Journal of Chromatography B 1073 (2018) 34–42

Contents lists available at ScienceDirect

Journal of Chromatography B

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

Influence of direct and sequential extraction methodology on metabolic profiling

K.M. Maria John, James Harnly, Devanand Luthria*

Food Composition Methods Development Lab., Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, 20705, United States

ARTICLE INFO

Keywords:
Sequential and direct extractions
Metabolic profiling
With and without derivatization
Multivariate analysis
GC–MS
HPLC-UV
HPLC–MS

ABSTRACT

A systematic comparison was made of the detected metabolite profiles for two plant materials (black beans and soybeans) and a dietary supplement (black cohosh) extracted using sequential (hexane, ethyl acetate, and 50% aqueous methanol) and direct extraction with three solvent systems (80% aqueous methanol, methanol/chloroform/water (2.5:1:1, v/v/v) and water). Extracts were analyzed by LC–MS (without derivatization) and GC-FID (with BSTFA/TMCS derivatizations). For sequential extraction, HPLC-UV and BSTFA/TMCS-derivatized GC-FID detection were more responsive to the polar molecules with a rough distribution of 10%, 10%, and 80% of the total signals in hexane, ethyl acetate, and 50% aqueous methanol, respectively. With HPLC–MS detection, the distribution of signals was more balanced, roughly 40%, 30%, and 30% for the same extracts (hexane, ethyl acetate, and 50% aqueous methanol). For direct extraction, HPLC-UV and BSTFA/TMCS-derivatized 4GC-FID provided signals between 60% and 150% of the total sequential extracted signals. The overlap of signals for the 3 sequential extracts ranged from 1% to 3%. The overlap of the signals for direct extraction with the total for sequential extraction ranged from 15% to 98%. With HPLC–MS detection, signals varied from 30% to 40% of the total signals for sequential extraction. Multivariate analysis showed that the components for the sequential and direct extracts were statistically different. However, each extract, sequential or direct, allowed discrimination between the 3 plant materials.

1. Introduction

In the past two decades, there has been significant research progress in the field of metabolite profiling and fingerprinting due to significant advancements in analytical technology and computational power. Metabolomics and metabolite profiling are comprehensive characterization of a tissue that focuses on identification and quantification of the low-molecular weight metabolites [1–5]. Conversely, metabolite fingerprinting uses the pattern of metabolites for comparison and discrimination analysis with no initial concern regarding identification or quantification [6]. Both approaches have found extensive application in the biological sciences.

A metabolite profile or fingerprint is a function of the solvent polarity and the detection system. Since metabolomics are generally intended to be as comprehensive as possible, the initial extraction of compounds from a solid sample is important. Plant materials are particularly challenging because they contain thousands of compounds with a wide range of chemistries, large dynamic ranges, and complex matrices [7]. Thus, regardless whether the goal is metabolomics or metabolite fingerprinting, extraction is a compromise. Classically, analytical methods were optimized for a single compound in a single matrix. This is not possible with metabolomics studies, especially for non-targeted methods. Since there is no universal solvent, the extraction method will display greatest efficiency for compounds with a polarity similar to the selected solvent. In this sense, every extraction method is targeted.

A truly comprehensive metabolomics method should employ multiple solvents. This problem has been discussed extensively in the literature [8–10]. Sequential and direct extraction methods have been employed; the method usually dependent on the purpose of the analysis. The most common extraction methods target either the polar and moderately polar secondary metabolites using different proportions of an aqueous methanol solvent mixtures. However for non-polar metabolites (lipidomics) hexane is either used separately or in combination with other non-polar solvents. Various ratios of methanol-chloroform-water have been used to accommodate a wider range of polar and non-polar compounds [11]. However, there have been no systematic studies comparing the application of sequential and direct extraction methods.

* Corresponding author at: Plant Composition Methods Development Lab, Bldg. 161, Lab 202, BARC (E), 10300, Baltimore Avenue, United States./Agricultural Research Service, US Department of Agriculture, Beltsville, MD, 20705, United States.
E-mail address: Dave.Luthria@ars.usda.gov (D. Luthria).

https://doi.org/10.1016/j.jchromb.2017.12.005
Received 5 September 2017; Received in revised form 1 December 2017; Accepted 3 December 2017
Available online 05 December 2017
1570-0232/ Published by Elsevier B.V.

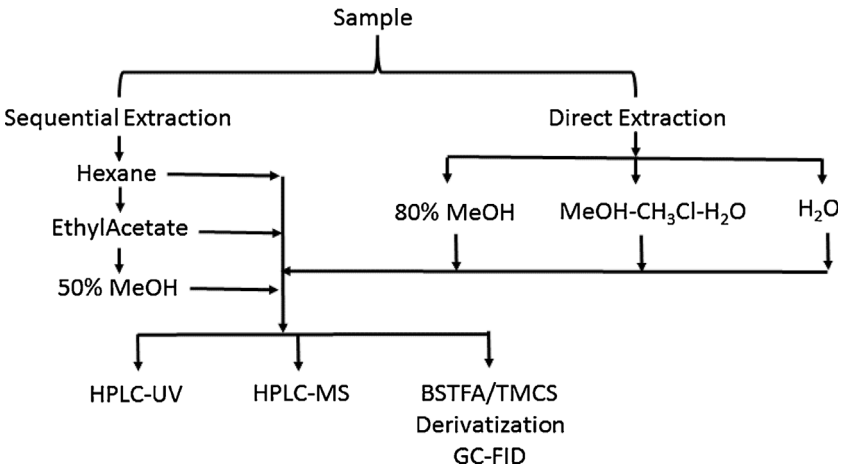

Fig. 1. Schematic diagram of extraction (sequential & direct) and analysis (GC, GC–MS, LC–MS) of metabolic profiles from three different plant samples.

The detection system is also an important determining factor for the metabolomic profiles or fingerprint. Metabolomic profiles require a separation/detection system, ideally UHPLC combined with HRAM/MS to obtain maximum detection and information [12,13]. Fingerprinting methods may skip the separation step and use flow injection MS or NMR. Fingerprints have also been acquired using IR, NIR, and UV spectrometry, but provide less information. Each instrument, however, has different sensitivities, measuring different physicals properties, e.g., mass, magnetic resonance, or vibrational transitions. Reponses of different molecules may vary significantly making detection and/or quantification problematic.

In this manuscript, a systematic comparison of metabolites extracted from two plants (black beans (BB) and soybeans (SB)) and a dietary supplement (black cohosh (BC)), was made using different extraction solvent strategies and analytical systems (Fig. 1). The three plant materials were selected as model substrates because they have been extensively studied and are known to contain wide array of metabolites with varying polarities [14–16]. Sequential extraction was accomplished using hexane (Hex), ethyl acetate (EtAc), and 50% aqueous methanol (50Me). These results were compared to three direct extraction methods using 80% aqueous methanol (80Me), methanol-chloroform-water (MCH), and 100% H_2O (H_2O). The sequential and direct extracts were analyzed by high performance liquid chromatography with photodiode array and mass spectrometric detection (HPLC-UV and HPLC–MS), and derivatized samples were analyzed by GC-FID. The chromatographic profiles of each of the solvents and the percentage of the analytical signal in each were compared. Principal component analysis (PCA) of the chromatograms was used to judge the similarity of the extracts and their ability to discriminate between the plant materials.

