Analytica Chimica Acta 1279 (2023) 341848

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Exploratory optimisation of a LC-HRMS based analytical method for untargeted metabolomic screening of *Cannabis Sativa* L. through Data Mining

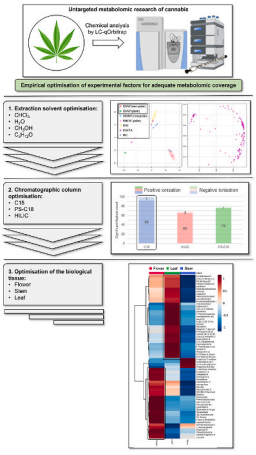
M. San Nicolas [a,b,c,*], A. Villate [a,b], M. Olivares [a,b], N. Etxebarria [a,b], O. Zuloaga [a,b], O. Aizpurua-Olaizola [c], A. Usobiaga [a,b]

[a] Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), 48940, Leioa, Basque Country, Spain
[b] Research Centre for Experimental Marine Biology and Biotechnology (PIE), University of the Basque Country (UPV/EHU), 48620, Plentzia, Basque Country, Spain
[c] Sovereign Fields S.L., 20006, San Sebastian, Basque Country, Spain

HIGHLIGHTS

- Empirical multistep optimisation for untargeted metabolomics study of cannabis.
- Solvent combinations inducing phase separation provide increased metabolite extraction capacity.
- The reverse-phase chromatographic column provided greater metabolic coverage while keeping good separation of the major polar compounds.
- Representative and complementary metabolic information was found in flower and leaf tissues, in contrast to the stem.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Plant metabolomics
Experimental factors
Cannabis sativa L.
LC-HRMS
Metabolomic coverage
Data mining

ABSTRACT

Background: Recent increase in public acceptance of cannabis as a natural medical alternative for certain neurological pathologies has led to its approval in different regions of the world. However, due to its previous illegal background, little research has been conducted around its biochemical insights. Therefore, in the current framework, metabolomics may be a suitable approach for deepening the knowledge around this plant species. Nevertheless, experimental methods in metabolomics must be carefully handled, as slight modifications can lead to metabolomic coverage loss. Hence, the main objective of this work was to optimise an analytical method for appropriate untargeted metabolomic screening of cannabis.

Results: We present an empirically optimised experimental procedure through which the broadest metabolomic coverage was obtained, in which extraction solvents for metabolite isolation, chromatographic columns for LC-

* Corresponding author. Barrio Sarriena s/n, 48940, Leioa, Vizcaya, Spain.
E-mail address: markel.sannicolas@ehu.eus (M. San Nicolas).

https://doi.org/10.1016/j.aca.2023.341848
Received 27 June 2023; Received in revised form 22 September 2023; Accepted 23 September 2023
Available online 27 September 2023

qOrbitrap analysis and plant-representative biological tissues were compared. By exploratory means, it was determined that the solvent combination composed of $CHCl_3$:H_2O:CH_3OH (2:1:1, v/v) provided the highest number of features from diverse chemical classes, as it was a two-phase extractant. In addition, a reverse phase 2.6 μm C18 100 Å (150 × 3 mm) chromatographic column was determined as the appropriate choice for adequate separation and further detection of the diverse metabolite classes. Apart from that, overall chromatographic peak quality provided by each column was observed and the need for batch correction methods through quality control (QC) samples was confirmed. At last, leaf and flower tissues resulted to provide complementary metabolic information of the plant, to the detriment of stem tissue, which resulted to be negligible. *Significance:* It was concluded that the optimised experimental procedure could significantly ease the path for future research works related to cannabis metabolomics by LC-HRMS means, as the work was based on previous plant metabolomics literature. Furthermore, it is crucial to highlight that an optimal analytical method can vary depending on the main objective of the research, as changes in the experimental factors can lead to different outcomes, regardless of whether the results are better or worse.

1. Introduction

Up to current times, *Cannabis Sativa* L. has been an unauthorised plant species in most countries of the world [1]. As a consequence, little research on its insights has been carried out in the past. However, in the USA, for instance, 28 states have already passed the Medical Marijuana Laws (MML) [2], and even though state and Federal laws currently appear to be in conflict, this outstands an increasing tendency of cannabis application for medical purposes, giving a slight notion of how widespread its use could be in the future. Due to this fact, literature around the inner biochemistry of cannabis is slowly expanding, since this plant species warrants further investigation, on behalf of further research in medicine of natural provenance. Thus, early documented research works in cannabis focused in the characterisation of its bioactive compounds, cannabinoids and terpenes [3–6], which seem to have a promising future in palliative medicine [7–11], nonetheless, as these compounds are naturally biosynthesised in the cannabis plant, current studies are also trying to document the distinctive features of its inner metabolic pathways. This way, some primary metabolites as amino acids, carbohydrates or organic acids have already been studied for knowledge deepening into the plant's behavioural response mechanisms or chemovars differentiation, whereas some secondary metabolites as alkaloids, plant hormones or phytosteroids have been acknowledged for participating in signalling mechanisms [12–21]. Nevertheless, established analytical procedures for the ensuring of quality control with full transferability of results are still lacking.

In this thread, "omics"-related studies occupy one of the most noted positions among current analytical trends, as the deconstruction of molecular mechanisms which make up a biological organism can help to totally comprehend its inner functioning [22]. Moreover, in the metabolome, the last downstream rung of the referred "ome"-s (genome, transcriptome, proteome and metabolome) [23], the fingerprint of both internal and external perturbations can be tracked, making metabolomics a very helpful tool to understand the minutiae of the environment that surrounds a living being [24]. Due to this fact, metabolomics is broadly applied for searching biomarkers in living organisms [25]. Nevertheless, in the plant kingdom, specially, the metabolic diversity is vast, as beyond 200,000 different metabolites can compose the metabolome of a plant individual [26]. Therefore, innovative biotechnological approaches are necessary, for instance, in projects focusing cultivations' adaptability to uneasy surroundings [27], and the application of metabolomics could be a major upgrade [28]. Nevertheless, unless proper analytical methods are applied for biomarkers search, metabolites of interest might be missed or misidentified.

In metabolomics, the dominating analytical platform is mass spectrometry, commonly coupled to a chromatographic system [29]. Among such separation techniques, liquid chromatography (LC) is the most widely used one, as it is a versatile tool for both polar and non-polar metabolite analysis, while gas chromatography (GC) is used for the analysis of volatile metabolites and primary metabolites after derivatisation [29]. In regard to mass spectrometry detection, high- and low-resolution platforms can be distinguished. Broadly speaking, low-resolution platforms (e.g., tandem mass spectrometry, MS/MS) are usually addressed for quantitative target metabolomics, addressing metabolites belonging to certain metabolic pathways, while high-resolution mass-spectrometry (e.g., Orbitrap) is used for unknown metabolite identification in non-targeted metabolomics, providing a broader scope concerning coverage of metabolic pathways [30]. Hence, LC coupled to High Resolution Mass Spectrometry (LC-HRMS) offers versatility and high resolution for an accurate description of the metabolome [31]. Nevertheless, this approach is never truly unbiased, as different factors or parameters must be defined prior to analysis, such as stationary phases or ionisation modes [30]. Consequently, properly developed analytical methods should be applied when performing metabolomics research.

A properly developed untargeted analytical method should primarily offer a broad metabolomic coverage. As stated, different factors of the whole experimental procedure must be optimised, in order to build a robust analytical method to be applied in future works related to this field. In this regard, works related to methodological optimisation actually exist in the field of metabolomics. For instance, there are examples in the literature focussed on the optimal selection of extraction solvents for the determination of as many metabolites as possible in the exploratory analysis of human fluids [32] or the standardisation of the derivatisation step of volatile metabolites in biological samples [33]. Indeed, this last optimisation approach was later applied in plant metabolic profiling studies [34]. Furthermore, standardised protocols for plant metabolomics have been previously developed based on the objective of covering as many metabolites as possible [35–38].

Among the experimental factors affecting the metabolite covering scope in cannabis, solvents for efficient metabolite extraction should be considered. Moreover, if the analytical platform of choice is LC-HRMS, the chromatographic stationary phase should also be taken into account, as efficient separation between compounds would be necessary. In addition, it is stated that plants possess a complex metabolome, thus, exploring the metabolic information in different biological tissues would also be a requirement, as in metabolomic studies organism-representative information should be retrieved.

Bearing all that in mind, multivariate data mining techniques, such as Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA), could be the suitable path to follow for corresponding experimental factor optimisation in the metabolome analysis of cannabis. Both techniques are unsupervised techniques, meaning the variance within the actual data of detected metabolites in the analysed samples is explored, excluding a prediction outcome [39]. Concretely, using those data mining techniques, trends explaining the variance of large datasets, which might not be noticed at first glance due to their dimensionality, can be spotted and relevant information extracted [40]. These are Thus, for factor optimisation, the applied selection criterion was based on the explained variance by the subjects under study. In this work, experimental factors such as extraction solvents, chromatographic columns and biological tissues were studied for appropriate untargeted

metabolomic screening of *Cannabis Sativa* L.

2. Materials and methods

2.1. Reagents and solvents

Solvents for chromatographic analysis, acetonitrile (ACN), methanol (MeOH) and water (H_2O) of HPLC grade, were acquired in Panreac/AppliChem (ITW Reagents, S.R.L., Italy). Formic acid (FA) for LC-MS was purchased from Fischer Chemicals (Thermo Fisher Scientific Inc., USA) and ammonium acetate (NH_4Ac, for molecular biology ≥98%) was acquired in Sigma Aldrich (Merck Group, USA). Calibration of the qOrbitrap was performed with Pierce™ LTQ Velos ESI Positive Ion Calibration Solution and Pierce™ ESI Negative Ion Calibration Solution by Thermo Scientific (Thermo Fisher Scientific Inc., USA).

For sample extraction, MeOH (anhydrous), trichloromethane ($CHCl_3$, synthesis grade) and methyl *tert*-butyl ether (MTBE, HPLC grade) were obtained in Macron Fine (Avantor, Inc., USA), Scharlau (Scharlab S.L., Spain) and Panreac/AppliChem (ITW Reagents S.R.L., Italy), respectively. Water was filtered to a resistivity of 18.2 MΩ cm @ 25 °C and a total organic content (TOC) less than 3 ng mL^{-1} using a Q-POD water dispenser and a Millipak express 40 (0.22 μm filter) by Merck (Merck Group, USA).

2.2. Plant samples

The samples were harvested from cannabis plants cultivated in the facilities of Sovereign Fields (Sovereign Fields S.L., Spain). The plants were cultivated indoor, in an isolated 24 m^2 (6 m × 4 m) growth room, built inside the greenhouse. Inside the room, 6 Ceramic Metal Halide (CMH) lamps (315 W/37000 lm) by Lumatek (Lumatek Ltd., UK) were installed in the roof, at 2.10 m height, uniformly distributed through the room, coupled to Adjust-A-Wings large enforcer reflectors.

