

## METABOLITE PROFILING OF SOY BY-PRODUCTS: A COMPREHENSIVE APPROACH

Felipe Sanchez Bragagnolo<sup>1,2</sup>, Gerardo Álvarez-Rivera<sup>2</sup>, Márcia Cristina Breitzkreitz<sup>3</sup>, Elena Ibáñez<sup>2</sup>,  
Alejandro Cifuentes<sup>2</sup>, Cristiano Soleo Funari<sup>1\*</sup>

<sup>1</sup> Green Biotech Network, School of Agricultural Sciences, São Paulo State University (UNESP),  
Botucatu, São Paulo - 18610-034, Brazil

<sup>2</sup> Laboratory of Foodomics, Institute of Food Science Research (CIAL-CSIC), Madrid - 28049, Spain

<sup>3</sup> Institute of Chemistry, University of Campinas (UNICAMP), Campinas, São Paulo - 13083-971,  
Brazil

\*E-mail: cristiano.funari@unesp.br

1 **ABSTRACT**

2 Soy is the major oilseed crop as soybeans are widely used to produce biofuel, food, and feed. Other  
3 parts of the plant have been left on the ground after the harvest. The accumulation of such by-products  
4 on the soil can cause environmental problems. This work presents for the first time a comprehensive  
5 metabolite profiling of soy by-products collected directly from the ground just after mechanical  
6 harvesting. A two-liquid-phase extraction using *n*-heptane and EtOH-H<sub>2</sub>O 7:3 (v/v) provided extracts  
7 with a complete characterization by gas chromatography and ultra-high-performance liquid  
8 chromatography both coupled to time-of-flight mass spectrometry. A total of 146 metabolites,  
9 including flavones, flavonols, isoflavonoids, fatty acids, steroids, mono-, sesqui-, di-, and  
10 triterpenoids, were tentatively identified in soy by-products and soybeans. These proved to be sources  
11 of a wide range of bioactive metabolites, thus suggesting that they could be valorized while reducing  
12 potential environmental damage in line with a circular economy model.

13

14

15

16

17

18

19

20 **Keywords:** Metabolomics; Waste valorization; Bioeconomy; Biorefinery; Green chemistry

21

22

## 23 INTRODUCTION

24 Plants are an important source of compounds with a wide range of therapeutic uses.<sup>1</sup> Agricultural  
25 by-products, a highly abundant source of natural products, can still be considered underexplored for  
26 this purpose.<sup>2-4</sup> Most of them are currently used for low value-added applications, such as animal feed,  
27 energy production, or simply left in the field with the aim of maintaining soil fertility.<sup>2-4</sup> However, an  
28 excessive accumulation of such by-products from consecutive harvests can also contribute to pest and  
29 weed infestation, greenhouse gas emission, soil microbiome deregulation, and water and soil  
30 contamination.<sup>4-7</sup> These problems are expected to increase due to the high demand for food caused by  
31 population growth.<sup>8</sup> Efforts to mitigate such problems through sustainable chemistry and engineering  
32 approaches are advocated by the United Nations (UN), the European Environment Agency, among  
33 other institutions. These institutions warn of the urgent need to convert agricultural by-products into  
34 raw-materials for high value-added products employing green technologies.<sup>9,10</sup>

35 The soy crop (*Glycine max* (L.) Merr.) deserves special attention, since it is the major oilseed crop  
36 worldwide.<sup>11</sup> The production of soybeans and associated by-products, which are left on the fields (i.e.,  
37 branches, leaves, pods, and roots), were expected to be around 362 and 597 million tonnes in 2020/21,  
38 respectively.<sup>12,13</sup> Chemical studies of soy plants at different growth stages reported a wide dynamic  
39 range of metabolites in their different organs, such as flavonoids, isoflavonoids, prenol lipids, fatty  
40 acids, and other organic acids, with potential applications in cosmetic, food, chemical, and  
41 pharmaceutical industries.<sup>14-16</sup> However, to the best of our knowledge, the literature does not report  
42 the untargeted metabolite profiling of soy by-products left on the ground after harvest. Previous works  
43 focused only in isoflavonoids and in an alkaloid (trigonelline) in such materials.<sup>17,18</sup>

44 Thus, a comprehensive extraction and identification procedure to tentatively identifying a widely  
45 range of metabolites in soy by-products would be desirable to estimate alternative applications for  
46 them. Previous works with sugarcane and coffee by-products evidenced that the best conditions for

47 one type of by-product would not necessarily be the best for another one.<sup>19,20</sup> In other words, an  
48 optimization should be pursued for each type of complex matrix. A Design of Experiment (DoE)  
49 should be preferred to a trial-and-error approach to achieve a global optimization of the extraction  
50 procedure in a minimum number of experiments.<sup>21</sup> Unlike DoE, a trial-and-error approach is very often  
51 a time and resource (i.e., solvents and energy) consuming strategy, involving a large number of  
52 experiments without leading to reliable optimization, in a clear contradiction with Green Chemistry  
53 principles.<sup>22</sup>