2. Materials and methods

2.1. Sample preparation

Freeze-dried ground material of black cohosh (BC) from root and rhizome (Wild Harvest Grade C Powder Run #947) was obtained from Dr. David Lytle of the Eclectic Institute, Oregon USA. Soybean (SB) seeds were purchased from a local grocery store (Mom in Beltsville, Maryland). Black bean (BB) samples were provided by Dr. M.A. Pastor-Corrales from USDA (Beltsville, Maryland). All samples were ground in a home style coffee grinder. All ground samples were passed through a standard 20-mesh sieve (particle size < 0.825 mm) and stored in a freezer < -20 °C until analyzed.

2.2. Sequential extraction

Each sample (250 mg) was weighed and sequentially extracted with three solvents in the following order: (i) hexane, (ii) ethyl acetate, and (iii) methanol:water (1:1 v:v). After each solvent extraction, the residue was placed under a stream of nitrogen gas to remove solvent traces before shifting to the next solvent system. The general procedure for each solvent: 5 mL of solvent was added to the ground sample and the mixture was vortexed followed by Ultra-Sonication for a period of 10 min. The extracts were centrifuged at 4000 rpm, supernatant was transferred to a separate vial. The residue was re-extracted with additional 5 mL of the same solvent. Collected supernatants from the two extractions were mixed and aliquoted into two separate vials. The solvent from two vials were evaporated to dryness using SpeedVac. The residue from one vial was re-dissolved in 1 mL of methanol: water (8:2, v/v) vortexed and filtered and the extract was used for LC–MS analysis. The residue in the second vial was derivatized and used for GC analysis.

For GC metabolic profiling the dried pellet was derivatized with methoxyamine hydrochloride in pyridine (20 mg/mL) of 200 µL. The mixture was incubated at 30 °C for 90 min. To this mixture, 100 µL of N,O-bis (trimethylsilyl) trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (TMCS) was added. The mixture was incubated for an additional 30 min at 37 °C. The mixture was assayed by GC-FID for metabolic profiling and GC–MS was used for the identification of metabolites [14]. Amino acids, organic acids and sugars were analyzed using an Agilent GC–MS system (7890A) attached to an Agilent (7693) autosampler. For metabolic profiling, derivatized sample (1 µL) was injected on a HP-5MS capillary column (30 m length × 0.25 mm i.d. × 0.25 µM film thickness-Agilent J&W GC column). The injector temperature was set at 250 °C, and the oven temperature was programmed as follows: 80 °C for 2 min, then ramped to 300 °C at a rate of 10 °C/min, and held at 300 °C for 3 min. The transfer line temperature was set at 250 °C. The ionization potential was set at 70 V (electron energy) with a source temperature of 200 °C. The detector voltage was 1450 V and the mass range was set at 50–600 m/z with an acquisition rate of 10 spectra per second.

2.3. Direct extraction

All three samples were directly extracted using three extraction solvents: methanol:water (8:2, v/v) and water (100%) frequently described in literature [17]. For the LC–MS analysis 250 mg of ground samples (BC, SB, BB) were extracted with 5 mL methanol: water (8:2, v/v) using ultrasonic assisted extraction. The extracts were centrifuged at 4000 rpm for 10 min and the supernatant was collected. This extraction was repeated one more time with 5 mL of fresh solvent. The pooled supernatants from two extractions were evaporated to dryness using a SpeedVac. Collected supernatants from the two extractions were mixed and aliquoted into two separate vials. The solvent from two vials were evaporated to dryness using SpeedVac. The residue from one vial was re-dissolved in 1 mL of methanol: water (8:2, v/v) vortexed and filtered

and the extract was used for LC–MS analysis. The residue in the second vial was derivatized and used for GC–MS analysis.

2.4. HPLC and GC analysis

UHPLC analysis was carried out using Agilent 1290 attached to a diode array and a G1956B Mass Spectrometer (Agilent). A Luna 5 μ C18 (2) 100A column (150 × 4.6 mm) was used with the gradient system of (A) water and (B) acetonitrile with 0.1% formic acid (v/v). The initial gradient of the mobile phase was maintained at 5% acetonitrile for 1 min, gradually increased to 100% acetonitrile at 60 min, maintained at 100% acetonitrile for 1 min, and then decreased to 5% acetonitrile and maintained for 1 min. Samples (20 μL) were injected with the flow rate of 0.7 mL/min. ESI was performed in the negative (−) and positive (+) ion mode over a range of 100–1000 *m/z*. The operating parameters were as follows: ion source temperature, 300 °C; capillary voltage, 70 V. Six replicate extractions and analyses were carried out for each sample.

For GC–MS metabolic profiling the dried pallet was derivatized with methoxyamine hydrochloride in pyridine (20 mg/mL) of 200 μL. The mixture was incubated at 30 °C for 90 min. To this mixture, 100 μL of N,O-bis (trimethylsilyl) trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (TMCS) was added. The mixture was incubated for an additional 30 min at 37 °C. The mixture was assayed by GC-FID and GC–MS analysis for the identification of metabolites [14]. Amino acids, organic acids and sugars were analyzed using an Agilent GC–MS system (7890A) attached to an Agilent (7693) autosampler. For metabolic profiling, derivatized sample (1 μL) was injected on a HP-5MS capillary column (30 m length × 0.25 mm i.d. × 0.25 μM film thickness-Agilent J&W GC column). The injector temperature was set at 250 °C, and the oven temperature was programmed as follows: 80 °C for 2 min, then ramped to 300 °C at a rate of 10 °C/min, and held at 300 °C for 3 min. The transfer line temperature was set at 250 °C. The ionization potential was set at 70 V (electron energy) with a source temperature of 200 °C. The detector voltage was 1450 V and the mass range was set at 50–600 *m/z* with an acquisition rate of 10 spectra per second.

2.5. Data processing and multivariate analysis

The GC–MS (*.qgd) and LC–MS (*.xms) data files were converted to CSV (*.csv) files followed by convert into Microsoft Excel. The resulting aligned file was processed for multivariate statistical analysis using principal component analysis using Solo software (eigenvector, Wenatchee, WA) and SIMCA-13 software (Umetrics, Umea, Sweden) [18]. The discriminate metabolites from the entire chromatogram were selected based on variable importance in the projection (VIP) values (value > 8.0) and p-value (< 0.05) statistics.

3. Results and discussion

3.1. Sequential extraction

Three plant materials, black cohosh (BC), black beans (BB), and soybeans (SB) were extracted sequentially; one extraction after the other from the same sub-sample. The solvents, in order, were Hex (Hex, polarity index 0.0), ethyl acetate (EtAc, polarity index 4.4), and 50% aqueous methanol (50Me, polarity index 7.1). Since there is no universal detector, 3 detection systems were used (LC-UV, LC–MS, and GC-FID after derivatization) to characterize the extracts. Thus, UV was sensitive to molecules with chromophores (conjugated bond systems), MS was sensitive to easily ionized molecules, and FID responded to derivatized polar molecules (without chromophores).