Two similar cultivations were run for the optimisation procedure. In the first one, just one plant was harvested for the first two experiments described in section 2.4, while in the second one, 10 plants were harvested for the third experiment described in section 2.4. A chemotype III cannabis cultivar was cultivated in both ($C_{total\ THC}/C_{total\ CBD} < 0.1$) [41].

Plants were grown in 11-L black pots containing a soil/hummus/nutrient mixture. Specifically, the mixture was composed of 80% of Light mix soil of Biobizz (Biobizz Worldwide S.L., Spain), 20% of hummus, and 10 g/L of farmer mix nutrient solution by Lurpe (Lurpe Natural Solutions, Spain), which is composed of bat guano, bone meal, kelp meal, Azomite®, organic alfalfa, insect frass, blood meal, dolomite, langbeinite humic and fulvic acids, and a complex blend of rhizobacteria and Trichoderma. The total cultivation time in both cases was 12 weeks, being the first 4 weeks the vegetative stage during which the plants grew, and the next 8 weeks the flowering stage. The vegetative and flowering stages were defined by the photoperiod regime. In the first 4 weeks of growth, the photoperiod was 18 h light/6 h dark, and in the flowering stage, 12 h light/12 h dark. Moreover, the lamps were set to work at 50% and 80% of the total lamp intensity during the vegetative and flowering stages, respectively. During the cultivation time, the room temperature was kept between 22 and 25 °C during the daytime, and it did not decrease below 18 °C at night. The relative humidity of the room was controlled at 60% (±5%).

Finally, when the cultivation time elapsed, the samples were harvested and instantly frozen in liquid nitrogen, following the quenching method. In the first cultivation, a flower sample was taken from a single plant, while in the second cultivation, leaf, flower, and stem samples were collected from each individual (n = 10). The collected samples had to be representative of the whole plant individual, as each one of them was a biological replicate. So, the following sampling design was defined:

- Leaf samples: A pool of a total 12 leaves from each plant individual were collected, 6 of them from the upper half part of the plant, and the 6 from the lower half, all of them uniformly distributed across the plant.
- Flower samples: Samples from each biological replicate were composed of 5 different flowers. The first flower sample was taken from the apical end of the plant, in the upper point of it. The second one was collected 30 cm lower than this point, next to the stem. From this point, the two nearest ramifications, which were opposite to each other, were followed until the end-point of them, and the flowers situated there were collected. Finally, the lowest flower situated next to the stem was collected. The collected 5 flowers were pooled to a representative flower sample of each plant individual.
- Stem sample: The lowest point of the plant stem was set as the reference point. 5 cm above the reference point, the stem was horizontally cut, cutting down the whole plant individual. From this point, a 10 cm long stem piece was taken as a sample. The possible ramifications on the stem sample were cut down, thus, only remaining the principal stem.

Afterwards, the samples were kept at −80 °C temperature prior to analysis.

2.3. General workflow

Fig. 1 shows the general workflow followed to achieve optimal parameters of an untargeted analytical method for the determination of as many metabolites as possible in *Cannabis Sativa* L.

The optimisation procedure was divided in three main steps.

The first step was the assessment of the extraction solvent for representative isolation of the metabolites from the plant matrix. In this context, 5 different solvents (i. e., methanol-water, methanol-water acidified with formic acid, methanol-chloroform, chloroform-methanol-water, methyl *tert*-butyl ether-methanol-water) and literature-based extraction protocols were compared on the same flower sample, using the same analysis conditions such as chromatographic separation and detection setup. Since most of the literature works regarding plant metabolomics use reverse phase liquid chromatography conditions, the extracts were analysed using a reverse phase column (i. e., 2.6 μm C18 100 Å (150 × 3 mm) column). Once the extraction protocol was set, the second step was the evaluation of three chromatographic columns (i.e., 2.6 μm C18 100 Å (150 × 3 mm), 2.6 μm PS-C18100 Å (150 × 3 mm) and BEH 1.7 μm Amide Hydrophilic Interaction Liquid Chromatography (100 × 2.1 mm) columns). In the last step, a small-scale cultivation experiment was designed for the metabolomic-like study of 10 plant-clones, in which leaf, flower and stem tissues of the plant were analysed using the conditions set in the previous experiments.

2.4. Experimental procedure

Prior to any extraction the frozen samples were ground by Spex SamplePrep Freezer Mill 6770 (Thermo Fisher Scientific Inc. USA), using liquid nitrogen, which worked under the following conditions: 1 min of sample precooling followed by 1 min grinding cycle at 8 cycles per second (cps). The result was a homogeneous-powder-frozen sample.

As previously mentioned, in the first optimisation step, 5 different literature-based metabolite extraction solvents (see Table 1) were compared to choose the extraction solvent or solvent combination that provides the broadest metabolomic coverage of cannabis. The extractions were carried out in a Precellys/Cryollys homogeniser system (Bertin technologies, France). 100 mg ground homogeneous flower sample was weighed, 8 ZrO beads (1.4 mm diameter) were added to each vial and 1.2 mL of the tested solvent were added according to the protocols specified in Table 1. All the assays were carried out in triplicate.

The extractions were performed at a 6400-rpm rate, with 3 cycles of

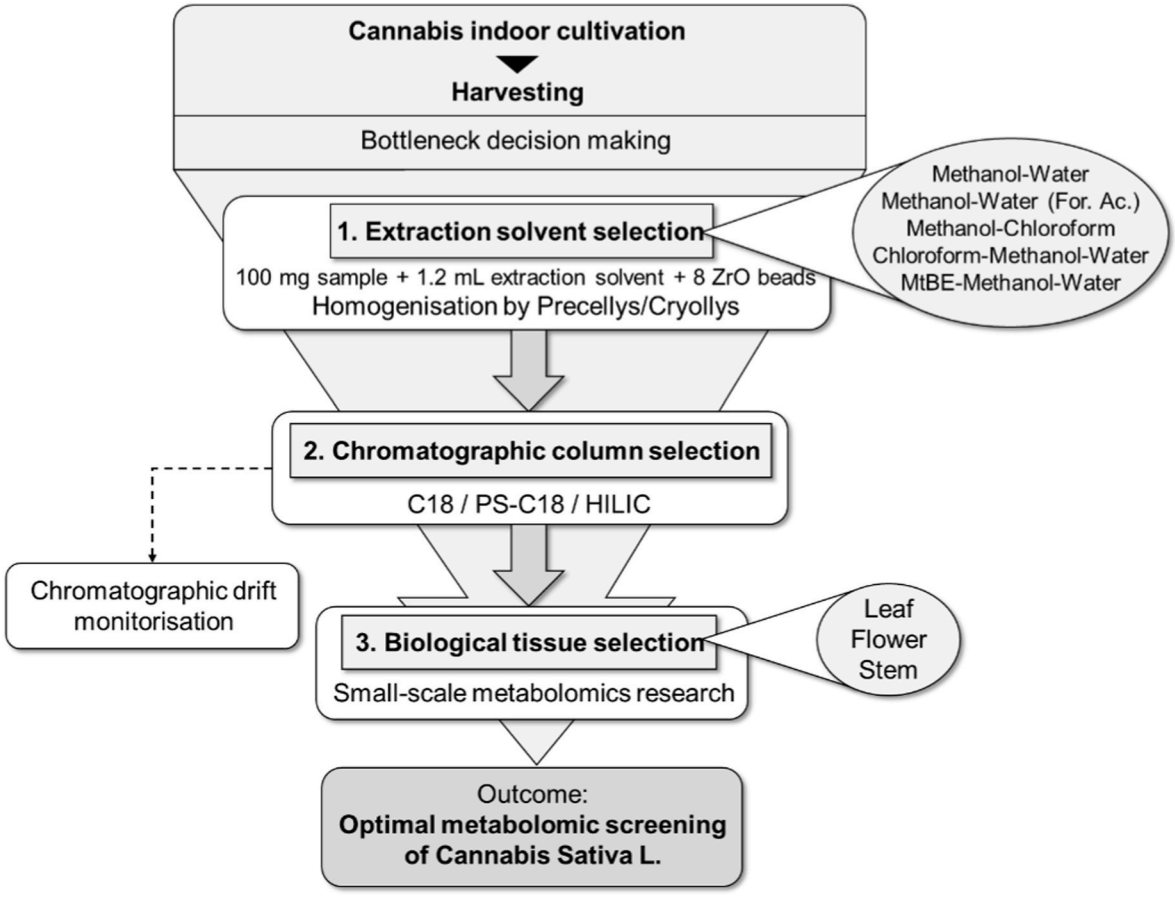


Fig. 1. General experimental workflow overview.

Table 1
Studied extraction-solvent combinations for suspect screening of Cannabis.

Extraction		Solvents (%, v/v)				Reference
		H_2O	MeOH	$CHCl_3$	MTBE	
1	MW (Single-phase)	20	80	–	–	[42]
2	MWFA (Single-phase)	20 (0.1% FA)[a]	80	–	–	[35]
3	MC (Single-phase)	–	80	20	–	[30]
4	CMW (Two-phases)	25	25	50	–	[43]
5	MtMW (Two-phases)	25	–	25	50	[44]

[a] The extraction procedure Nº 2 is equal to the Nº 1, but acidified at 0.1% (v/v) with FA. Acronyms: M: Methanol; W: Water; FA: Formic Acid; C: Chloroform; Mt: Methyl *tert*-Butyl Ether.

60 s run and 15 s rest between each run (6400-3x60-15). The temperature was kept at 4 °C with liquid nitrogen during the extraction time. In study cases Nº 1, Nº 2 and Nº 3 (see Table 1), the respective mixture of solvents resulted in total miscibility at the specified contents. On the contrary, in study cases Nº 4 and Nº 5 (see Table 1), the contents of their main non-polar solvents ($CHCl_3$ and MTBE, respectively) caused their immiscibility with the H_2O:MeOH mixture. As a consequence, phase separation occurred in the later cases, where 50% (v/v) of the total extractant consisted of the non-polar phase ($CHCl_3$ in case Nº 4 and MTBE in case Nº 5) and the other 50% (v/v) consisted of the polar phase (H_2O:MeOH, 1:1 v/v). In the single-phase extractant cases (cases Nº 1, Nº 2 and Nº 3), the respective solvents were added in order of polarity, from the least polar to the most polar. In the case of the two-phase extractions (i.e., Nº 4 and Nº 5, see Table 1), the procedure was performed in two steps. First, the sample was treated with 300 μL MeOH and 300 μL H_2O (the polar phase), and after one extraction run, 600 μL of the corresponding main non-polar solvent ($CHCl_3$ or MTBE) were subsequently added for sequential extraction of the sample.

Once the extraction was completed, the vials were centrifuged in the Allegra X–30R centrifuge (Beckman Coulter, USA) for 5 min, at 21,000 g force and 4 °C. The supernatants were decanted to another 2 mL vial for a second centrifugation run (15 min at 21,000 g and 4 °C). At last, prior to analysis, the polar and non-polar extracts were separated in the case of two-phase extractions, and they were transferred to chromatographic vials and diluted at a 1:20 ratio in MeOH, as well as the single-phase extracts. All the extracts in these assays were analysed by means of LC-qOrbitrap using a reverse phase Kinetex 2.6 μm C18 100 Å (150 × 3 mm) with AJ0-8782 C-18 pre-column (Phenomenex, USA), as detailed in section 2.5.