54 Therefore, this work aimed to conduct metabolite profiling for the first time of soy by-products  
55 collected directly from the ground just after mechanical harvesting (branches, leaves, pods, and roots).  
56 To achieve this goal, DoE was initially employed to optimize a (i) an ultra-high performance liquid  
57 chromatography-PDA/UV-mass spectrometry (UHPLC-PDA/UV-MS) method, (ii) a gas  
58 chromatography-mass spectrometry (GC-MS) method, and (iii) a two-liquid phase extraction to extract  
59 simultaneously polar, medium polarity, and nonpolar compounds. Subsequently, (iv) the optimized  
60 procedures were applied for tentative identification of metabolites in each soy by-product (and in the  
61 commodity soybeans, which were used as reference material) by employing a dereplication strategy  
62 that crosses chromatographic, spectral, and chemotaxonomic information. Global Natural Product  
63 Social Molecular Networking (GNPS) was employed for tentative identification of high to medium  
64 polarity compounds in hydroethanolic extracts, while GC-MS data were matched against NIST  
65 (National Institute of Standards and Technology) and Fiehn libraries.

66

## 67 **MATERIALS AND METHODS**

### 68 *Reagents and materials*

69 HPLC grade ethanol and MS grade acetonitrile and methanol were purchased from Merck  
70 (Germany). P.A. (practical grade) formic acid and *n*-heptane were from Sigma Aldrich (Germany).

71 *Plant material*

72 Soy by-products from NA5909 cultivars were collected directly from the ground after mechanical  
73 harvesting of soybean at the School of Agricultural Sciences of São Paulo State University in Botucatu  
74 city, São Paulo State, Brazil (-22.8296354, -48.42553) and Santa Fé Farm in Pardinho city, São Paulo  
75 State, Brazil (-23.0268717, -48.3859183). The collected materials were then separated into branches,  
76 leaves, pods, and roots. Similar mass portions of the same type of material (e.g., pods) from the two  
77 collection sites were mixed and grounded in a basic analytical mill (IKA® A11, Germany). For  
78 extraction optimization purposes, the resulting material was separated by ranges of granulometry on  
79 an electromagnetic sieve shaker (Bertel, Brazil).

80

81 *Optimization of methods by Design of Experiments (DoE)*

82 Optimizations were pursued using Protimiza Experimental Design® software (Protimiza  
83 Experimental Design, Brazil) and GraphPad Prism 9 software (GraphPad Software, USA).

84

85 *Method optimization of ultra-high-performance liquid chromatography coupled to photodiode*  
86 *array and mass spectrometry (UHPLC-PDA/UV-MS)*

87 The liquid chromatography analyses were performed using a UHPLC Nexera UC system  
88 (Shimadzu, Japan), consisting of two LC-30AD pumps, SIL-30AC auto-injector, CTO-20AC column  
89 oven, SPD-M20A photodiode network detector, unit degasser DGU-20A5R, coupled to a LCMS-2020  
90 single quadrupole mass spectrometer equipped with an ESI ionization source and a CBM-20A  
91 communications module. The separations were achieved in a Zorbax Eclipse Plus C18 column, 150  
92 mm × 2.1 mm; 1.8 µm (Agilent Technologies, USA), a widely used column that has already been  
93 employed for soy metabolomics.<sup>23-25</sup> A 1 µL UltiMate™ 3000 Pre-Heater (Thermo Fisher Scientific,  
94 USA) was installed inside the column compartment. A two-factor central composite rotatable design

95 (CCRD,  $\alpha = (2^k)^{1/4}$ ) was employed. Initial % of B (MeOH) ( $x_1$ ) and % of formic acid (HCOOH) in the  
96 mobile phases ( $x_2$ ) were the selected factors. The flow rate and analysis temperature were kept at 0.3  
97 mL/min and 30 °C, respectively. An aliquot of 2  $\mu$ L of a non-concentrated EtOH:H<sub>2</sub>O 7:3 (v/v)  
98 (prepared as the central point described in the section “method optimization of two-liquid-phase  
99 dynamic maceration (2-Mac)”) was filtered with a syringe-filter of 0.22- $\mu$  nylon prior to injection.  
100 The CCDR (Central Composite Rotatable Design) itself as well as the outcomes are shown in Table  
101 1. The monitored responses were the total number of peaks detected by PDA/UV at 254 nm ( $y_1$ ) (the  
102 samples analyzed here were expected to be rich in subclasses of flavonoids which absorbs at this  
103 wavelength)<sup>26</sup> and the sum of the peaks in positive (ESI+) and negative (ESI-) ionization mode for MS  
104 analysis ( $y_2$  and  $y_3$ , respectively). Only peaks with signal/noise (S/N) greater than 3 were considered.  
105 Mass spectrometer parameters were set as: nebulization gas flow at 1.5 L/min; drying gas flow at 15  
106 L/min; heat block temperature at 400 °C; dissolution line temperature at 250 °C, and a voltage detector  
107 of 0.1 kV. The optimized condition was H<sub>2</sub>O (A) and MeOH (B), both acidified with 0.1% formic acid  
108 (v/v) in the following gradient elution: 15–100% B from 0 to 60 min, at 0.3 mL/min and 30 °C. The  
109 data were processed using Shimadzu’s LabSolutions software.