Fig. 2 shows the chromatographic profiles of six extracts obtained for BC (Fig. 2A), BB (Fig. 2B), and SB (Fig. 2C) using UHPLC with diode-array detection. These traces provide a qualitative comparison of the molecules extracted with six different extraction solvents. The traces for all 3 materials are similar, showing considerable overlap for Hex and EtAc at longer retention times (more non-polar molecules) and a shift to more moderate retention times (more polar molecules) for 50Me, as expected. It should be noted that the plots for Hex and EtAc have been expanded by a factor of 5.

Significant overlap of peaks in the chromatograms for Hex and EtAc extracts were observed for all three materials with diode-array detection. In theory, repeat extractions with Hex would have efficiently removed the non-polar compounds. A possible explanation is that most of the compounds were not efficiently extracted by Hex due to their moderate polarity. Among the three materials investigated, soybean contains approximately 20% oil as compared to trace quantity in black beans and insignificant amount in black cohosh. These differences were not captured in UHPLC chromatographic traces with diode array detection due to lack of chromophores in the triglyceride molecules (oil).

A more quantitative evaluation of the sequential extraction of the 3 plant materials with measurements by the 3 detection methods (HPLC-UV, HPLC–MS and GC-FID) is presented in Table 1 which consists of 9 sub-tables (3 plant materials × 3 detection methods). For each chromatogram, the signals (absorbance, total ion counts, or counts) were summed for the full interval (0–70 min). In each table, the sum of the signals for the 3 sequential extracts was used as the basis for evaluating the fraction of molecules extracted with each solvent. Thus, the sum of the values for Hex, EtAc, and 50Me in the top-left of each table is 100%. For example, Table 1A shows that for BC, 10%, 11%, and 79% of the total signal were attributed to Hex, EtAc, and 50Me, respectively. This is consistent with the traces in Fig. 2 where the scales for Hex and EtAc have been expanded by a factor of 5.

Table 1 also shows data for the fraction of overlapping signals (bottom-left), again expressed as a percentage of the sum of the signals for the 3 sequential extracts. Thus, in Table 1A, 4% of the signals for Hex and EtAc were observed for molecules at the same retention time. Since Hex and EtAc constituted 10% and 11% of the total signal, respectively, approximately half their signals occurred over the same interval. This suggests that Hex did not efficiently remove all the molecules or that EtAc was more efficient in interacting with the sample matrix to remove the remaining molecules.

The fraction of the summed signal for sequential extraction was approximately the same for all 3 materials. Hex and EtAc provided only 10% to 30% of the total signal. The overlap of signals for the 3 solvents was also similar, ranging from 1% to 7%.

3.2. Direct extraction

Sequential extraction was compared to direct extraction using three solvents frequently reported in the literature: 80% aqueous methanol (80Me, polarity index 5.9), methanol-chloroform-water (MCH, 2.5:1:1,v/v/v, polarity index 6.0), and water (H_2O, polarity index 10.2) [14,19]. In each case, a separate sub-sample was used for each solvent. The chromatograms for the 3 direct extracts are shown in Fig. 2A–C. The chromatograms show considerable overlap with those for 50Me and also, as expected, more peaks at shorter retention times (more polar molecules).

Quantitatively, the fraction of the signals obtained for each of the direct extracts, based on the total signal for the sequential extracts, is shown in Table 1A–I (top-right). Thus, in Table 1A, 80Me, MCH, and H_2O provided 87%, 116%, and 71% of the total sequential signal, respectively. The data suggest that either MCH extracts the same molecules as the sequential series in a more efficient manner or extracts molecules beyond the polarity range of the sequential solvents. The data in Table 1A–C and Fig. 2 suggest that the bulk of the molecules extracted are polar (t_R = 0–30 min) or intermediately polar (t_R = 30–50 min). Alternatively, a much larger fraction of the polar and intermediately polar molecules possess chromophores.

Table 1A–C also present the overlap of the signals for the direct extracts with the sequential extracts (bottom-right), i.e., overlap of the signals from the direct extracts with the 3 signals from the sequential

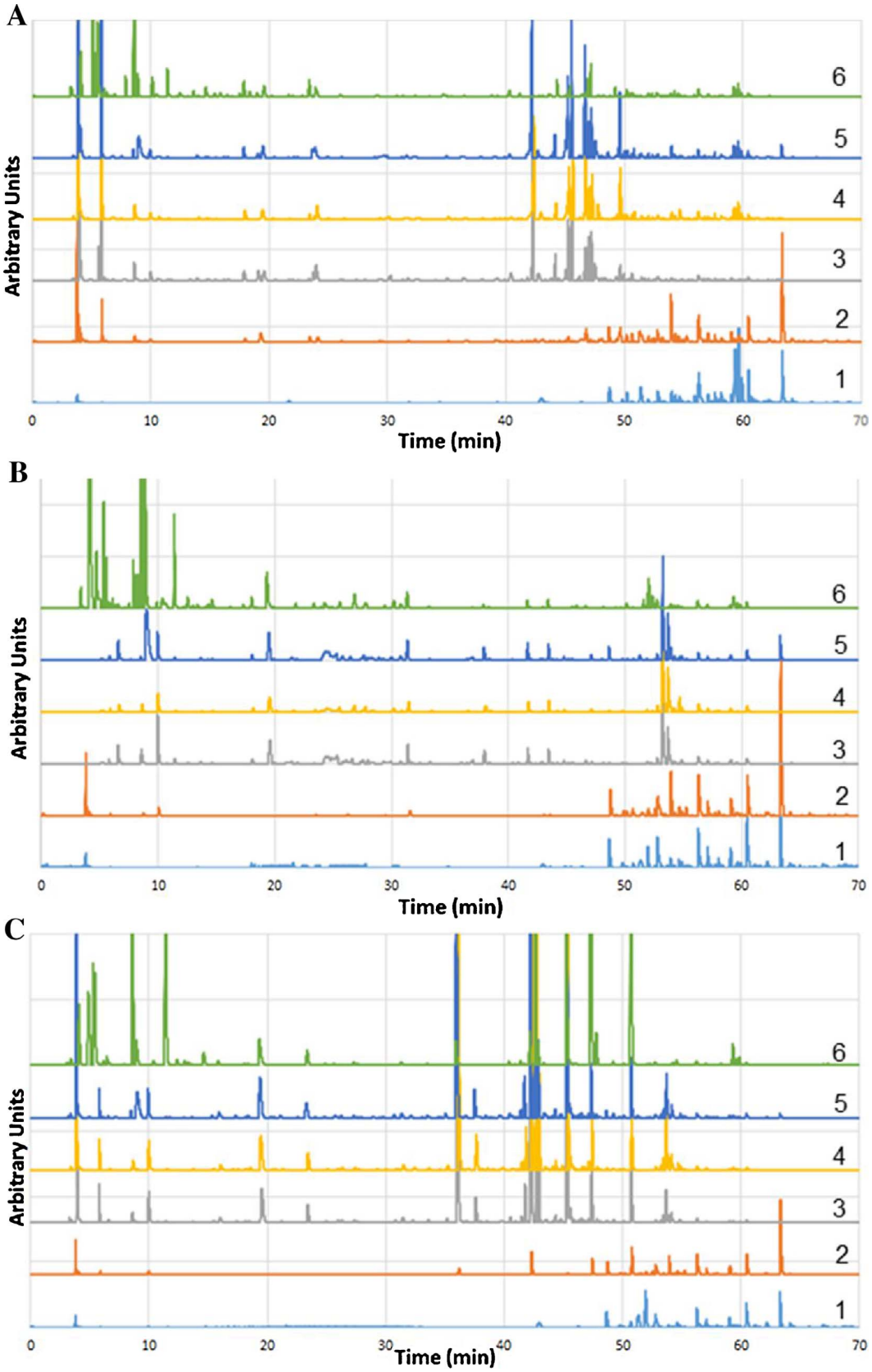

Fig. 2. LC–MS chromatograms of six extracts (1. hexane, 2.ethyl acetate, 3. 50% methanol: water, 4. 80% methanol: water, 5. methanol: chloroform: water, 6. water) from three different plant samples (A. BC, B. BB and C. SB).