Having set the extraction procedure, the appropriate chromatographic column was selected through exploratory analysis of a flower sample. The compared columns were: i) the previously used reverse phase Kinetex 2.6 μm C18 100 Å (150 × 3 mm) with AJ0-8782 C-18 pre-column (Phenomenex, USA) for enhanced hydrophobic retention, ii) an aqueous stable with positive surface charged reverse phase Kinetex 2.6 μm PS-C18 100 Å (150 × 3 mm) with AJ0-8950 PS-C18 pre-column (Phenomenex, USA) for improved hydrophilic affinity, and, iii) a normal phase Acquity UPLC BEH 1.7 μm Amide Hydrophilic Interaction Liquid Chromatography (HILIC) 130 Å (100 × 2.1 mm) with Acquity UPLC BEH 1.7 μm Amide Vanguard 130 Å (2.1 × 5 mm) pre-column (Waters, USA) for enhanced separation and selectivity of polar compounds.

Apart from that, the chromatographic drift and signal fluctuation of

the most suitable column sequence were explored across the analysis sequence. For doing so, three replicates of the flower sample were randomly analysed 10 times each, and a pooled sample was periodically analysed between them, serving as a quality control (QC) reference. Variances of the samples were explored throughout the analysis sequence, comparing the results obtained with QC correction and without correction.

In the third optimisation step, as the extractant and the chromatographic column are set up, the small-scale metabolomic study of the 10 plant-clones from the second cultivation (see 2.2. Plant samples section) was conducted for representative biological tissue exploration, comparing leaf, stem and flower samples.

2.5. Chromatographic methods and general MS settings

Untargeted analysis in the plant extracts was performed by Thermo Scientific Dionex Ultimate 3000 liquid chromatograph coupled to a Thermo Scientific Q Exactive Focus quadrupole-Orbitrap mass spectrometer (UHPLC-q-Orbitrap), equipped with a heated electrospray ionisation source (HESI, Thermo-Fisher Scientific, CA, USA). Analysis were performed with 5 μL injection from each extract in both ionisation modes, positive and negative, with the automatic injector (at 5 °C). To avoid any carryover, the injection needle was cleaned before and after every injection with 50 μL of H_2O:MeOH (90:10, v/v).

Regardless of the column used, chromatographic analysis consisted of two mobile phases, water (solvent A) as polar solvent and ACN (solvent B) as non-polar solvent, both of them containing either 0.1% (v/v) of FA or 10 mM NH_4Ac, for respective measurements in positive or negative ionisation.

In the case of the chromatographic separation using the C18 column, the eluent flow rate was kept at 0.3 mL/min. During the analysis time, the solvent gradient started at 95% A (held for 1 min), then linearly decreased to 5% An until minute 16 (held for 10 min) and finally returned to the initial conditions in 5 min, where they were maintained for 1 min. Regarding the separation conditions using the PS-C18 column, the eluent flow rate was also 0.3 mL/min and the gradient was stablished as follows. The mobile phase gradient started at 97% A, which was held for the first 2 min, then decreased to 85% An until minute 5, followed by a change to 5% An until minute 11 (held for 14 min), before returning to initial conditions in 4 min (held for 1 min). At last, for the separations using HILIC column, both mobile phase flow and composition were changed. The flow gradient started at 0.2 mL/min (held for 3 min), then it increased to 0.3 mL/min (held for 21 min) and it was again decreased to 0.2 mL/min until minute 35, when the analysis run ended. Regarding the mobile phase composition gradient, the analysis method started at 3% A for the first 3 min, then increased to 15% An until minute 5, changing to 25% An until minute 14 and a final increase to 60% An until minute 17 (held for 3 min), before returning to initial conditions in 3 min (held for 12 min).

Regardless of the ionisation mode, the HESI ionisation source was operated under the following parameters: spray voltage of 3.20 kV, spray current of 0.50 μA, the capillary temperature was kept at 320 °C and S-lens RF level at 55.0, the sheath gas (nitrogen) flow rate at 48 arbitrary units (au), auxiliary gas flow rate at 11 (au), sweep gas flow rate at 2 (au), and auxiliary gas heater temperature at 310 °C. Every three days Pierce LTQ ESI Calibration Solutions were used for external calibration of the instrument.

Data acquisition in the high-resolution mass spectrometer (qOrbitrap) was done in Full MS-Data dependant MS^2 (Full-MS-dd-MS^2) discovery acquisition mode for every analysis of the optimisation process. Full-scan mass spectrum was collected in a scan range between 70 and 1000 *m/z* with a resolution of 70,000 FWHM for an *m/z* of 200. Three additional scans were performed in dd-MS^2 mode with a resolution of 17,500 FWHM, an isolation window of 0.8 *m/z*, and applying an stepped normalized collision energy (nce) of 10%, 35% and 75% in the collision cell, where the first mass fixed was 50.0 *m/z* and intensity of AGC target of $2.00{\cdot}10^5$. The software used was Xcalibur 4.0 (Thermo-Fisher-Scientific).

2.6. Data processing

Data processing was performed using Compound Discoverer 3.3 software by Thermo Fischer. A minimum signal-to-noise ratio (S/N) of 3 was stablished in order to consider a chromatographic peak feasible and subsequently integrate its corresponding under-peak area. In the detected chromatographic features, MS1 was used as precursor for compound detection, with a mass tolerance of 5 ppm after performing 5 scans per feasible peak. Compounds were also grouped across samples with a mass tolerance of 5 ppm and a retention time (RT) tolerance of 0.2 min based on the preferred ions $[M+H]^{+1}$ and $[M-H]^{-1}$ for positive ionisation and negative ionisation, respectively. Moreover, a maximum threshold of 1 min peak width at half peak height was stablished in every detected feature. Additionally, possible gaps were filled with 1.5 S/N ratio and also 5 ppm mass tolerance.

In the study cases of this work, only significant features were considered. To be deemed significant, the detected features had to pass the following filters: greater chromatographic peak area than 1,000,000 intensity counts, less variance than 30% between corresponding replicates and, at least, partial putative spectral match with a minimum single candidate compound among provided spectral fragmentation libraries. These referred libraries were Cannabis Sativa L. endogen suspect list, retrieved from Plant Metabolic Network database [40], LipidMaps structure database [41] and Endogenous metabolites database of 4400 compounds provided by Thermo Scientific. Furthermore, detected features were also filtered according to chromatographic peak quality, which was defined by four criteria: the jaggedness, the zig-zag index, the FWHM2base and the modality of the peaks [45]. Each of these metrics was measured at 5/10, being this value, the representation of its contribution compared to the other parameters contributors. Each measured metric contributed equally to the overall peak quality. Therefore, the peak quality filter threshold was set at 5/10, as only peaks surpassing this overall value would be accepted as significant for the study. For metabolite identification in the significant chromatographic features, mzLogic data analysis algorithm and Mass Frontier 7.0 spectral interpretation software were used, both from Thermo Fisher Scientific. In each of the significant chromatographic features, metabolite candidates from spectral libraries were ranked according to the similarity between experimental and estimated MS2 spectrum and the Fragment Ion Search (FISh) score was calculated in the five candidates with highest similarity score for structural elucidation and putative metabolite identification. The FISh score was calculated with a high accuracy mass tolerance of 2.5 mmu, low accuracy mass tolerance of 0.5 Da and a S/N threshold of 3. Feature annotation was performed according to the confidence levels defined in *Schrimpe-Rutledge* et al. [46]. Since no pure reference standard was used, the highest confidence level was 2 (putative identification, MS2 match). The features fulfilling all those previous defined criteria were later analysed through unsupervised multivariate data-analysis to determine the variances and trends between the analysed samples and annotated significant features. These data-analysis approaches were carried out using Metaboanalyst 5.0 [47–49].

3. Results and discussion

3.1. Extractant selection

To determine the most suitable extractant for optimal metabolomic screening of cannabis, the range of coverage of each extraction solvent combination was explored. Nonetheless, to avoid the assessment of hypothetical false positives, the processed data was filtered according to the constraints stated in section 2.6., so only chromatographic areas of significant features were evaluated through multivariate data analysis. The results are presented in Table S1 of the Supporting Material (SM),

where a total of 171 features were detected in positive ionisation mode and 35 features in negative ionisation mode. The area values were auto scaled and transformed to logarithmic scale (Log_{10}), prior to analysing the data by Principal Component Analysis (PCA), and the trends of greater variance were determined though exploratory means (see Fig. 2).

According to the PCA scores plot, PC1 explained 71.3% and 82.1% of the total variance in the positive and negative ionisation, respectively. Regarding PC1, single-phase extractants (MC, MW and MWFA) seemed to have no relevance at all since most of the variables are related to the antagonist phases (polar and non-polar) of the two-phase extractions (CMW and MtMW) that located at opposed endpoints of PC1. PC2, which explained 17.6% and 9.4% of the total variance in positive and negative ionisations, respectively, seemed to be related with the distinction of single-phase and two-phase extractants. According to these facts, antagonist phase extracts of the two-phase extractants presented greater significance over the total variance in the study, which suggested that a two-phase extractant would be more appropriate for representative metabolomic coverage of cannabis. Indeed, phase differentiation could be the reason for the enhancement of the extraction yield, at the expense of single-phase extractions, as both very polar and very non-polar metabolites could be quantitatively isolated from the

