acid by LC-MS n, J. Agric. Food Chem. 53 (2005) 3821–3832, https://doi.org/10.1021/jf050046h.
[52] J.L. Willems, M.M. Khamis, W. Mohammed Saeid, R.W. Purves, G. Katselis, N. H. Low, A. El-Aneed, Analysis of a series of chlorogenic acid isomers using differential ion mobility and tandem mass spectrometry, Anal. Chim. Acta 933 (2016) 164–174, https://doi.org/10.1016/j.aca.2016.05.041.
[53] S. Kumar, A. Yadav, M. Yadav, J.P. Yadav, Effect of climate change on phytochemical diversity, total phenolic content and in vitro antioxidant activity of Aloe vera (L.) Burm.f, BMC Res. Notes 10 (2017) 60, https://doi.org/10.1186/s13104-017-2385-3.
[54] S. Kabtni, D. Sdouga, I. Bettaib Rebey, M. Save, N. Trifi-Farah, M.-L. Fauconnier, S. Marghali, Influence of climate variation on phenolic composition and antioxidant capacity of Medicago minima populations, Sci. Rep. 10 (2020) 8293, https://doi.org/10.1038/s41598-020-65160-4.
[55] E. De Meijer, Fibre hemp cultivars: a survey of origin, ancestry, availability and brief agronomic characteristics, J. Int. Hemp Assoc. (1995).