Table 1
Fractions of compounds extracted and overlapping signals for sequential and direct methods.

	Liquid Chromatography (UV)	Liquid Chromatography (MS)	Gas Chromatography (Derivatized) (FID)
Black Cohosh	Table 1A	Table 1D	Table 1G
Black Bean	Table1B	Table 1E	Table 1H
Soy Bean	Table 1C	Table 1F	Table 1I

Table 1A

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	10%	11%	79%	87%	116%	71%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	4%	---	80%Me	41%	---	
50%Me	3%	5%	MCH	58%	41%	---
			H2O	21%	13%	19%

Table 1D

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	51%	15%	34%	38%	---	39%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	16%	---	80%Me	12%	---	
50%Me	26%	6%	MCH	---	---	---
			H2O	22%	6%	---

Table 1G

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	6%	6%	88%	107%	96%	128%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	4%	---	80%Me	71%	---	
50%Me	2%	3%	MCH	53%	82%	---
			H2O	53%	80%	81%

Table1B

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	14%	15%	71%	53%	95%	156%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	7%	---	80%Me	41%	---	
50%Me	7%	5%	MCH	61%	41%	---
			H2O	23%	16%	28%

Table 1E

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	34%	36%	30%	34%	--	37%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	16%	---	80%Me	18%	---	
50%Me	7%	15%	MCH	---	---	---
			H2O	18%	7%	---

Table 1H

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	10%	3%	87%	98%	57%	79%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	2%	---	80%Me	98%	---	
50%Me	3%	1%	MCH	56%	54%	---
			H2O	74%	69%	52%

Table 1C

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	3%	3%	94%	108%	110%	95%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	1%	---	80%Me	64%	---	
50%Me	1%	1%	MCH	57%	36%	---
			H2O	20%	14%	26%

Table 1F

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	44%	32%	24%	31%	--	31%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	15%	---	80%Me	15%	---	
50%Me	7%	14%	MCH	---	---	---
			H2O	19%	6%	---

Table 1I

Fraction Extracted						
Peak	Sequential Extraction			Direct Extraction		
Area	Hex	EtAc	50%Me	80%Me	MCH	H2O
100%	8%	3%	89%	86%	97%	95%

Overlapping Signals						
Beween Seq Extractions			Seq Extraction vs Dir Extraction			
	Hex	EtAc		Seq	80%Me	MCH
Hex	---		Seq	---		
EtAc	2%	---	80%Me	86%	---	
50%Me	3%	2%	MCH	97%	68%	---
			H2O	82%	55%	77%

extracts treated as a single unit. Thus, in Table 1A, the series of sequential extracts corresponds most closely with MCH (58%). Correlation is least for H_2O with any of the other extracts, direct or sequential. The H_2O extract has 71% of the signals found for sequential extraction but only a 21% overlap. This suggests that signals equivalent to half that for the sequential extraction are coming from molecules not extracted by sequential extraction.

The data suggest that, with the use of a reverse phase column, two extractions with either Hex and EtAc and 50Me, 80Me, or MCH are sufficient to characterize the non-polar and moderately polar components of the samples. The polar fraction would be better characterized using a normal phase column. This would provide separation of compounds found in the solvent front as demonstrated by Buszewski and Noga, with a HILIC column [20]. In the past, 50Me and a reverse phase column were effective in identifying amino acids, organic acids, and simple sugars [11,21–23]. These molecules proved more effective in discriminating between cultivars and farming mode for broccoli than the polyphenols. However, the polar compounds may not have been extracted efficiently by the 50Me and chances to identify many of these compounds may have been lost when they eluted in the solvent front.

3.3. Comparing detection systems

The results for GC-FID after derivatization of sugars, organic acids and amino acids (Table 1G–I) were similar to those for HPLC-UV (Table 1A–C). This is not surprising as the BSTFA/TMCS derivatization targets polar molecules with labile hydrogens (e.g., alcohols, phenols, carboxylic acids, amines, amides, thiols, steroids, and alkaloids), a subset of the polar and moderately polar molecules. Consequently, high signals (percentages) for 50Me. 80Me, MCH, and H_2O are not unexpected. Only 3 of 24 values were less than 79%.

By contrast, HPLC–MS presents a dramatically different picture. For all 3 materials, the average extracted fraction for Hex, EtAc, and 50Me were 45%, 28%, and 29% (based on total ion counts), respectively. This compares to averages of 9%, 10%, and 81% for HPLC-UV and 7%, 4%, and 88%, for GC-FID (after derivatization). For direct extraction, MS provides averages for all 3 materials of 34% and 36% for 80Me and H_2O, respectively compared to 83% and 107% for HPLC-UV and 97% and 93% for GC-FID. Obviously, UV absorbance and GC-FID (after derivatization) favors the detection of polar compounds with chromophores whereas MS appears to provide more balanced detection for all molecules [24]. Non-polar molecules such as oils and triterpenoids which are abundant in soybeans and black cohosh are not detected due to lack of chromophore and targeted derivations for amino acids, sugars and organic acids but are easily detected with MS [25].

3.4. Profile similarity

The principal component analysis score plots in Fig. 3 show the similarity of the chromatograms for each of the solvents and detection systems. Fig. 3A shows the PCA score plot for sequential extraction (Hex, EtAc, and 50Me) with HPLC-UV detection. The relatively non-polar solvents (Hex and EtAc) cluster close to one another on the right side of the plot and the chromatograms for 50Me extraction are on the left side. Each data point represents the average of 4 chromatograms. Fig. 3B presents the same sequential extraction data plus the results for the 3 direct extractions (80Me, MCH, and H_2O). The relatively non-polar extracts are in the lower-right quadrant and more polar extracts are on the left side. Clusters for the sequential extracts are determined by the solvent while the clusters for the direct extracts are determined by the plant material.