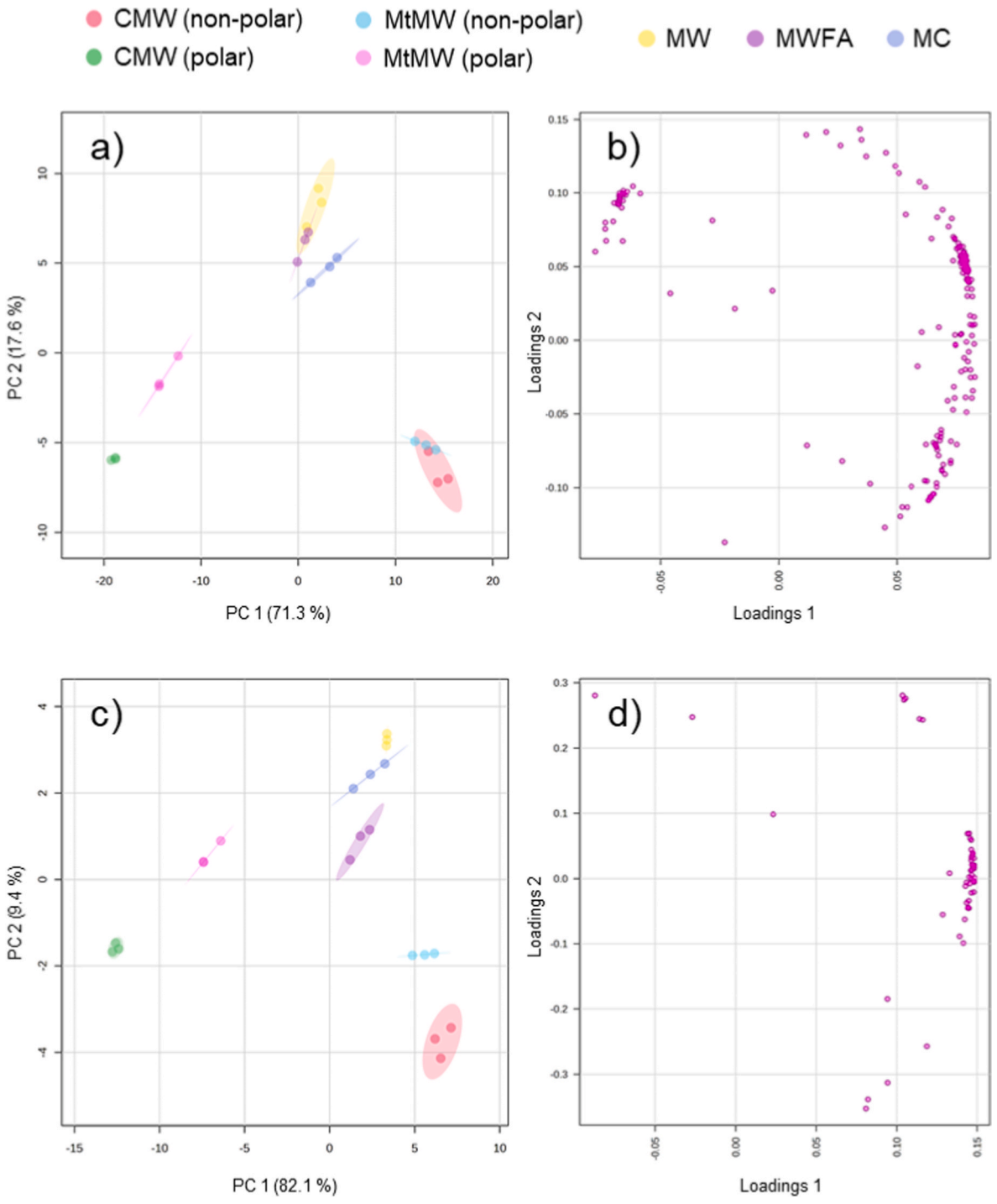


Fig. 2. PCA scores and loadings of the filtered results in the extractant optimisation step a) Scores plot of positive ionisation results b) Loadings plot of positive ionisation results c) Scores plot of negative ionisation results d) Loadings plot of negative ionisation results.

plant matrix.

Going deeper into the results obtained using two-phase extractants, it could be noted that scores corresponding to the extracts of CMW appear to be further than the corresponding MtMW extracts from the null relevance point of the variable space according to PC1, which could indicate a greater stronger statistical influence for the CMW extraction. Nevertheless, at first glance, it was difficult to assert that there is actually a significant difference between both two-phase extractions. Thus, the relative extraction yield of the significant chromatographic features got using two-phase extractants was assessed (see Fig. 3). According to these results, even though the difference among the yield of both extractants was barely significant, in every study case greater average peak area was obtained using CMW as extractant. As the composition of the polar phase was equal in both cases (50% methanol and 50% water, v/v), the difference resided in the corresponding non-polar solvent. Chloroform would present greater extraction yield for non-polar compounds than methyl *tert*-butyl ether, as chloroform possesses greater partition coefficient ($LogP_{CHCl3}$ = 2.3, $LogP_{C5H12O}$ = 0.9). Therefore, theoretically, chloroform presents 2.55 times greater affinity towards non-polar compounds, resulting in an enhanced extraction yield for metabolites of this nature. On the other hand, greater polarity gradient between polar and non-polar phases could also enhance polar compound extraction capability in its corresponding phase, resulting in a greater extraction yield of these metabolites of this nature as well. Hence, based on the exploratory results, the extraction solvent combination composed of chloroform-methanol-water (50 %-25 %–25%, v/v) was considered the most suitable for representative untargeted metabolomics screening of cannabis. It should also be noted that greater count of significant features was detected in positive ionisation analysis (171) than in negative ionisation analysis (35), thus giving the former case greater statistical weight in the exploratory optimisation process.

3.2. Chromatographic column selection

After performing the data processing and filtering of the results of the studied chromatographic conditions as described in section 2.6, obtained results for polar and non-polar extracts are resumed in Table S2 and Table S3 of SM, respectively. Detected feature counts and corresponding areas obtained with each of the studied chromatographic columns are displayed in Fig. 4.

The first fact to note was that, either in positive or negative ionisation modes, greater number of significant peaks were annotated when using the C18 chromatographic column. A total of 195 significant features were detected using the C18 column, whereas the HILIC and the PS-C18 columns yielded 85 and 117 significant features, respectively. Secondly, we observed that, regardless of the chromatographic column used, a higher number of features were detected in the positive ionisation mode, as it happened in the previous optimisation step. At first sight, this fact acknowledges that analysis through positive ionisation mode, regardless of the analytical conditions applied, could provide broader information, as more chromatographic features could be annotated.

In addition to the number of features, the areas of the chromatographic peaks were also a factor to be considered. As expected, the HILIC column provided good separation and large peak areas for characteristic polar compounds, such as amino acids, among others. Concretely, choline (*m/z* 104.1071), proline (*m/z* 116.0707), betaine (*m/z* 118.0863), leucine (*m/z* 132.1018), asparagine (*m/z* 133.0607), trigonelline (*m/z* 138.0549) and glutamine (*m/z* 147.0763) and were detected with the largest peak areas in the positive ionisation mode acquisition, all of them putatively identified (identification confidence level 2), while in the results of negative ionisation mode hydroxy threonine (*m/z* 134.0460), calditol (*m/z* 253.0936) and coutaric acid (*m/z* 295.0457) stood out, identified as tentative structures (identification confidence level 3). On the other hand, the C18 column provided large chromatographic peak areas in the analysis of the non-polar extract, where cannabinoids highlighted, along with lipids and steroids. In this case, the largest peak areas belonged to α-farnesene (*m/z* 205.1948), 6-[(1E,3E,5E,7E,9E,11E)-9,11-dimethyltetradeca-1,3,5,7,9,11-hexaenyl]-5-ethyloxane-2,4-dione (*m/z* 341.2104), cannabidiolic acid (*m/z* 359.2210), armillarin (*m/z* 415.2109) and myxopyronin B (*m/z* 432.2373), putatively identified in positive ionisation (identification

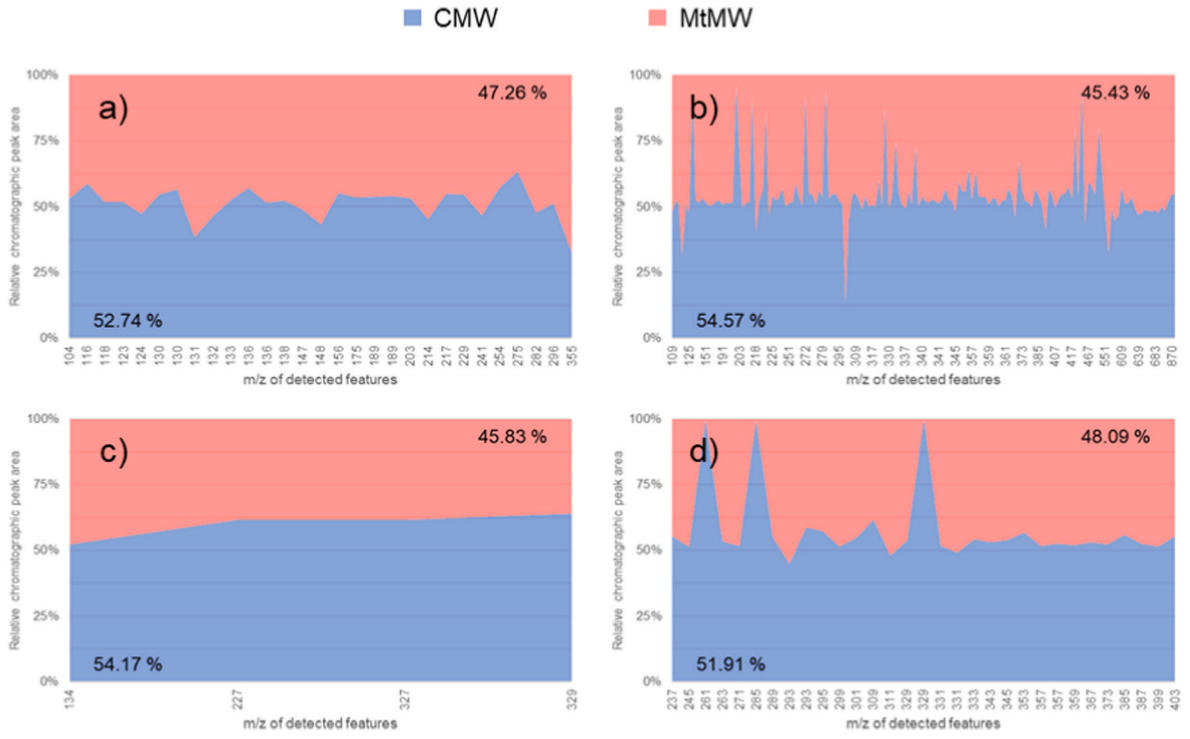


Fig. 3. Relative coverage of the detected significant features of the polar and non-polar extracts of the CMW and MtMW extractants: a) Polar extracts through positive ionisation b) Non-polar extracts through positive ionisation c) Polar extracts through negative ionisation d) Non-polar extracts through negative ionisation.