110

#### 111 *Method optimization of gas chromatography coupled to mass spectrometry (GC-MS)*

112 Gas chromatography analyses were performed using a GC-MS Clarus 680 chromatograph (Perkin  
113 Elmer, USA) coupled to a Clarus SQ 8 T mass Detector. Chromatographic separations were performed  
114 using a PerkinElmer Elite-5MS column (30 m  $\times$  0.25 mm, 0.25  $\mu$ m). A two-factor CCDR design was  
115 used to optimize the chromatographic separation, where  $x_1$  = temperature rate and  $x_2$  = carrier gas flow  
116 (Table 2). The total number of peaks was the monitored response. The final optimized separation was:  
117 the oven temperature was held isothermal at 150 °C for 2 min and then increased to 350 °C at the rate  
118 of 7 °C/min. The GC injector and MS transfer line temperatures were set at 330 °C and 310 °C,  
119 respectively. The injection volume of the non-concentrated *n*-heptane extract (central point described  
6

120 on 2.3.3) was 2  $\mu$ L with a split flow of 5 ml/min. Helium gas was used as the carrier gas at a constant  
121 flow rate of 1.6 ml/min, and 6 min solvent delay. The ionization of the sample was performed in EI  
122 mode (70 eV), and the acquisition mass range was set at 50–600 amu.

123

#### 124 *Method optimization of two-liquid-phase dynamic maceration (2-Mac)*

125 Two-liquid-phase dynamic maceration (2-Mac) with magnetic stirring with a temperature sensor  
126 (Heidolph MR Hei-Tec, Germany) was selected as the extraction technique to be optimized.<sup>27</sup> A mix  
127 of the different parts of the soy by-products collected directly from the ground on the day of mechanical  
128 harvesting, plus soybeans (60 mg of beans, branches, leaves, pods, and roots, in a total of 300 mg) was  
129 added to a 22.5 mm inner diameter beaker. Then, 3 mL of EtOH-H<sub>2</sub>O 7:3 (v/v) and 3 mL of *n*-heptane  
130 were added into the beaker. The other parameters varied according to a four-factor Doehlert design as  
131 shown in Table 3. All produced extracts were centrifuged at 4000 rpm for 5 min (Eppendorf 5810 R,  
132 Germany) to speed phase separation. The simultaneously produced EtOH-H<sub>2</sub>O 7:3 (v/v) and *n*-heptane  
133 extracts were then syringe filtered (0.22  $\mu$ m nylon filter) and analyzed by UHPLC-PDA/UV-MS and  
134 GC-MS, respectively (see final conditions of 2.3.1 and 2.3.2). The sum of peaks observed in both types  
135 of chromatograms was selected as the response (*y*) to be monitored in each experiment presented in  
136 Table 3. The final optimized condition of extraction was Time: 35 min; Temperature: 45 °C; Magnetic  
137 stirring: 1400 rpm; Granulometry of the plant material: 48–115 Mesh. For peak annotation purposes,  
138 this condition was applied to the individual parts of soy by-products or soybeans which were then  
139 analyzed by UHPLC-QToF-MS/MS and GC-QToF MS.

#### 140 *Ultra-high-performance liquid chromatography coupled to time-of-flight mass spectrometry*

##### 141 *(UHPLC-QToF-MS/MS)*

142 Metabolite profiling of concentrated EtOH-H<sub>2</sub>O 7:3 (v/v) extracts (10 mg/mL) was carried out in  
143 an Agilent 1290 UHPLC system (Agilent Technologies, USA) coupled to an Agilent 6540 quadrupole-

144 time-of-flight mass spectrometer (QToF-MS), equipped with an orthogonal ESI source (Agilent Jet  
145 Stream, USA). The UHPLC condition was the optimized condition described in previously.  
146 Additionally, a shorter UHPLC-QToF-MS/MS analysis in negative mode was performed using a  
147 Zorbax Eclipse Plus C18 column, 100 × 2.1 mm; 1.8 μm (Agilent Technologies, USA), and H<sub>2</sub>O and  
148 ACN, both acidified with 0.01% formic acid (v/v) (A and B, respectively) in the following gradient  
149 elution: 0–30% B in 0–7 min, 30–80% B in 7–9 min, 80–100% B in 9–11 min, 100% B in 11–13 min,  
150 and 0% B in 13–14 min. The flow rate and analysis temperature were kept at 0.5 mL/min and 30 °C,  
151 and the sample injection volume was 5 μL. MS parameters were capillary voltage, 4000 V; nebulizer  
152 pressure, 40 psi; drying gas flow rate, 10 L/min; gas temperature, 350 °C; skimmer voltage, 45 V;  
153 fragmentor voltage, 110 V. Both MS and Auto MS/MS modes were acquired in m/z values of 50 and  
154 1100 and 50–800, respectively, and at a scan rate of 5 spectra per second. Agilent Mass Hunter  
155 Qualitative analysis software (B.07.00) was applied for postacquisition data processing.<sup>28</sup> Metabolites  
156 present in the EtOH-H<sub>2</sub>O 7:3 (v/v) extracts were annotated using Global Natural Product Social  
157 Molecular Networking (GNPS) (<http://gnps.ucsd.edu>) and MZmine 2.53 software  
158 (<https://mzmine.github.io/>).<sup>29–31</sup> First, MS/MS data were converted to mzML format with  
159 ProteoWizard 3.0.6002 package MSConvert software (ProteoWizard, USA).