Fig. 3C presents a PCA score plot similar to Fig. 3B for sequential and direct extraction with GC–MS detection (after derivatization for target molecules amino acids, organic acids and sugars). The extracts

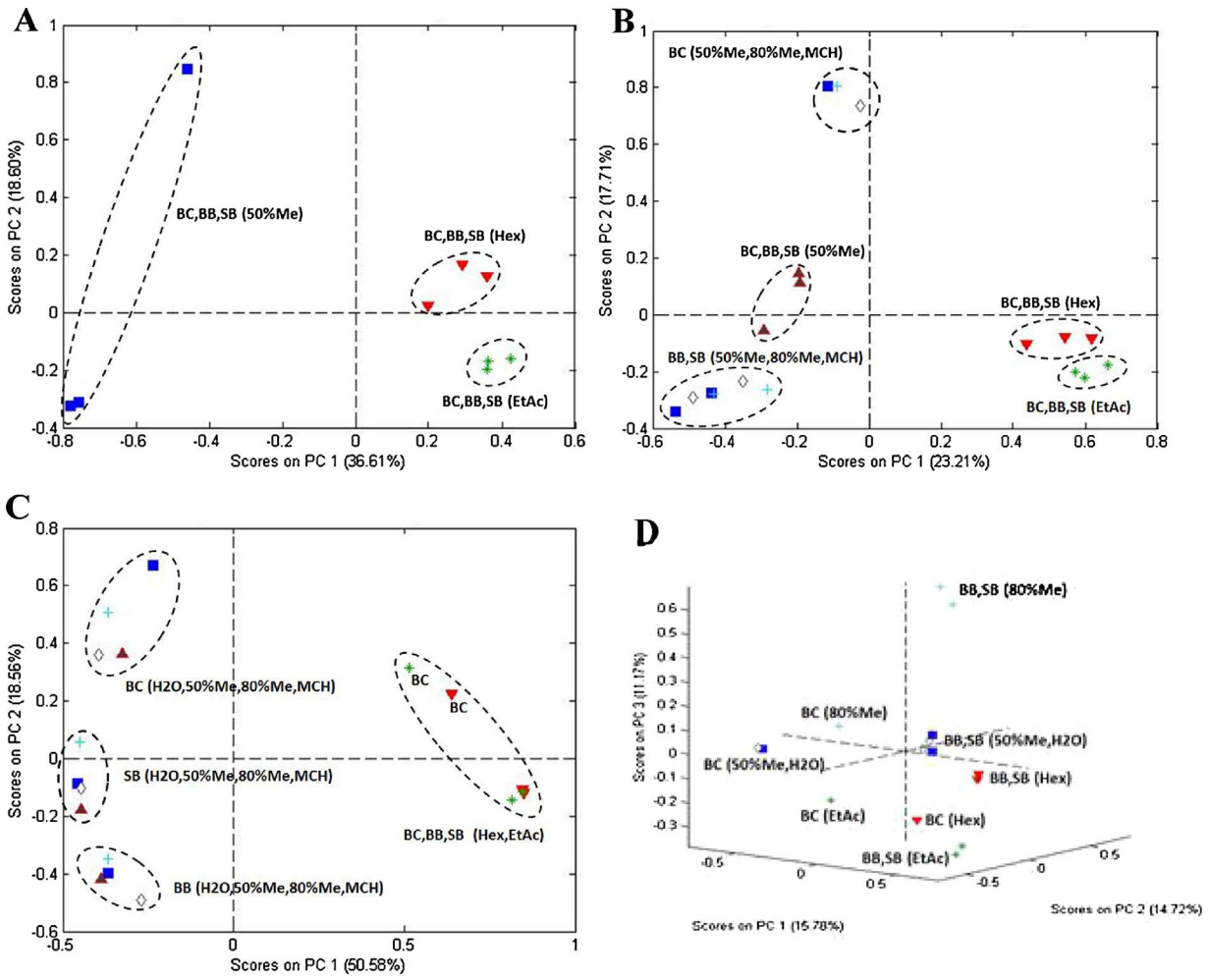

Fig. 3. PCA score plot for LC-UV chromatograms for (A) hexane, ethyl acetate, and 50% methanol extractions and (B) hexane, ethyl acetate, methanol-chloroform-water, 80&% methanol, 50% methanol, and H_20. PCA score plot for LC–MS chromatograms for (C) hexane, ethyl acetate, and 50% methanol extractions and (D) hexane, ethyl acetate, methanol-chloroform-water, 80% methanol, 50% methanol, and H_20.

for Hex and EtAc cluster together on the right with the BC chromatograms showing some difference. On the left side of the score plot, the clusters for 50Me, 80Me, MCH, and H_2O are based on the plant materials.

Fig. 3D shows the PCA score plot for sequential and direct extraction with HPLC–MS detection. The clusters observed for HPLC–MS and GC–MS are no longer present. All the chromatograms for BC, with the exception of 50Me and H_2O, are well separated from each other and those for BB and SB. Similarly, all the chromatograms for BB and SB are well separated from each other, again with the exception of 50Me and H_2O. Thus, MS detection appears to be more sensitive to the chemical differences of the extracts. MS clearly establishes the difference between each solvent for the same plant material and the difference between each plant material for the same solvent.

The close clusters for hex and EtAc in Fig. 3 suggest that these solvents could not be used in conjunction with chemometrics to discriminate between the different plant materials because of the similarity of their chromatograms. However, it must be remembered that the size of the PCA plot is determined by the total variance of the data set. Fig. 4A shows the PCA score plot of the individual chromatograms (each point in Fig. 3A is the average of 4 chromatograms) for the 3 plant materials obtained with EtAc extraction. Similar results were obtained for Hex (data not shown). It can be seen that the clusters for each of the 3 plant materials are well separated. Fig. 4B shows a similar plot for extraction with H_2O (at the other end of the polarity spectrum) that shows similar separation of the plants. Similar results were obtained for 50Me, 80Me, and MCH (data not shown).

It was found, that any of the 6 solvents, used either sequentially or directly, could be used to differentiate between BC, BB, and SB. Thus, almost any sub-sample of the plant metabolome can be used to differentiate plants from each other. This has been reported in previous studies However, in cases of closely related cultivars or marginal differences in the growing conditions, metabolomics variation may not be clearly visible with single extractions. Thus, detailed sequential or direct extractions with solvents of varying polarities may provide improved potential to classify and differentiate samples. This can also be achieved with detail chromatographic separation and analysis of different metabolites.

3.5. Identification of metabolites

An attempt was made to identify the major compounds found in the extracts of each of the plant materials using GC–MS (after BSTFA/TMCS derivatization) and HPLC–MS. GC–MS data proved easier to analyze since the extraction process resulted in fewer ions in the chromatogram. HPLC–MS provided many more ions, many of which could not be identified.