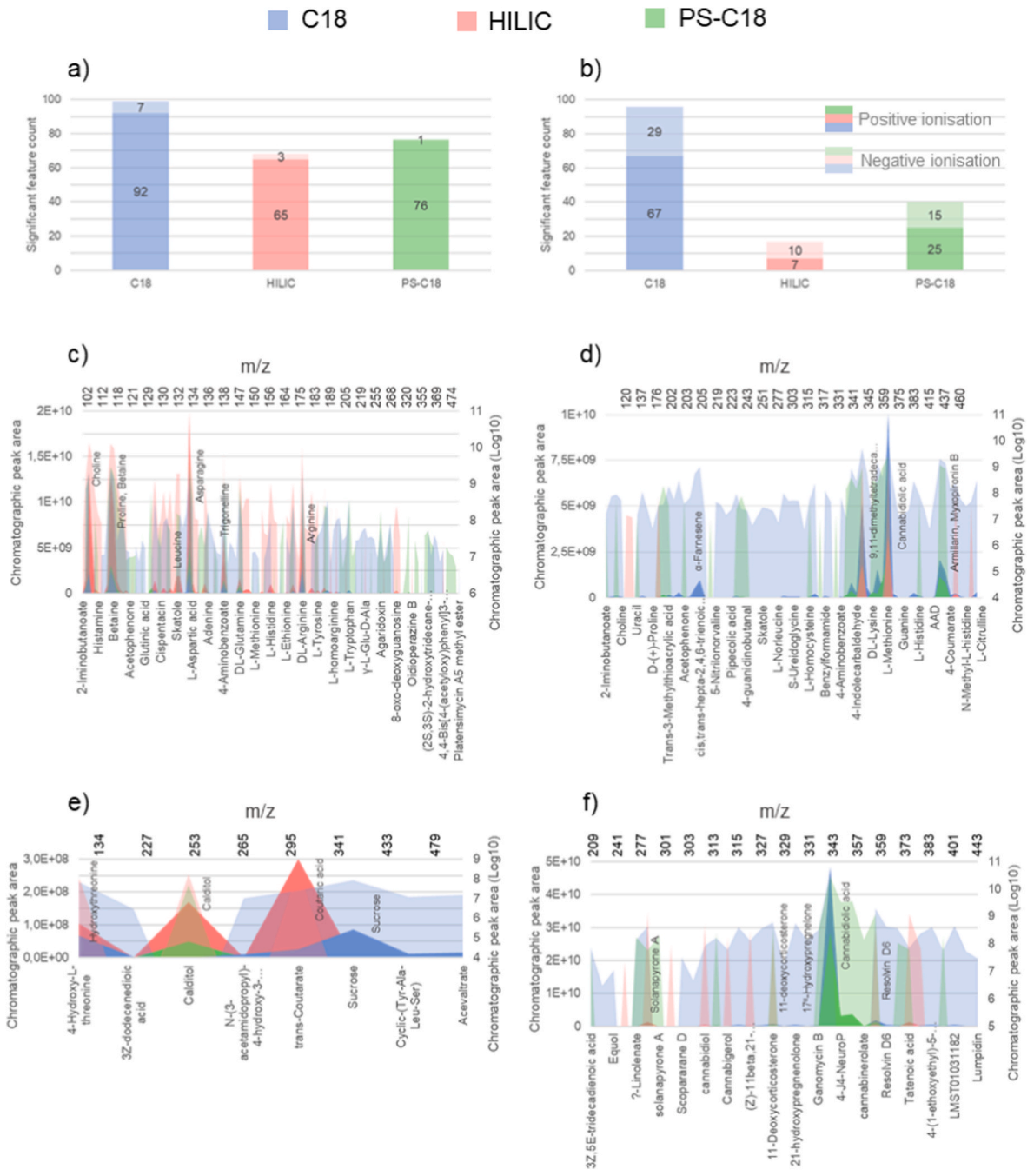


Fig. 4. Number of significant feature count (SFC) detected by each of the chromatographic columns and their corresponding peak areas (CPA). a) SFC in polar phase, positive and negative ionisation modes b) SFC in non-polar phase, positive and negative ionisation modes c) CPA vs SFC in polar phase, positive ionisation mode d) CPA vs SFC in non-polar phase, positive ionisation mode e) CPA vs SFC in polar phase, negative ionisation mode f) CPA vs SFC in non-polar phase, negative ionisation mode.

confidence level 2), while 11-Deoxycorticosterone (*m/z* 329.2123, level 3), 17α-hydroxypregnelone (*m/z* 331.2280, level 3) cannabidiolic acid (*m/z* 357.2071, level 2) and resolvin D6 (*m/z* 359.2230, level 2) highlighted in negative ionisation.

According to Fig. 5, it was also determined that the PS-C18 column, could be deemed as the least relevant among the studied columns to be used in untargeted metabolomics. As its stationary phase relied on a midway polarity nature between the C18 and the HILIC column, it failed to provide more information than the already obtained with the other columns. In particular, the PS-C18 column does not allow the detection of non-polar compounds either in number or type of those already detected by the C18 column, while for the detection of polar compounds it is preferable to use the HILIC column.

Moreover, it is worth mentioning that most of the polar compounds detected in the polar extract using the HILIC were also detected in the measurements using C18 column. Of course, it should not be ignored the fact that the polarity of the reversed-phase column implied a lower retention affinity towards compounds of polar nature, so that both the areas of their corresponding chromatographic peaks and the difference in separation time between them were significantly lower with the C18 column; nevertheless, a higher number of compounds were detected. For instance, some amino acids such as L-methionine (*m/z* 150.0582), aminoadipic acid (*m/z* 162.0760) or L-tryptophan (*m/z* 205.0970), alkaloid derivatives such as nicotinamide (*m/z* 123.0553), 4-Indolecarbaldehyde

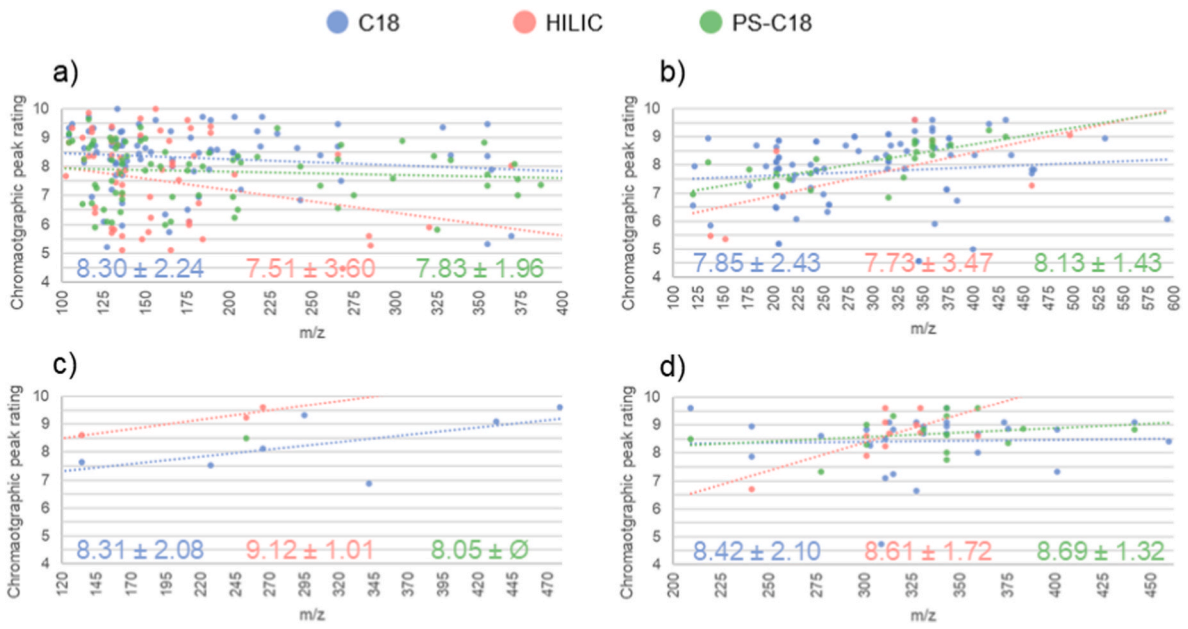


Fig. 5. Peak qualities of the detected significant features and their average values in each of the studied cases, in a 95% confidence interval a) polar phase, positive ionisation mode b) non-polar phase, positive ionisation mode c) polar phase, negative ionisation mode d) non-polar phase, negative ionisation mode.

(*m/z* 146.0599) or *trans*-3-indoleacrylic acid (*m/z* 188.0704), metabolites that play an important role in the inner regulation of the plants (all of them were identified at confidence level 2) were not detected in the polar extract by the HILIC column, instead they were annotated with the C18 column. That did not happen in the case of the analysis of the non-polar extract with the HILIC column, as only 7 and 10 features were detected in positive and negative ionisation modes, respectively.

Apart from the annotated feature number and their corresponding peak areas, their quality should also be a factor to consider. Further data-analysis of the results of untargeted metabolomic research would depend on the obtained peak areas, hence, in order to avoid undue deviations, detected peak should pass a minimum quality threshold, as poor chromatographic peaks may deviate results from their true values. In Fig. 6, peak qualities of the detected features are displayed, and the average peak quality was calculated at 95% confidence for each chromatographic column. The average peak qualities offered by each column were similar in the four study cases shown in the figure. In fact, the internal deviation within each column was greater than the difference between columns, thus, the quality of the detected features was fully comparable in the different study cases. The PS-C18 column offered the smallest deviation in the different study cases, nonetheless, it was previously observed that its contribution, according to the objective of the work, was the least among the three columns. Apart from that, the C18 column had greater statistical weight in the peak quality exploration, as it provided a higher number of significant features.

Hence, based on all those results and observations, analysis through the C18 column is suggested for the intended purpose, bearing in mind that its stationary phase will present greater retention affinity towards non-polar compounds. Therefore, it was concluded that C18 column would be the most suitable for a broader metabolomic coverage.

Moreover, it is known that in metabolomic studies, where large sample sets are analysed in a single run, chromatographic signal drift can occur along the sequence. This phenomenon can lead to misleading conclusions since the chromatographic signals may depend on the acquisition-time. Hence, to ensure accurate data acquisition, the chromatographic area drift was assessed suing the C18 column to define the QC correction effect through the analysis time. The data of unfiltered Log_{10} transformed results can be seen in Fig. 6. This unfiltered raw data is presented in Table S4 (Polar extract, QC corrected data), Table S5 (Polar extract, non-corrected data), Table S6 (Non-polar extract, QC corrected data) and Table S7 (Non-polar extract, non-corrected data) of SM. Since no filter was applied to the detected chromatographic features, the feature annotation in this section was limited to tentative structure of the metabolites (level 3), molecular formula match (level 4) or to just a unique chromatographic feature (level 5), according to the metabolite identification level specified in *Schrimpe-Rutledge* et al. [46].

At first glance, an apparent difference between QC-corrected and non-corrected results can be distinguished. On the one hand, the Compound Discoverer 3.3 software allowed the correction of the retention time (RT) shifts in the alignment of chromatographic peaks across the samples through ChromAlign algorithm [50], having the QC sample as a reference. As it can be seen in the RT shifts correction plots, in some of the cases RT shifts of up to 0.2 min were corrected, from where it could be deduced that the lack of RT shifts correction could also have a significant impact on the subsequent chromatographic peak area integration, thus, also affecting the area values and, therefore, the subsequent data-analysis. On the other hand, it could also be observed that the cumulative non-corrected signal of the QC sample exhibited significant fluctuations through the time span of the analysis sequence, which appeared to be independent of the data acquisition time, as they did not follow a defined trend. Signal correction through the SERRF (Systematical Error Removal using Random Forest) QC normalization method [51], which was the one implemented in the Compound Discoverer 3.3 data processing software, resulted in more constant area signal results in the chromatographic features through sequence time. Hence, these two correction methods led to lower intra-group variance of the data, as it can be seen in the corresponding PCA score plots. In the case of the polar extract, it was clear that correcting the data using QC samples allowed a clustering of replicates with reduced variance. Since the analysed samples were technical replicates of three biological replicates of the same flower sample, ideal results would be expected to present null variance between either of the replicates, so, based on the empirical results, the smaller the variance between replicates, the more reliable the obtained results would be. In the case of the non-polar extract, this was not so evident, as the variance between replicates did not differ so much between QC-corrected and non-corrected data, so it would be difficult to ascertain from just the PCA plots that there was an apparent difference between the two cases. Nevertheless, as stated before, the QC signal fluctuation plot revealed fluctuations in the summed signals of the QC sample in different analysis times. What it could be determined was that