160 The converted files were uploaded to GNPS platform, and a molecular network was created using  
161 the online workflow (<https://ccms-ucsd.github.io/GNPSDocumentation/>). The data were filtered by  
162 removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were  
163 window filtered by choosing only the top six fragment ions in the +/- 50Da window throughout the  
164 spectrum. The precursor ion mass and MS/MS fragment ion tolerances were set to 0.02 Da. A cosine  
165 score above 0.65 and more than four matched peaks were used to create the edges of the network. The  
166 spectra in the network were searched against spectral libraries of GNPS. The library spectra were  
167 filtered in the same manner as the input data. The matches between network spectra and library spectra  
168 were required to have a score above 0.65 and at least four matched peaks. MolNetEnhancer, a tool of



169 GNPS, and Cytoscape version 3.7.2 (Cytoscape Consortium, USA) were used for the network  
170 visualization. The product ion spectra presented in GNPS were manually verified with previous  
171 literature annotation. The mass error tolerance adopted was  $\leq 10$  ppm (Table 4). The final classification  
172 of the compounds was performed using JChem for Excel 21.1.0.787 - ChemAxon  
173 (<https://www.chemaxon.com>) and NPClassifier, with the final classification using ClassyFire.<sup>32,33</sup>

174

#### 175 *Gas chromatography coupled to time-of-flight mass spectrometry (GC-QToF-MS)*

176 Metabolite profiling of concentrated *n*-heptane extracts (10 mg/mL) was done using a 7890B  
177 Agilent system (Agilent Technologies, USA) coupled to a quadrupole time-of-flight (QToF) 7200  
178 (Agilent Technologies, USA) equipped with an electronic ionization (EI) interface. An Agilent Zorbax  
179 DB5- MS + 10 m Duragard Capillary Column (30 m  $\times$  250  $\mu$ m  $\times$  0.25  $\mu$ m), similar to the column used  
180 in the GC-MS method optimization, was employed for chromatographic separation. The sample  
181 injection volume was 1  $\mu$ L using a split flow of 8.4 mL/min, and the other parameters were the same  
182 as in the optimized GC-MS method. Metabolites present in the *n*-heptane extracts were annotated using  
183 the match of mass spectra in the Agilent Mass Hunter Unknown Analysis tool and mass spectral  
184 databases (i.e., NIST MS Search v.2.0 and Fiehn Lib).

## 185 **RESULTS AND DISCUSSION**

### 186 *Optimization of Analytical methods*

187 As the main goal of this work was to identify the greatest possible number of metabolites in the  
188 different soybean crop by-products, the methods were optimized to obtain the greatest number of  
189 chromatographic peaks with signal/noise  $\geq 3$ .<sup>19,34</sup> For the optimization of separations by employing a  
190 two-factor central composite rotatable design (CCRD), UHPLC-PDA/UV-MS and GC-MS systems  
191 were used due to their higher availability when compared with the systems later employed in the  
192 application step, when high-resolution mass spectrometers were used for tentative identification of

193 compounds. At this optimization step, the injected samples in the UHPLC-PDA/UV-MS and GC-MS  
194 systems were a EtOH-H<sub>2</sub>O 7:3 (v/v) and a n-heptane extract, respectively, prepared with a mix of soy  
195 by-products and soybeans (reference material) by Two-Liquid-Phase Dynamic Maceration (2-Mac).  
196 The extraction condition was that of the central point of the experimental design that should be used  
197 later for the optimization of the extraction condition (Table 3). Once the UHPLC-PDA/UV-MS and  
198 GC-MS optimizations were achieved, they were used to monitor the extraction efficiency during the  
199 optimization of 2-Mac extraction itself.<sup>21</sup> In both cases, the simultaneously produced EtOH-H<sub>2</sub>O 7:3  
200 (v/v) and n-heptane extracts by 2-Mac were injected into the chromatographic systems just after  
201 filtration. This means without any concentration and resuspension step to enhance sample throughput  
202 while saving solvents and energy, in line with principles 1, 2, 6, 7, 9, 10, 11, and 12 of the 12 Principles  
203 of Green Analytical Chemistry (GAC).<sup>35</sup>

#### 204 *Optimization of the UHPLC-PDA/UV-MS method*

205 The variables and levels selected for the CCRD (Table 1) were based on preliminary tests and on a  
206 literature review.<sup>23-25,36</sup> Initial % of B (MeOH) was set as a factor in the CCRD (Table 1), because  
207 preliminary tests with mobile phase ranging from 5 to 100% of MeOH led to peaks mainly spread out  
208 in the second half of the chromatogram (late elutions - data not shown). Therefore, finding the gradient  
209 of B that best distributed the peaks across the chromatogram would lead to the highest number of  
210 peaks. On the other hand, % HCOOH in the mobile phase was selected as the other factor in the CCRD,  
211 because this acid has been used in soy metabolomics works and could improve chromatography  
212 separation and ionization efficiency in mass spectrometry.<sup>23-25,37-40</sup> Thus, selecting the best % of  
213 HCOOH would be also important. Methanol was chosen as the organic modifier since it is available  
214 as a MS grade solvent, being classified as a “recommended” solvent in the CHEM21 solvent selection  
215 guide.<sup>41</sup> Although acetonitrile is also available as a MS grade solvent, it is classified as a “problematic”  
216 one according to Prat et al. (2016).<sup>41</sup>