3.5.1. GC–MS

Table 2 lists the 10 compounds identified by GC–MS analysis of BB extracts, namely, L-valine, glycerol, L-methionine, L-tyrosine, L-proline, D-galactose, D-mannose, galactaric acid, *myo*-inositol, and sucrose

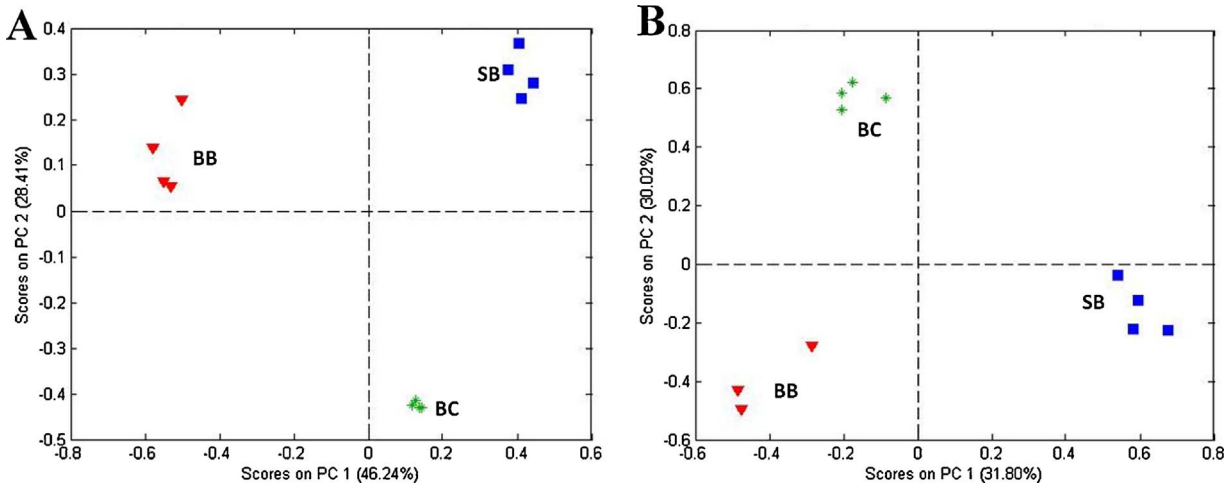

Fig. 4. PCA score plot for LC-UV chromatograms for (A) ethyl acetate and (B) H_2O.

Table 2
Metabolites discriminating between sequential and conventional extraction methods analyzed by LC–MS.

S.No	t_R	Name	BB	BC	SB	Reference
1	6.85	L-valine	D	x	x	Standard, NIST
2	7.39	L-serine	x	D	x	Standard, NIST
3	7.63	Glycerol	D	D	D	Standard, NIST
4	7.92	L-threonine	D	x	x	Standard, NIST
5	8.11	Butanedioic acid	x	D	D	Standard, NIST
6	9.19	L-methionine	D	x	x	Standard, NIST
7	9.6	L-glycine	x	D	D	Standard, NIST
8	9.66	L-tyrosine	D	x	x	Standard, NIST
9	10.93	L-proline	D	D	x	Standard, NIST
10	10.97	L-glutamic acid	x	x	D	Standard, NIST
11	11.2	L-phenylalanine	x	D	x	Standard, NIST
12	11.2	L-alanine	x	x	D	Standard, NIST
13	15.13	Sorbose	x	D	x	Standard, NIST
14	15.28	D-galactose	D	D	x	Standard, NIST
15	15.05	D-fructose	x	D	x	Standard, NIST
16	15.34	D-mannose	D	x	D	Standard, NIST
17	15.73	D-mannitol	x	x	D	Standard, NIST
18	15.9	Inositol	x	x	D	Standard, NIST
19	16.37	D-Gluconic acid	x	D	D	Standard, NIST
20	16.54	Galactaric acid	D	x	D	Standard, NIST
21	17.16	Myo-inositol	D	D	D	Standard, NIST
22	17.99	Octadecadienoic acid	x	x	D	Standard, NIST
23	21.97	Sucrose	D	D	D	Standard, NIST
24	22.38	Maltose	x	D	D	Standard, NIST

t_R: retention time, BB: black bean, BC: black cohosh, SB: soybean, D: detected, x: not detected.

causing differentiation among various extracts (Table 2). These compounds were annotated based on the comparison of experimentally determined retention times and mass spectral details with values for standards in the NIST library. The same amino acid compounds were reported by Słupski using an amino acid analyzer in a study of the genotypic variation and processing changes in the amino acid content of three different bean varieties [26]. Higher amino acid content was observed in the samples extracted with methanol:chloroform:water (2.5:1:1, v/v/v) as compared to the water extract. However, the content of sugars and its derivatives were higher in water extract as compared to samples extracted with methanol:chloroform:water (2.5:1:1, v/v/v) [11].

Similarly GC–MS analysis of SB identified glycerol, butanedioic acid, L-glycine, L-glutamic acid, L-alanine, D-mannose, D-mannitol, inositol, D-gluconic acid, galactaric acid, *myo*-inositol, octadecadienoic acid, sucrose, and maltose. The same metabolites (sugars, free amino acid and organic acids) were previously reported by Song et al. in several Chinese soybean varieties using Nuclear magnetic Resonance Spectroscopy [27]. The authors reported that nutrition profiles of different soybean varied significantly. Some varieties were enriched in asparagine and malate, whereas others had higher sucrose levels or contained greater amounts of alanine, glutamate, glycine, and serine. In a separate study by Moldes et al. authors investigated changes in the free amino acid content in glyphosate resistant and susceptible soybean lines [28]. Similar amino acids were reported as observed in the present study. The authors showed that with PCA, cluster analysis (CA), and linear discriminant analysis (LDA) could differentiate between soybean samples of different genetic origin.

3.5.2. HPLC–MS

Table 3 lists the major metabolites identified in the sequential and direct extracts that were analyzed by HPLC–MS without derivatization. Compounds in the water extract were distinctly different from those in the methanol: water (1:1, v/v), methanol: water (8:2, v/v) and methanol (50Me and 80Me) extracts. The variation in the BC aqueous alcohol and water extracts may be attributed to the presence of triterpene glycosides which are absent in the other two plant substrates (BB and SB). The phenolic metabolites that contributed to variations between the extracts were putatively identified based on comparison of their mass, retention times, and UV spectra. However the tentative identification of triterpenoids in BC, was achieved primarily through the comparison of parent and product ions with previously reported data [15,16].

The variation in the 80Me extract may be attributed to efficient extraction of phenolic acids. Higher yields of phenolic acids were extracted from black cohosh with varying methanol water mixtures [29]. The authors observed highest yields with 50% DMSO. However, extraction, evaporation, and odor of DMSO limits its use as a solvent. In a separate study on extraction of phenolic acids from eggplant, the authors investigated various proportion of methanol water solvent mixtures on extraction efficacy of phenolic acids. The results showed that 80% methanol was optimum for extracting the phenolic acids [29].

Results from the present study revealed that 18 metabolites showed variations with BB. Similar compounds have been reported in the previous studies of dry beans [22,30]. In the case of SB, daidzin, glycitin, genistin, M-genistein, daidzein, and glycitein were identified as contributing to the variation among the extracts. Several publications are reported in the literature indicating the presence of similar metabolites in soybeans. In a previous study, this laboratory observed that the efficiency of extraction was significantly influenced by the technique and solvent of the method [31]. Most studies on BC have focused on the isolation and identification of triterpenoids, saponins and phenolics [15,16,32–34]. Detection was highest in BC with positive ionization.

Table 3
Metabolites discriminating between sequential and conventional extraction methods analyzed by LC–MS.