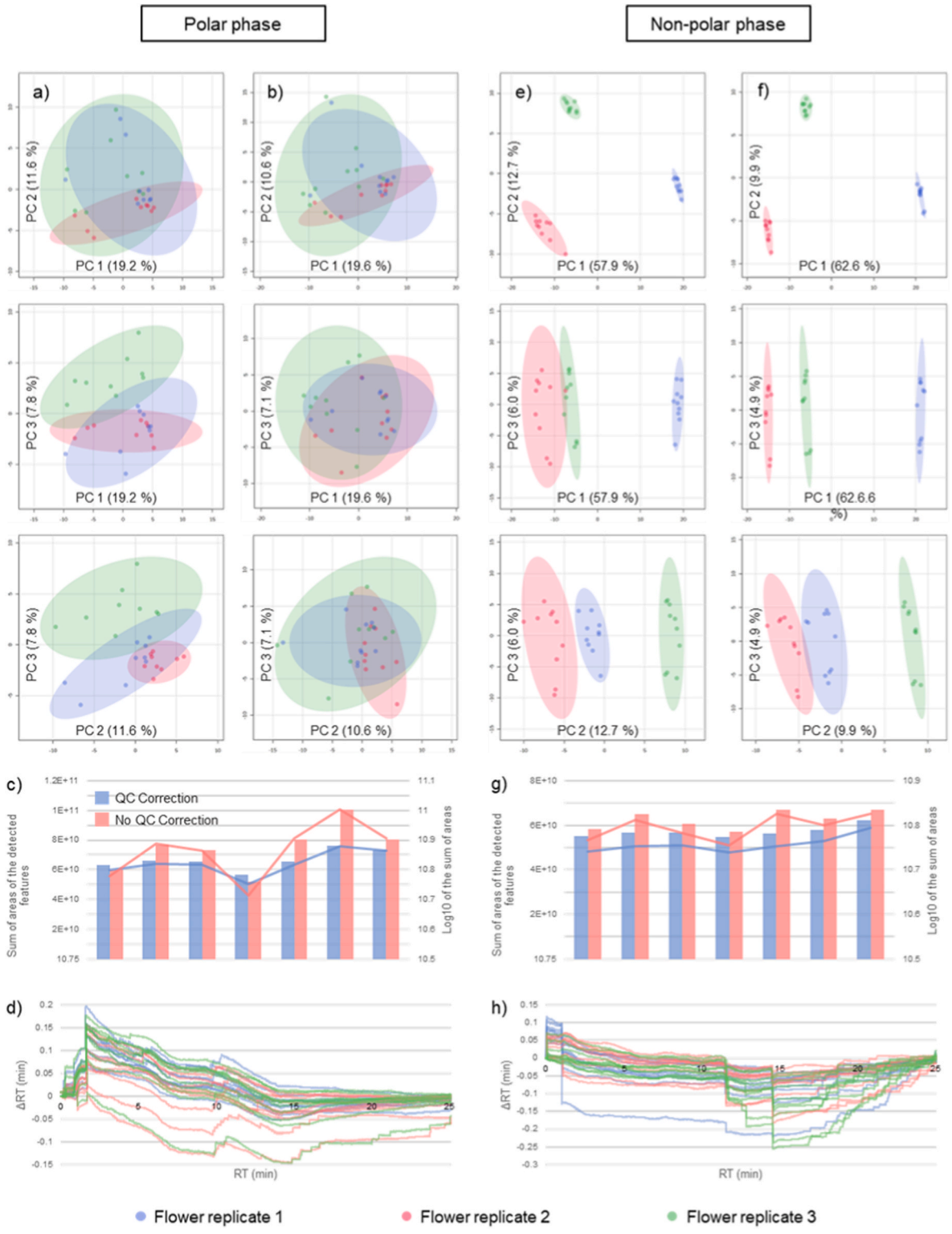


Fig. 6. Chromatographic drift monitorisation results in randomly analysed flower sample's biological and technical replicates: a) PCA scores of the corrected polar extract results (displayed at 95% confidence regions) b) PCA scores of the non-corrected polar extract results c) Fluctuation of the sum of areas of the features in the polar QC sample across the analysis sequence d) RT shift corrections in chromatographic features of the polar extracts through analysis time e) PCA scores of the corrected non-polar extract results f) PCA scores of the non-corrected non-polar extract results g) Fluctuation of the sum of areas of the features in the non-polar QC sample across the analysis sequence h) RT shift corrections in chromatographic features of the non-polar extracts through analysis time.

the non-corrected area signals of the QC sample followed a similar overall trend as the QC corrected signals through the analysis time span, but this could be a matter of chance in this exact case, as it did not happen the same in the case of the polar extract. Accordingly, the implementation of signal drift correction and RT shifts correction through QC samples for further research on untargeted cannabis metabolomics is strongly suggested, regardless of the specific research objectives, as the interpretation of non-corrected results could be biased

and unreliable.

3.3. Biological tissue selection

The analysis of leaf, stem and flower tissues from 10 plant clones of the same phenotype was carried out simulating a small-scale experiment for metabolomics study. Samples were extracted using the CMW solvent combination and they were subsequently analysed using the C18 chromatographic column. Periodic analysis of pooled QC samples was performed to enable later signal drift correction. A comparison between different sample classes (leaf, stem and flower) was performed based on unsupervised variance analysis of the filtered results, which can be found in Table S8 (Polar extract results) and Table S9 (Non-polar extract results) of SM. Therefore, a hierarchical cluster analysis (HCA) heatmap was generated with the detected significant features to determine the correlation between the samples and features. The data was auto scaled and transformed to logarithmic scale (Log_{10}), and Ward clustering method was applied using the Euclidean distance measure. The corresponding results, depicting the average peak area values of the detected features for both phases and ionisation modes of the 10 clones, are shown in Fig. 7, as comparison between the three studied tissues. A total of 166 and 57 significant features were identified in positive ionisation mode analysis of the polar and non-polar extracts, respectively. In the case of negative ionisation mode results, 23 and 24 significant features were identified in the polar and non-polar extracts, respectively. Thereby, to determine significant differences among biological tissues, ANOVA test was performed between tissue-wise sample averages.

Based on the information shown in Fig. 7 it was concluded that most of the identified compounds highly correlate with either flower or leaf tissues, while, in contrast, few of the detected features were directly correlated with the stem tissue, as it is summed in Table 2.

Among the detected significant features in the stem tissue, various compounds were identified as metabolites that play diverse roles in plant intern metabolism. For instance, p-coumaroyl tyramine, a derivative of p-coumaric acid, acts as a secondary metabolite involved in defence against stress and pathogens. It also participates in the synthesis of lignin, a characteristic compound that provides structural support to plant cell walls, which explains its abundance in the stem tissue [52]. Moreover, other highlighted metabolites observed in the stem happened to be lipids, such as linolenic acid or diacylglycerol (18:3(6Z,9Z, 12Z)/18:3(9Z,12Z,15Z)/0:0). Linolenic acid (Ω-3 fatty acid) plays a key role in plant growth, being part of cell membranes or participating in the synthesis of other growth regulator metabolites such as jasmonic acid. On the other hand, diacyl glycerol contributes to lipid metabolism, playing diverse roles as signal transduction, energy metabolism, and cell-integrity. It is also a component of the hydrophobic cell membranes [53,54].

Nevertheless, in leaf and flower tissues a greater number of key metabolites were identified, which play equal and more significant metabolic roles compared to those identified in the stem. For instance, p-coumaraldehyde, p-coumaric acid and p-coumaroyloctopamine were detected in leaf tissue. p-coumaric acid is created from p-coumaraldehyde, which is an essential part of the phenylpropanoid pathway, the same pathway p-coumaroyl tyramine (detected in the stem) is involved. Coumarins are also a class of phytoalexins, which are defence compounds produced by plants to deal with stress or pathogen attacks [55]. Additionally, metabolites with plant-hormone role such as abscisic acid or some lactones (trihomononactic acid lactone and xanthatin) were also identified. These compounds also significantly impact plant development regulation and response to environmental stimuli [56]. Furthermore, a wide range of secondary metabolites were also annotated, such as alkaloids (hordenine and 2-methylindole), quinones (lapachol) or fatty acids derivatives (ascr#13), metabolites that possess diverse roles as antioxidant, antimicrobial, signalling, plant defence, growth regulation and interaction functions with the environment.

In the case of flower tissue, other essential metabolites were also identified. As in the case of the leaf tissue, some growth regulation plant hormones and defence-related metabolites were observed, such as gibberellins (gibberellin A15) and phytotoxins (vomifoliol), respectively. Regarding secondary metabolites, flavonoids (flaviogeranin, quercetol C, robinetinidol) and their derivatives (kaempferol-7-alloside, quercetin-3-glucoside, nictoflorin) were identified, as well as some phenolic compounds (salicylaldehyde and caffealdehyde), which, similar to their counterparts in leaf tissue, are involved in a wide range of metabolic roles.

Furthermore, several amino acids were identified in both leaf and flower tissues, compounds that play essential roles in protein biosynthesis and signalling. Histidine, isoleucine or N6,N6,N6-trimethyl-lysine could be noted in the leaf tissue, whereas, alanine, choline, proline, aspartic acid or arginine could be observed in the flower, among others. Moreover, some dipeptides were also detected, which can act as signalling actors or storage forms of amino acids, for instance, valine-leucine or alanine-leucine were annotated in the leaf, and leucine-proline or glutamine-alanine in the flower. Essential aromatic amino acids were also detected: tryptophan in leaf and tyrosine (alongside corresponding derivatives tyrosilalanine and tyrosiltyrosine) and N-(Carboxyacetyl)phenylalanine (a phenylalanine derivative) in flower. Tryptophan serves as a precursor to previously stated metabolite classes as alkaloids, phytoalexins and indole glucosinolates [57]; while tyrosine acts as a precursor to isoquinoline alkaloids and quinones [58]. In parallel, flavonoids, tannins, lignins, and many other metabolites originate from phenylalanine [52], as this amino acid is believed to be the precursor of compounds than can constitute up to a third part of the organic compounds present in some plant species [59,60], therefore, it is crucial to identify them in the tissues under study. These three aromatic amino acids are downstream products of chlorismate, the final compound of the shikimate pathway [61], from which other metabolites, as some vitamins, are also derived. In line with this, it is worth mentioning that pantothenic acid, also known as vitamin B_5, was detected in the flower tissue, which is the acetyl CoA source, an indispensable component in various metabolic pathways [62].

Hence, based on the presence of specific metabolites in each biological tissue, it was determined that stem did not offer a significant contribution in terms of plant-representative metabolic information, whereas flowers and leaves seemed to be essential for the untargeted metabolomics study of the cannabis plant (all the stated metabolites were putatively identified, at identification confidence level 2).