217 Considering the UHPLC-PAD/UV outcomes at 254 nm (Table 1), factors  $x_1^2$ ,  $x_2$ , and  $x_2^2$  showed a  
218 significant influence in the number of chromatographic peaks ( $y_1$ ) at 95% confidence level, with  $p$ -  
219 value of  $5.7 \times 10^{-5}$ ,  $8.1 \times 10^{-5}$ , and  $8.5 \times 10^{-3}$ , respectively. When considering the UHPLC-MS  
220 chromatograms at 95% confidence level,  $x_1^2$  and  $x_2$  were significant factors, with  $p$ -value of  $2.7 \times 10^{-2}$   
221 and  $5.2 \times 10^{-3}$  to ESI+ ( $y_2$ ), and  $3.9 \times 10^{-2}$  and  $1.7 \times 10^{-3}$  to ESI- ( $y_3$ ), respectively. The resulting  
222 mathematical models from UHPLC-PAD/UV at 254 nm, UHPLC-MS in positive and negative mode  
223 outcomes (Table 1) are presented in Equations 1, 2, and 3, respectively.

$$224 \quad y_1 = 40.60 - 4.74 x_1^2 - 4.22 x_2 - 2.24 x_2^2 \quad (\text{Equation 1})$$

$$225 \quad y_2 = 74.48 - 1.90 x_1^2 - 2.47 x_2 \quad (\text{Equation 2})$$

$$226 \quad y_3 = 37.52 - 2.60 x_1^2 - 2.94 x_2 \quad (\text{Equation 3})$$

227 The determination coefficients ( $R^2$ ) of the models described by Equations 1, 2, and 3 were 0.92,  
228 0.66, and 0.77, respectively. The statistical significance of the models was verified by the ANOVA F  
229 test. The regressions were significant ( $p < 0.05$ ) without lack of fit; therefore, the experimental data  
230 were successfully described by the models. Figure S1 illustrates the response surface of the model for  
231 the response  $y_1$ , which predicts the highest number of peaks (43) when using 15% of MeOH and 0.16–  
232 0.30% of HCOOH. Thus, the lowest concentration of HCOOH in the optimum range (0.16%) was  
233 preferred to reduce the amount of HCOOH used, in line with principle 7 of GAC.<sup>35</sup>. Furthermore, the  
234 predicted optimal conditions for the negative and positive modes of MS ( $y_2$  and  $y_3$ , respectively) were  
235 15% of MeOH and 0.10% of HCOOH, which would lead to 42 and 78 peaks, respectively. Both  
236 optimal predicted conditions were tested experimentally in triplicate, resulting in  $40 \pm 3$ ,  $42 \pm 0$ , and  
237  $79 \pm 1$  for UHPLC-PAD/UV at 254 nm, and UHPLC-MS in negative and positive modes, respectively.  
238 All these values matched the predicted ones, evidencing that the optimizations for both detectors were  
239 achieved. Representative chromatograms of the optimized conditions are shown in Figure 1, with clear  
240 distributions of peaks across the whole chromatogram.

241

242 *Optimization of the GC-MS method*

243 Another two-factor CCRD was applied to optimize the GC-MS separation for the analysis of *n*-  
244 heptane extracts obtained by 2-Mac. The experiments of the CCRD and their outcomes are presented  
245 in Table 2.

246 At 95% confidence level, factors  $x_1$ ,  $x_1^2$ ,  $x_2^2$ , and their interaction ( $x_1x_2$ ) showed significant influence  
247 in the number of peaks, with *p*-value of  $3.5 \times 10^{-3}$ ,  $7.9 \times 10^{-3}$ ,  $7.9 \times 10^{-3}$ , and  $2.4 \times 10^{-2}$ , respectively.  
248 Equation 4 is the resulting mathematical model.

249  $y_4 = 63.60 + 5.74 x_1 - 5.30 x_1^2 - 5.30 x_2^2 + 5.50 x_1 x_2$  (Equation 4)

250 The  $R^2$  of the model described by Equation 2 was 0.85. The statistical significance of the model  
251 was verified by the ANOVA F test and evaluation of *p*-values. The regression was significant  
252 (*p*=0.002) without lack of fit, indicating that the models describe well the experimental data. Figure  
253 S2 illustrates the response surface of the model for  $y_4$ , which predicted the highest number of peaks  
254 (66) when using 7 °C/min and 1.6 mL/min. Thus, the optimum predicted condition was tested in  
255 triplicate, resulting in 66±3 peaks, matching the predicted value and evidencing again that  
256 optimization was achieved. A representative chromatogram with the optimum condition is provided  
257 in Figure 2.