S.No	R.Time	[M−H]−/[M+H]+	MS^2	Putative metabolite	Reference
Discriminating black bean (BB) metabolites					
1	3.84	341.1087/-	179	Caffeoylglucose	Abu-Reidah et al. [35]
2	10.10	425.1216/-	263	Dalbinol	Abu-Reidah et al. [35]
3	11.43	639.1729/641.1902	320	Quercetin 3-glucuronide-7-glucoside	Abu-Reidah et al. [35]
4	14.40	521.1533/-	503	Isolariciresinol 9-*O*-B-D-glucopyranoside II	Abu-Reidah et al. [35]
5	16.99	593.1265/-	285	Kaempferol 3-rutinoside I	Abu-Reidah et al. [35]
6	17.60	431.1179/-	137	Primeveroside salicylic acid II	Abu-Reidah et al. [35]
7	18.59	461.1278/-	167	Sibiricose A3	Abu-Reidah et al. [35]
8	22.75	355.0655/	191	Ferulic acid β-glucoside II	Abu-Reidah et al. [35]
9	24.07	491.1385/493.1332	370	Isorhamnetin 3-glucuronide	Abu-Reidah et al. [35]
10	25.88	477.1019/479.1176	315	Quercetin 3-O-glucuronide	Abu-Reidah et al. [35]
11	27.37	563.1589/565.1759	387	Schaftoside III	Abu-Reidah et al. [35]
12	35.60	741.1840/-	299	Quercetin 3-O-xylosylrutinoside I	Abu-Reidah et al. [35]
13	38.22	609.1429/-	301	Quercetin rutinoside I	Abu-Reidah et al. [35]
14	39.80	429.1744/-	387	Formononetin 7-o-glucoside	Abu-Reidah et al. [35]
15	40.11	755.2358/-	417	Kaempferol 3-*O*-rutinoside-7-*O*-glucose III	Abu-Reidah et al. [35]
16	41.09	463.0859/465.1024	301	Hesperetin 7-glucoside	Abu-Reidah et al. [35]
17	41.92	579.1687/-	271	Kaempferol 3-sambubioside I	Abu-Reidah et al. [35]
18	42.88	447.0915/449.1077	284	Kaempferol 3-O-glucuronide	Abu-Reidah et al. [35]
Discriminating black cohosh (BC) metabolites					
1	08.18	-/144.1129	112	Proline betaine	Nikolić et al. [25]
2	08.36	-/160.108	101	δ-guanidinovaleric acid	Nikolić et al. [25]
3	09.47	-/268.104	136	Adenosine	Nikolić et al. [25]
4	10.64	-/220.1178	202	Panthotenic acid	Nikolić et al. [25]
5	12.44	-/186.1235	144	Cimipronidine methyl ester	Nikolić et al. [25]
6	12.61	-/166.086	120	Norsalsolinol	Nikolić et al. [25]
7	24.39	-/351.1663	334	N-feruloyl arginine	Nikolić et al. [25]
8	25.03	-/370.2071	162	Allocryptopine	Nikolić et al. [25]
9	27.51	-/342.1698	297	Magnoflorine	Nikolić et al. [25]
10	30.12	-/365.1823	172	N-isoferuloyl arginine methylester	Nikolić et al. [25]
11	33.70	-/504.3372	459	N-feruloyl phenylalanine-4′-O hexoside	Nikolić et al. [25]
12	38.34	-/337.1395	–	Caffeoyl arginine	Nikolić et al. [25]
13	48.66	-/476.1558	177	N-feruloyl tyramine-4‴-O-hexoside	Nikolić et al. [25]
Discriminating Soybean (SB) metabolites					
1	27.78	431.1180/433.1320	269	Genistin	Simons et al. [36]
2	31.58	517.2280/519.1115	–	malonyl-genistin	Simons et al. [36]
3	35.26	[a]461.1084/417.1168	253	Daidzin	Simons et al. [36]
4	36.55	[a]491.1180/447.1264	283	Glycitin	Simons et al. [36]
5	39.51	415.1019/	253	Daidzin	Simons et al. [36]
6	41.38	253.0494/255.0644	–	Daidzein	Simons et al. [36]
7	42.15	253.0503/	–	Daidzein	Simons et al. [36]
8	42.32	283.0602/285.0747	151	Glycitein	Simons et al. [36]
9	44.8	269.0453/271.0596	225	Genestein	Simons et al. [36]
10	47.58	-/285.0747	–	Biochanin A	Simons et al. [36]
11	50.52	269.0449/271.0592	–	Genestein	Simons et al. [36]

Results also revealed that 13 metabolites significantly varied between the direct and sequential extraction process.

4. Conclusion

The identity and quantity of compounds extracted from black cohosh, black beans, and soy beans were strongly influenced by the solvent and the method of analysis. General conclusions are difficult since the extraction efficiencies and the instrument responses are strongly convoluted. It would appear the each pair of conditions (extraction and instrument) analyze different compounds with a less than 50% overlap. However, each pair of conditions was successful in differentiating between the three plant materials. The methodology used in the present study may prove to be of value in differentiating complex substrates as it provides a holistic approach to extract and analyze metabolites of varying polarities.

Acknowledgements

The authors wish to acknowledge Mrs. Samina Shami of USDA-ARS and student Mr. David Darko for their assistance in carrying out experimentation. This work was supported financially by the Agricultural Research Service of the U.S. Department of Agriculture and by an InterAgency Agreement with the Office of Dietary Supplements of the National Institutes of Health.

References

[1] S. Bajad, V. Shulaev, Highly-parallel metabolomics approaches using LC–MS for pharmaceutical and environmental analysis, TrAC, Trends Anal. Chem. 26 (2007) 625–636.
[2] Y. Okazaki, K. Saito, Recent advances of metabolomics in plant biotechnology, Plant Biotechnol. Rep. 6 (2012) 1–15.
[3] D.F. Gomez-Casati, M.I. Zanor, M.V. Busi, Metabolomics in plants and humans: applications in the prevention and diagnosis of diseases, BioMed Res. Int. (2013) ID 792527, 11 pages.
[4] I. Tzoulaki, T.M. Ebbels, A. Valdes, P. Elliott, J.P. Ioannidis, Design and analysis of metabolomics studies in epidemiological research: a primer on—omic technologies, Am. J. Epidemiol. 180 (2014) 129–139.
[5] T. Zhang, D.G. Watson, A short review of applications of liquid chromatography mass spectrometry based metabolomics techniques to the analysis of human urine, Analyst 140 (2015) 2907–2915.
[6] K.N. Patel, J.K. Patel, M.P. Patel, G.C. Rajput, H.A. Patel, Introduction to hyphenated techniques and their applications in pharmacy, Pharm. Methods 1 (2010) 2–13.
[7] S. Wernisch, S. Pennathur, Evaluation of coverage retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics, Anal. Bioanal. Chem. 408 (2016) 6079–6091.
[8] D.L. Luthria, S. Mukhopadhyay, Influence of sample preparation on assay of