4. Conclusion

We found that an optimisation of the analytical method was necessary for untargeted studies to achieve meaningful results. For untargeted metabolomic screening of cannabis, we identified chloroform:methanol: water (50:25:25, v/v) as the most suitable extractant, as it offered greater metabolomic coverage by enabling efficient extraction and phase separation. We also determined that LC-qOrbitrap analysis using the Kinetex 2.6 μm C18 100 Å (150 × 3 mm) column was more appropriate for the defined purpose, as it offered a higher compound number detection capability in both polar and non-polar extracts, with representative metabolomic coverage, and providing correct chromatographic peak qualities for accurate metabolite annotation and further data analysis. Nonetheless, it was found that, even though the C18 column yielded appropriate results, the occurrence of chromatographic drift needed the implementation of batch correction methods during data processing of untargeted metabolomics studies, which are characterised by the analysis of a large number of samples in one analysis-run. Moreover, it was determined that the stem tissue lacks significance in terms of representative metabolic information in comparison to leaf and flower tissues. The conclusions defined in this experimental procedure optimisation path can be contrasted with the methods followed in other research works related to cannabis metabolomics. For instance, in *Li* et al. [18], a methanol-water (75:25, v/v) solvent combination was used

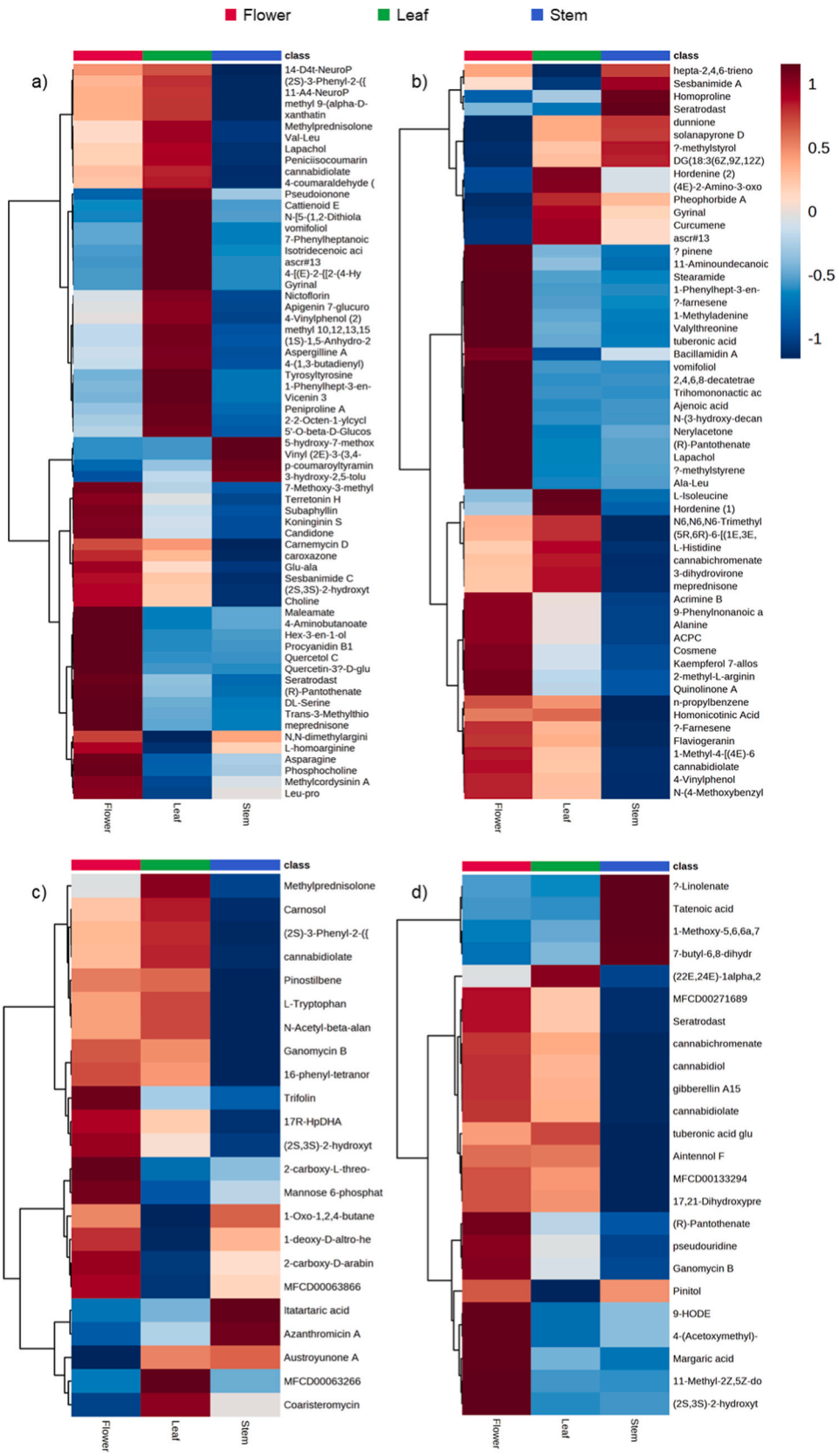


Fig. 7. HCA on average values of significant features in flower, leaf and stem samples a) polar extracts, positive ionisation (Top 65 features). Complete results of the features detected in the polar phase can be found in Fig. S1 of SM b) non-polar extracts, positive ionisation c) polar extracts, negative ionisation d) non-polar extracts, negative ionisation.

Table 2
Significantly correlated features in the studied biological tissues.

Ionisation mode	Extract Phase	Detected features			
		Flower	Leaf	Stem	Total
Positive	Polar	77	68	21	166
	Non-polar	34	15	8	57
Negative	Polar	10	9	4	23
	Non-polar	18	2	4	24

for the extraction of the metabolites, method that is very similar to study case Nº 1 of the extraction solvent optimisation step of this work, and as it could be empirically observed, solvent combination of study case Nº 4 ($CHCl_3$:H_2O:CH_3OH, 2:1:1, v/v) could offer broader metabolomic coverage. Furthermore, in *Zheng* et al. [13] it was stated that identification of more polyphenols in cannabis leaves was missed in their targeted assay, as this would turn to be very useful. Fortunately, polyphenols as robitenidol, kaempferol-7-alloside, quercetin-3-glucoside or nictoflorin, among others, were putatively identified in cannabis flower tissue.

Thus, based on the observations in this work, it is suggested that future research on understanding inner functioning of the cannabis plant is conducted according to experimental designs based on empirical evidence, with the purpose of covering the widest possible range of metabolic information. However, this would not mean that the experimental conditions discarded in this optimisation process (i.e., the rest of the tested solvents for metabolite extraction, the studied chromatographic columns and the stem tissue of the plant) are negligible for other works, but they should be optimised according to the specific objective of each research study instead. Therefore, unsupervised multivariate data analysis approaches were used in this work, in order to compare the actual data based on a certain objective, which was to optimise an experimental procedure for an untargeted metabolomics screening analytical method that offered the greatest metabolomic coverage, while maintaining data quality intact.

In this line, it is important to note that each study is unique and may have specific considerations that were not discussed in this work. For instance, factors such as dimensionality of the research plays an important role according to functionality, accessibility or economic matters in experimental design. The sample size to be processed and analysed can be a key factor. In certain studies, a large number of samples may need to be analysed, making it economically and temporally challenging to perform multiple extractions per sample. Thus, in such cases, a single-phase extraction procedure would be preferable. Similarly, constraints may prevent the analysis of extracts through two ionisation modes, leading to a loss of metabolomic coverage. Therefore, the success of a study resides in designing experiments that maximise resulting information within the available resources.

On the contrary, if hypothetical scenarios where a small sample quantity must be studied and the stated factors are feasible, experiments should be conducted using more than one chromatographic column, without discarding the excluded biological tissue, in order to gather the widest possible metabolic coverage. Ultimately, the design of appropriate experimental approaches should align with the specific research objectives, taking into account the available resources and leveraging empirical results to make informed decisions.

Funding

This work was financially supported by the Education Department of the Basque Country as a consolidated group of the Basque Research System (IT1213-19) and by Sovereign Fields S.L., in the framework of the project *Metabolomic study of Cannabis Sativa* L. *cultivations and determination of contaminants in medical cannabis plants*.

CRediT authorship contribution statement

M. San Nicolas: Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **A. Villate:** Investigation. **M. Olivares:** Conceptualization, Methodology, Writing – review & editing. **N. Etxebarria:** Conceptualization. **O. Zuloaga:** Writing – review & editing, All authors have read and agreed to the published version of the manuscript. **O. Aizpurua-Olaizola:** Conceptualization, Methodology, Investigation, Supervision, Project administration, Writing – review & editing. **A. Usobiaga:** Conceptualization, Methodology, Investigation, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

I have shared my data at the Attach File step

Acknowledgments

M. San Nicolas and A. Villate are grateful to the IBeA consolidated research group and to the UPV/EHU for their predoctoral fellowships.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2023.341848.

References

[1] D.C. Simiyu, J.H. Jang, O.R. Lee, Understanding cannabis sativa L.: current status of propagation, use, legalization, and haploid-inducer-mediated genetic engineering, Plants 11 (9) (May 2022) 1236, https://doi.org/10.3390/plants11091236.
[2] H. Carliner, Q.L. Brown, A.L. Sarvet, D.S. Hasin, Cannabis use, attitudes, and legal status in the U.S.: a review, Prev. Med. 104 (Nov. 2017) 13–23, https://doi.org/10.1016/j.ypmed.2017.07.008.
[3] O. Aizpurua-Olaizola, et al., Evolution of the cannabinoid and terpene content during the growth of cannabis sativa plants from different chemotypes, J. Nat. Prod. 79 (2) (2016) 324–331, https://doi.org/10.1021/acs.jnatprod.5b00949.
[4] W.A. Devane, F.A. Dysarz III, M.R. Johnson, L.S. Melvin, A.C. Howlett, Determination and characterization of a cannabinoid receptor in rat brain, Mol. Pharmacol. 34 (5) (1988) 605–613.
[5] P.F. Whiting, et al., Cannabinoids for medical use: a systematic review and meta-analysis, JAMA, J. Am. Med. Assoc. 313 (24) (2015) 2456–2473, https://doi.org/10.1001/jama.2015.6358.
[6] 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 (17–18) (2010) 2058–2073, https://doi.org/10.1016/j.phytochem.2010.10.001.
[7] I. Galve-Roperh, C. Sánchez, M.L. Cortés, T.G. Del Pulgar, M. Izquierdo, M. Guzmán, Anti-tumoral action of cannabinoids: involvement of sustained ceramide accumulation and extracellular signal-regulated kinase activation, Nat. Med. 6 (3) (2000) 313–319, https://doi.org/10.1038/73171.
[8] S. Blasco-Benito, et al., Appraising the "entourage effect": antitumor action of a pure cannabinoid versus a botanical drug preparation in preclinical models of breast cancer, Biochem. Pharmacol. 157 (2018) 285–293, https://doi.org/10.1016/j.bcp.2018.06.025.
[9] S. Zou, U. Kumar, Cannabinoid receptors and the endocannabinoid system: signaling and function in the central nervous system, Int. J. Mol. Sci. 19 (3) (2018), https://doi.org/10.3390/ijms19030833.
[10] M.A. ElSohly, M.M. Radwan, W. Gul, S. Chandra, A. Galal, Phytochemistry of cannabis sativa L, Prog. Chem. Org. Nat. Prod. 103 (2017) 1–36, https://doi.org/10.1007/978-3-319-45541-9_1.
[11] O. Hardiman, et al., Amyotrophic lateral sclerosis, Nat. Rev. Dis. Prim. 3 (2017), https://doi.org/10.1038/nrdp.2017.71.
[12] A. Cerrato, et al., Phytocannabinomics: untargeted metabolomics as a tool for cannabis chemovar differentiation, Talanta 230 (2021), https://doi.org/10.1016/j.talanta.2021.122313.
[13] J. Zheng, M. Johnson, R. Mandal, D.S. Wishart, A comprehensive targeted metabolomics assay for crop plant sample analysis, Metabolites 11 (5) (2021), https://doi.org/10.3390/metabo11050303.