258  
259 *Optimization of a two-liquid-phase dynamic maceration (2-Mac)*

260 A Doehlert design was applied to optimize a 2-Mac. The experimental conditions and the  
261 corresponding responses are listed in Table 3. The monitored response ( $y_5$ ) was the sum of the peaks  
262 of the EtOH-H<sub>2</sub>O 7:3 (v/v) and *n*-heptane extracts as determined with the previously optimized  
263 UHPLC-PDA/UV-MS and GC-MS methods, respectively.<sup>19</sup>

264 At 95% confidence level, experimental factors  $x_1$ ,  $x_2$ , and  $x_4$  exhibit a significant influence in the  
265 total number of peaks, with  $p$ -value of  $2.5 \times 10^{-3}$ ,  $3.1 \times 10^{-3}$ , and  $1.1 \times 10^{-6}$ , respectively, leading to the  
266 mathematical model shown in Equation 5:

$$267 \quad y_5 = 241.88 + 19.80 x_1 + 22.20 x_2 + 35.55 x_4 \quad (\text{Equation 5})$$

268 The  $R^2$  was 0.77, and the statistical significance of the model was confirmed by the ANOVA F test  
269 and evaluation of  $p$ -values. The regression was significant ( $p=8.2 \times 10^{-7}$ ) without lack of fit, indicating  
270 that the model satisfactorily describes the experimental data. Figure S3 illustrates the response surface  
271 of the model for  $y_5$ . The coefficients of the variables were all positive; therefore, the highest response  
272 would be achieved by setting the levels of the variables  $x_1$ ,  $x_2$ , and  $x_4$  to their highest, which corresponds  
273 to 35 min of extraction time, 45 °C, and particle sizes of 48–115 Mesh. The rotation of the magnetic  
274 bar ( $x_3$ ) was fixed at 1400 rpm. The predicted optimum condition would lead to 270 peaks. Then, this  
275 optimal condition and the best experimental condition of the original CCRD itself (exp. 5 of Table 3)  
276 were performed in triplicate on the same day, leading to  $275 \pm 6$  and  $270 \pm 13$ , respectively. This result  
277 is compatible with the optimum predicted response, indicating that another optimization was achieved  
278 successfully. Chromatograms of UHPLC-PDA/UV-MS and GC-MS obtained for the EtOH-H<sub>2</sub>O 7:3  
279 (v/v) and *n*-heptane extracts produced from the optimum extraction condition are presented in Figures  
280 1 and 2, respectively.

### 281 *Applying the optimized conditions for metabolite profiling of the soy by-products*

282 Once the separation and extraction methods were optimized, the next step was to apply the  
283 optimal separation conditions for the first untargeted metabolite profiling of soy by-products from  
284 collected directly from the ground after mechanical harvesting. Due to the fact that high-resolution  
285 mass spectrometers were used in this step, adjustments in some parameters were necessary to get to  
286 the best identification conditions, as mentioned in UHPLC-QToF-MS/MS and GC-QToF-MS  
287 methods. Two NA5909 cultivars were sampled since this cultivar has been widely cropped in Brazil,  
288 which is the major producer of soybean. The outcomes confirmed that a comprehensive and optimized

289 procedure was achieved. Considering the annotation of compounds in the simultaneously produced  
290 EtOH-H<sub>2</sub>O 7:3 (v/v) and n-heptane extracts, a total of 76, 76, 66, and 79 compounds were tentatively  
291 identified for branches, leaves, pods, and roots, respectively. These were 50 in the commodity  
292 soybeans (Tables 4 and 5). The tentatively annotated compounds were grouped according to  
293 Classyfire<sup>33</sup> and summarized in Figure 3. Considering all matrices together, fatty acyls was the most  
294 representative group, with 30 metabolites. Fatty acyls are formed by the chain-elongation of an acetyl-  
295 CoA primer with malonyl-CoA or methylmalonyl-CoA and contain a vast functionality as biological  
296 lipids.<sup>42</sup> Flavonoids and isoflavonoids were the second and third most represented groups, with 16 and  
297 15 metabolites, respectively (Figure 3, Table 4). Both have been widely known activities in plant  
298 defense, in addition to human health benefits.<sup>43,44</sup> Then, prenol lipids (14), steroids and steroid  
299 derivatives (10), and carboxylic acids and derivatives (8) appeared (Figure 3, Tables 4 and 5). These  
300 correspond to fundamental substances in plants, being the precursor of several specialized metabolites  
301 and constituting vitamins, mono-, sesqui-, and triterpenes, and other high value-added compounds.<sup>42</sup>  
302 The abundance of such metabolites in soy by-products increase their potential as a source of bioactive  
303 compounds, especially for the food and pharmaceutical industry. A more detailed analysis by  
304 extract/matrix follows.

305

#### 306 *Annotation of compounds in the EtOH-H<sub>2</sub>O 7:3 (v/v) extracts by UHPLC- ESI-QToF-MS/MS*

307 The EtOH-H<sub>2</sub>O 7:3 (v/v) extracts of the individual by-products as well of soybeans were analyzed  
308 by UHPLC-ESI-QToF-MS. The annotation of the ions was performed comparing the MS/MS data  
309 against the GNPS spectral reference library.<sup>29</sup> Additionally, the candidates obtained by GNPS had their  
310 compatibility challenged with the acquired high-resolution masses as well as with previous reports of  
311 their occurrence in soy by-products as organized in a database published elsewhere.<sup>14</sup>

312 A total of 67 metabolites were tentatively identified in EtOH-H<sub>2</sub>O 7:3 (v/v) extracts of soy parts as  
313 summarized in Table 4. Sixteen of them were isoflavonoids, a class of compounds that have been