phenolic acids from eggplant, J. Agric. Food Chem. 54 (2006) 41–47.
[9] C. Hu, G. Xu, Mass-spectrometry-based metabolomics analysis for plantomics, TrAC, Trends Anal. Chem. 52 (2013) 36–46.
[10] Y. Shen, T. Fatemeh, L. Tang, Z. Cai, Quantitative metabolic network profiling of Escherichia coli: an overview of analytical methods for measurement of intracellular metabolites, TrAC, Trends Anal. Chem. 75 (2016) 141–150.
[11] K.M. John, D. Luthria, Amino acid organic acid, and sugar profiles of 3 dry bean (*Phaseolus vulgaris* L.) varieties, J. Plant Sci. 80 (2015) C2662–C2669.
[12] B.M. Hounoum, H. Blasco, P. Emond, S. Mavel, Liquid chromatography–high-resolution mass spectrometry-based cell metabolomics: experimental design, recommendations, and applications, TrAC, Trends Anal. Chem. 75 (2016) 118–128.
[13] Y. Liu, L. Li, Y. Xiao, J.Q. Yao, P.Y. Li, D.R. Yu, Y.L. Ma, Global metabolite profiling and diagnostic ion filtering strategy by LC-QTOF MS for rapid identification of raw and processed pieces of Rheum palmatum L, Plant Chem. 192 (2016) 531–540.
[14] K.M.M. John, E.S. Jung, S. Lee, J.S. Kim, C.H. Lee, Primary and secondary metabolites variation of soybean contaminated with *Aspergillus sojae*, Food Res. Int. 54 (1) (2013) 487–494.
[15] B. Jiang, F. Kronenberg, P. Nuntanakorn, M.H. Qiu, E.J. Kennelly, Evaluation of the botanical authenticity and phytochemical profile of black cohosh products by high-performance liquid chromatography with selected ion monitoring liquid chromatography-mass spectrometry, J. Agric. Plant Chem. 54 (2006) 3242–3253.
[16] K. He, B. Zheng, C.H. Kim, L. Rogers, Q. Zheng, Direct analysis and identification of triterpene glycosides by LC/MS in black cohosh, *Cimicifuga racemosa*, and in several commercially available black cohosh products, Planta Med. 66 (2000) 635–640.
[17] N.G.T. Meneses, S. Martins, J.A. Teixeira, S.I. Mussatto, Influence of extraction solvents on the recovery of antioxidant phenolic compounds from brewer's spent grains, Sep. Purif. Technol. 108 (2013) 152–158.
[18] J. Suberu, P.S. Gromski, A. Nordon, A. Lapkin, Multivariate data analysis and metabolic profiling of artemisinin and related compounds in high yielding varieties of Artemisia annua field-grown in Madagascar, J. Pharm. Biomed. Anal. 117 (2016) 522–553.
[19] K. Dettmer, N. Nürnberger, H. Kaspar, M.A. Gruber, M.F. Almstetter, P.J. Oefner, Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols, Anal. Bioanal. Chem. 399 (2011) 1127–1139.
[20] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique, Anal. Bioanal. Chem. 402 (2012) 231–247.
[21] D.L. Luthria, Optimization of extraction of phenolic acids from a vegetable waste product using a pressurized liquid extractor, J. Funct. Plants 4 (2012) 842–850.
[22] L.Z. Lin, J.M. Harnly, M.S. Pastor-Corrales, D.L. Luthria, The polyphenolic profiles of common bean (*Phaseolus vulgaris* L.), Plant Chem. 107 (2008) 399–410.
[23] A. Oikawa, T. Otsuka, R. Nakabayashi, Y. Jikumaru, K. Isuzugawa, H. Murayama, K. Saito, K. Shiratake, Metabolic profiling of developing pear fruits reveals dynamic variation in primary and secondary metabolites, including plant hormones, PLoS One 10 (2015) e0131408.
[24] Z.Y. Cao, L.H. Sun, R.X. Mou, L.P. Zhang, X.Y. Lin, Z.W. Zhu, M.X. Chen, Profiling of phytohormones and their major metabolites in rice using binary solid-phase extraction and liquid chromatography-triple quadrupole mass spectrometry, J. Chromatogr. A 1451 (2016) 67–74.
[25] D. Nikolić, T. Gödecke, S.N. Chen, J. White, D.C. Lankin, G.F. Pauli, R.B. Van Breemen, Mass spectrometric dereplication of nitrogen-containing constituents of black cohosh (*Cimicifuga racemosa* L.), Fitoterapia 83 (2012) 441–460.
[26] J. Słupski, Effect of cooking and sterilisation on the composition of amino acids in immature seeds of flageolet bean (*Phaseolus vulgaris* L.) cultivars, Plant Chem. 121 (2010) 1171–1176.
[27] J. Song, C. Liu, D. Li, Z. Gu, Evaluation of sugar free amino acid, and organic acid compositions of different varieties of vegetable soybean (*Glycine max* [L.] Merr), Ind. Crops Prod. 50 (2013) 743–749.
[28] C.A. Moldes, J.M. Camiña, L.O. Medici, S.M. Tsai, R.A. Azevedo, Physiological effects of glyphosate over amino acid profile in conventional and transgenic soybean (*Glycine max*), Pestic. Biochem. Physiol. 102 (2012) 134–141.
[29] S. Mukhopadhyay, D.L. Luthria, J.R. Robbins, Optimization of extraction process for phenolic acids from black cohosh (*Cimicifuga racemose*) by pressurized liquid extraction, J. Sci .Plant Agric. 86 (2006) 156–162.
[30] S. Aisyah, H. Gruppen, S. Andini, M. Bettonvil, E. Severing, J.P. Vincken, Variation in accumulation of isoflavonoids in Phaseoleae seedlings elicited by *Rhizopus*, Plant Chem. 196 (2016) 694–701.
[31] D.L. Luthria, R. Biswas, S. Natarajan, Comparison of extraction solvents and techniques used for the assay of isoflavones from soybean, Plant Chem. 105 (2007) 325–333.
[32] B. Avula, Y.H. Wang, C.S. Rumalla, Z. Ali, T.J. Smillie, I.A. Khan, Analytical methods for determination of magnoflorine and saponins from roots of *Caulophyllum thalictroides* (L.) Michx. Using UPLC, HPLC and HPTLC, J. Pharm. Biomed. Anal. 56 (2011) 895–903.
[33] C. Ma, A.R. Kavalier, B. Jiang, E.J. Kennelly, Metabolic profiling of Actaea species extracts using high performance liquid chromatography coupled with electrospray ionization time-of-flight mass spectrometry, J. Chromatogr. A 1218 (2011) 1461–1476.
[34] B. Avula, Y.H. Wang, T.J. Smillie, I.A. Khan, Quantitative determination of triterpenoids and formononetin in rhizomes of black cohosh (*Actaea racemosa*) and dietary supplements by using UPLC-UV/ELS detection and identification by UPLC–MS, Planta Med. 75 (2009) 381–386.
[35] I.M. Abu-Reidah, D. Arráez-Román, J. Lozano-Sánchez, A. Segura-Carretero, A. Fernández-Gutiérrez, Phytochemical characterisation of green beans (*Phaseolus vulgaris* L.) by using high-performance liquid chromatography coupled with time-of-flight mass spectrometry, Phytochem. Anal. 24 (2013) 105–116.
[36] R. Simons, J.P. Vincken, N. Roidos, T.F. Bovee, M. Van Iersel, M.A. Verbruggen, H. Gruppen, Increasing soy isoflavonoid content and diversity by simultaneous malting and challenging by a fungus to modulate estrogenicity, J. Agric. Food Chem. 59 (2011) 6748–6758.