[14] P.G. Vásquez-Ocmín, et al., Cannabinoids vs. whole metabolome: relevance of cannabinomics in analyzing Cannabis varieties, Anal. Chim. Acta 1184 (2021), https://doi.org/10.1016/j.aca.2021.339020.
[15] T.M. Sirangelo, R.A. Ludlow, N.D. Spadafora, Multi-omics approaches to study molecular mechanisms in cannabis sativa, Plants 11 (16) (2022), https://doi.org/10.3390/plants11162182.
[16] A. Shapira, P. Berman, K. Futoran, O. Guberman, D. Meiri, Tandem mass spectrometric quantification of 93 terpenoids in cannabis using static headspace injections, Anal. Chem. 91 (17) (2019) 11425–11432, https://doi.org/10.1021/acs.analchem.9b02844.
[17] 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 (13) (2022) 3847–3862, https://doi.org/10.1007/s00216-022-04026-2.
[18] C.-R. Li, et al., LC-MS-based untargeted metabolomics reveals chemical differences of Cannabis leaves from different regions of China, Ind. Crop. Prod. 176 (2022), https://doi.org/10.1016/j.indcrop.2021.114411.
[19] F. Thoma, A. Somborn-Schulz, D. Schlehuber, V. Keuter, G. Deerberg, Effects of light on secondary metabolites in selected leafy greens: a review, Front. Plant Sci. 11 (2020), https://doi.org/10.3389/fpls.2020.00497.
[20] 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.
[21] 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 (Mar. 2021), 128076, https://doi.org/10.1016/j.foodchem.2020.128076.
[22] Y. Zhou, et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (1) (Apr. 2019), https://doi.org/10.1038/s41467-019-09234-6. Art. no. 1.
[23] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (1) (2007) 51–78, https://doi.org/10.1002/mas.20108.
[24] A. Villate, et al., Review: metabolomics as a prediction tool for plants performance under environmental stress, Plant Sci. 303 (2021), https://doi.org/10.1016/j.plantsci.2020.110789.
[25] C.H. Johnson, J. Ivanisevic, G. Siuzdak, Metabolomics: beyond biomarkers and towards mechanisms, Nat. Rev. Mol. Cell Biol. 17 (7) (2016) 451–459, https://doi.org/10.1038/nrm.2016.25.
[26] E.T. Wurtzel, T.M. Kutchan, Plant metabolism, the diverse chemistry set of the future, Science 353 (6305) (2016) 1232–1236, https://doi.org/10.1126/science.aad2062.
[27] A. Raza, et al., Impact of climate change on crops adaptation and strategies to tackle its outcome: a review, Plants 8 (2) (2019), https://doi.org/10.3390/plants8020034.
[28] R. Kumar, A. Bohra, A.K. Pandey, M.K. Pandey, A. Kumar, Metabolomics for plant improvement: status and prospects, Front. Plant Sci. 8 (2017), https://doi.org/10.3389/fpls.2017.01302.
[29] T. Cajka, O. Fiehn, Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics, Anal. Chem. 88 (1) (Jan. 2016) 524–545, https://doi.org/10.1021/acs.analchem.5b04491.
[30] A. Ribbenstedt, H. Ziarrusta, J.P. Benskin, Development, characterization and comparisons of targeted and non-targeted metabolomics methods, PLoS One 13 (11) (Nov. 2018), e0207082, https://doi.org/10.1371/journal.pone.0207082.
[31] J.-L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, J. Chromatogr. A 1382 (2015) 136–164, https://doi.org/10.1016/j.chroma.2014.10.091.
[32] C. Martias, et al., Optimization of sample preparation for metabolomics exploration of urine, feces, blood and saliva in humans using combined NMR and UHPLC-HRMS platforms, Molecules 26 (14) (Jan. 2021), 14, https://doi.org/10.3390/molecules26144111.
[33] M.-E.P. Papadimitropoulos, C.G. Vasilopoulou, C. Maga-Nteve, M.I. Klapa, Untargeted GC-MS metabolomics, Methods Mol. Biol. 1738 (2018) 133–147, https://doi.org/10.1007/978-1-4939-7643-0_9.
[34] G. Tooulakou, et al., Alarm photosynthesis: calcium oxalate crystals as an internal CO2 source in plants, Plant Physiol. 171 (4) (Aug. 2016) 2577–2585, https://doi.org/10.1104/pp.16.00111.
[35] R.C.H. De Vos, S. Moco, A. Lommen, J.J.B. Keurentjes, R.J. Bino, R.D. Hall, Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry, Nat. Protoc. 2 (4) (2007) 778–791, https://doi.org/10.1038/nprot.2007.95.
[36] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based metabolomic analysis of plants, Nat. Protoc. 5 (3) (2010) 536–549, https://doi.org/10.1038/nprot.2009.237.
[37] T. Fujii, et al., Direct metabolomics for plant cells by live single-cell mass spectrometry, Nat. Protoc. 10 (9) (2015) 1445–1456, https://doi.org/10.1038/nprot.2015.084.
[38] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A.R. Fernie, Gas chromatography mass spectrometry–based metabolite profiling in plants, Nat. Protoc. 1 (1) (Jun. 2006) 387–396, https://doi.org/10.1038/nprot.2006.59.
[39] J.M. Amigo, Data mining, machine learning, deep learning, chemometrics: definitions, common points and trends (Spoiler Alert: VALIDATE your models!), Brazilian J. Anal. Chem. 8 (32) (2021) 22–38, https://doi.org/10.30744/BRJAC.2179-3425.AR-38-2021.
[40] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507, https://doi.org/10.1126/science.1127647.
[41] E. Small, H.D. Beckstead, Cannabinoid phenotypes in cannabis sativa, Nature 245 (5421) (Sep. 1973), https://doi.org/10.1038/245147a0. Art. no. 5421.
[42] A. Rivas-Ubach, et al., 'Topsoil depth substantially influences the responses to drought of the foliar metabolomes of Mediterranean forests', Perspectives in Plant Ecology, Evol. System. 21 (2016) 41–54, https://doi.org/10.1016/j.ppees.2016.06.001.
[43] H.K. Kim, R. Verpoorte, Sample preparation for plant metabolomics, Phytochem. Anal. 21 (1) (2010) 4–13, https://doi.org/10.1002/pca.1188.
[44] S. Wu, et al., Mapping the arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions, Mol. Plant 11 (1) (2018) 118–134, https://doi.org/10.1016/j.molp.2017.08.012.
[45] K. Chetnik, L. Petrick, G. Pandey, MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data, Metabolomics 16 (11) (2020), https://doi.org/10.1007/s11306-020-01738-3.
[46] A.C. Schrimpe-Rutledge, S.G. Codreanu, S.D. Sherrod, J.A. McLean, Untargeted metabolomics strategies—challenges and emerging directions, J. Am. Soc. Mass Spectrom. 27 (12) (Dec. 2016) 1897–1905, https://doi.org/10.1007/s13361-016-1469-y.
[47] Z. Pang, et al., MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic Acids Res. 49 (W1) (Jul. 2021) W388–W396, https://doi.org/10.1093/nar/gkab382.
[48] Z. Pang, et al., Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data, Nat. Protoc. 17 (8) (Aug. 2022), 8, https://doi.org/10.1038/s41596-022-00710-w.
[49] Y. Lu, Z. Pang, J. Xia, Comprehensive investigation of pathway enrichment methods for functional interpretation of LC–MS global metabolomics data, Briefings Bioinf. 24 (1) (Jan. 2023) bbac553, https://doi.org/10.1093/bib/bbac553.
[50] R.G. Sadygov, F.M. Maroto, A.F.R. Hühmer, ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces, Anal. Chem. 78 (24) (2006) 8207–8217, https://doi.org/10.1021/ac060923y.
[51] S. Fan, et al., Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data, Anal. Chem. 91 (5) (2019) 3590–3596, https://doi.org/10.1021/acs.analchem.8b05592.
[52] T. Vogt, 'Phenylpropanoid biosynthesis', Mol. Plant 3 (1) (2010) 2–20, https://doi.org/10.1093/mp/ssp106.
[53] Y. Yang, C. Benning, Functions of triacylglycerols during plant development and stress, Curr. Opin. Biotechnol. 49 (Feb. 2018) 191–198, https://doi.org/10.1016/j.copbio.2017.09.003.
[54] H.U. Kim, 'Lipid metabolism in plants', Plants 9 (7) (Jul. 2020) 7, https://doi.org/10.3390/plants9070871.
[55] R.A. Dixon, Natural products and plant disease resistance, Nature 411 (6839) (2001) 843–847, https://doi.org/10.1038/35081178.
[56] M. Chadwick, H. Trewin, F. Gawthrop, C. Wagstaff, Sesquiterpenoids lactones: benefits to plants and people, Int. J. Mol. Sci. 14 (6) (2013) 12780–12805, https://doi.org/10.3390/ijms140612780.
[57] E.R. Radwanski, R.L. Last, Tryptophan biosynthesis and metabolism: biochemical and molecular genetics, Plant Cell 7 (7) (1995) 921–934.
[58] T.M. Kutchan, Alkaloid biosynthesis - the basis for metabolic engineering of medicinal plants, Plant Cell 7 (7) (1995) 1059–1070.
[59] R.A. Razal, S. Ellis, S. Singh, N.G. Lewis, G.H.N. Towers, Nitrogen recycling in phenylpropanoid metabolism, Phytochemistry 41 (1) (1996) 31–35, https://doi.org/10.1016/0031-9422(95)00628-1.
[60] T. Tohge, M. Watanabe, R. Hoefgen, A.R. Fernie, Shikimate and phenylalanine biosynthesis in the green lineage, Front. Plant Sci. 4 (MAR) (2013), https://doi.org/10.3389/fpls.2013.00062.
[61] H. Maeda, N. Dudareva, The shikimate pathway and aromatic amino acid biosynthesis in plants, Annu. Rev. Plant Biol. 63 (2012) 73–105, https://doi.org/10.1146/annurev-arplant-042811-105439.
[62] R. Leonardi, Y.-M. Zhang, C.O. Rock, S. Jackowski, 'Coenzyme A: back in action', Prog. Lipid Res. 44 (2–3) (2005) 125–153, https://doi.org/10.1016/j.plipres.2005.04.001.