314 widely known for their bioactive properties. These were 11, 10, 10, 9, and 4 in pods, branches,  
315 soybeans, roots, and leaves, respectively (Table 4). Formononetin 7-O-glucoside (ononin) and  
316 genistein, which are isoflavonoids, were identified in all soy by-products (Table 4). Ononin has been  
317 reported as a potential anti-angiogenesis and neuroprotective effects, while genistein administration  
318 has been related to the reduction of some cardiovascular risk factors.<sup>45-47</sup> Additionally, a systematic  
319 review about phytoestrogens and menopausal symptoms pointed to the potential of extracts with high  
320 levels of genistein (> 30 mg/d) in the treatment for hot flushes.<sup>48</sup> Flavonoids corresponded to 16  
321 compounds, with 11, 9, 8, 6, and 2 identified in pods, branches, soybeans, leaves, and roots,  
322 respectively, with four of them only identified in soybeans (Table 4). Flavonoid glycosides, such as  
323 kaempferol 3-O-glucoside (astragalín), kaempferol 3-O-sophoroside (sophoraflavonolósido), and  
324 luteolin 7-O-glucoside (cynaroside), were identified in at least one soy by-product. Among them,  
325 sophoraflavonolósido was only present in soybeans (Table 4). These compounds have shown  
326 antitumor and anti-inflammatory properties.<sup>49-51</sup>

327 It is important to highlight that compounds identified in EtOH-H<sub>2</sub>O 7:3 (v/v) extracts of soy by-  
328 products, such as daidzein, daidzin, genistein, genistin, and glycitin are classified as phytotoxins.<sup>6</sup>  
329 Zoccolo (2010)<sup>73</sup> and later Hama et al. (2021)<sup>6</sup> reported the presence of such compounds in agricultural  
330 soils and streams close to soy crops, which in turn were close to human settlements. Thus, recovering  
331 such isoflavonoids from at least part of soy by-products would be an advantage from an environmental  
332 point of view to decrease the potential undesirable effects associated with large build-up of  
333 isoflavonoids in the environment.<sup>17</sup> This would be in line with a “responsible production” of soybeans,  
334 as recommended by the number 12 of the 17 Sustainable Development Goals of the UN. Additionally,  
335 this practice might potentially lead to economic gains in soy production chain by developing value-  
336 added commercial extracts standardized in bioactive isoflavones from by-products in a biorefinery  
337 approach.

338

340 The *n*-heptane extracts of the individual by-products as well of soybeans were analyzed by GC-  
341 QToF-MS (Figure S4). MS data were matched against NIST and Fiehn libraries. High match factors  
342 (>70) were achieved between the MS spectrum of the candidates and of these libraries, as demonstrated  
343 in Table 5. A total of 79 metabolites were tentatively identified in *n*-heptane extracts of soy by-products  
344 or soybeans, as summarized in Table 5. Most of them (50) were identified for the first time in soy plant  
345 tissues, which could be because few works in the literature have used nonpolar solvents for the  
346 extraction and subsequent analyses of soy parts or by-products. Nonpolar solvents, such as hexane,  
347 have been used mainly to degrease plant matrices by solid-liquid extractions or liquid-liquid  
348 partitioning in a so called “clean-up” sample pre-treatment.<sup>74,75</sup> The produced nonpolar extract or  
349 fraction are often discarded without analysis of their composition, potentially losing valuable  
350 phytochemical information<sup>19,20</sup> as confirmed in this work. Twenty-six metabolites tentatively  
351 identified in *n*-heptane extracts were fatty acyls, which is the group with the highest number of  
352 annotated compounds (Table 5). Most of these compounds was identified in soy roots (18), followed  
353 by leaves (13), branches (11), pods (8), and soybeans (6). Fatty acyls, such as hexadecenoic,  
354 octadecanoic, and eicosanoic acids, have been used in dermatologic, nutraceutical, and other industries  
355 since they present anti-inflammatory properties and antibacterial activities.<sup>76,77</sup> The second most  
356 representative group of compounds present in *n*-heptane extracts was steroids and their derivatives,  
357 with 14 annotated compounds. Particularly,  $\alpha$ -tocopherol, campesterol, and stigmasterol were  
358 identified in all soy by-products.  $\alpha$ -Tocopherol has been used in nutraceutical and dermatological  
359 applications, as it presents antioxidant, anti-inflammatory, and other human health benefits.<sup>78,79</sup> On the  
360 other hand, campesterol and stigmasterol have been explored for potential treatment of human ovarian  
361 cancer and neuroprotective properties, respectively.<sup>80,81</sup> Moreover, prenol lipids, phytosteroids,  
362 unsaturated, and saturated hydrocarbons were identified in soy by-products, representing a diverse  
363 range of compounds in such materials.



364

## 365 **ABBREVIATIONS USED**

366 United Nations (UN), Design of Experiment (DoE), Ultra-High Performance Liquid Chromatography-  
367 PDA/UV-Mass Spectrometry (UHPLC-PDA/UV-MS), Global Natural Product Social Molecular  
368 Networking (GNPS), Gas Chromatography- Mass Spectrometry, NIST (National Institute of  
369 Standards and Technology), HPLC (High Performance Liquid Chromatography), P.A. (practical  
370 grade), Two-Liquid-Phase Dynamic Maceration (2-Mac), Ethanol (EtOH), Methanol (MeOH), Formic  
371 Acid (HCOOH), Electrospray Ionization (ESI), Central Composite Rotatable Design (CCDR), Ultra-  
372 High-Performance Liquid Chromatography coupled to Time-Of-Flight Mass Spectrometry (UHPLC-  
373 QToF-MS/MS), Gas Chromatography coupled to Time-of-Flight Mass Spectrometry (GC-QToF-MS).

## 374 **SUPPORTING INFORMATION DESCRIPTION**

375 Response surfaces and contour plots of all optimized methods (Figures S1-S3) and GC-QToF-MS  
376 chromatogram of the individual soy by-products (Figure S4).

377

## 378 **REFERENCES**

- 379 (1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four  
380 Decades from 01/1981 to 09/2019. *Journal of Natural Products*. 2020, pp 770–803.  
381 <https://doi.org/10.1021/acs.jnatprod.9b01285>.
- 382 (2) Ranaivoson, L.; Naudin, K.; Ripoche, A.; Affholder, F.; Rabeharisoa, L.; Corbeels, M. Agro-  
383 Ecological Functions of Crop Residues under Conservation Agriculture. A Review. *Agronomy*  
384 *for Sustainable Development* **2017**, 37 (4). <https://doi.org/10.1007/s13593-017-0432-z>.
- 385 (3) Rakkar, M. K.; Blanco-Canqui, H. Grazing of Crop Residues: Impacts on Soils and Crop  
386 Production. *Agriculture, Ecosystems and Environment*. 2018, pp 71–90.  
387 <https://doi.org/10.1016/j.agee.2017.11.018>.
- 388 (4) Reinoso Moreno, J. v.; Pinna-Hernández, G.; Fernández Fernández, M. D.; Sánchez Molina, J.  
389 A.; Rodríguez Díaz, F.; López Hernández, J. C.; Acién Fernández, F. G. Optimal Processing

- 390 of Greenhouse Crop Residues to Use as Energy and CO<sub>2</sub> Sources. *Industrial Crops and*  
391 *Products* **2019**, *137*, 662–671. <https://doi.org/10.1016/J.INDCROP.2019.04.074>.
- 392 (5) Carvalho, J. L. N.; Nogueirol, R. C.; Menandro, L. M. S.; Bordonal, R. de O.; Borges, C. D.;  
393 Cantarella, H.; Franco, H. C. J. Agronomic and Environmental Implications of Sugarcane  
394 Straw Removal: A Major Review. *GCB Bioenergy* **2017**, *9* (7), 1181–1195.  
395 <https://doi.org/10.1111/GCBB.12410>.
- 396 (6) Hama, J. R.; Kolpin, D. W.; LeFevre, G. H.; Hubbard, L. E.; Powers, M. M.; Strobel, B. W.  
397 Exposure and Transport of Alkaloids and Phytoestrogens from Soybeans to Agricultural Soils  
398 and Streams in the Midwestern United States. *Environmental Science & Technology* **2021**,  
399 [acs.est.1c01477](https://doi.org/10.1021/ACS.EST.1C01477). <https://doi.org/10.1021/ACS.EST.1C01477>.
- 400 (7) Mateo-Sagasta, J.; Zadeh, S. M.; Turrall, H. *Water Pollution from Agriculture: A Global*  
401 *Review*; 2017. <https://www.fao.org/3/i7754e/i7754e.pdf> (accessed 2021-02-02).
- 402 (8) ONU. *World Population Prospects 2019*; 2019. <https://population.un.org/wpp/> (accessed  
403 2021-02-02).
- 404 (9) United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*;  
405 2018. <https://doi.org/10.1891/9780826190123.ap02>.
- 406 (10) European Environment Agency. *Towards a Green Economy in Europe - EU Environmental*  
407 *Policy Targets and Objectives 2010-2050 — European Environment Agency*; 2013.
- 408 (11) Food and Agriculture Organization of the United Nations. *FAOSTAT*. Crops - Production  
409 Quantity. <http://www.fao.org/faostat/en/#data/QC> (accessed 2021-02-02).
- 410 (12) Krisnawati, A.; Adie, M. M. Variability of Biomass and Harvest Index from Several Soybean  
411 Genotypes as Renewable Energy Source. In *Energy Procedia*; Elsevier Ltd, 2015; Vol. 65, pp  
412 14–21. <https://doi.org/10.1016/j.egypro.2015.01.023>.
- 413 (13) United States Department of Agriculture. *Oilseeds: World Markets and Trade*; 2021.  
414 [https://downloads.usda.library.cornell.edu/usda-](https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/2v23wp329/6t054b55w/oilseeds.pdf)  
415 [esmis/files/tx31qh68h/2v23wp329/6t054b55w/oilseeds.pdf](https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/2v23wp329/6t054b55w/oilseeds.pdf) (accessed 2021-05-03).
- 416 (14) Bragagnolo, F. S.; Funari, C. S.; Ibáñez, E.; Cifuentes, A. Metabolomics as a Tool to Study  
417 Underused Soy Parts: In Search of Bioactive Compounds. *Foods* **2021**, *Vol. 10*, Page 1308  
418 **2021**, *10* (6), 1308. <https://doi.org/10.3390/FOODS10061308>.
- 419 (15) Feng, Z.; Ding, C.; Li, W.; Wang, D.; Cui, D. Applications of Metabolomics in the Research  
420 of Soybean Plant under Abiotic Stress. *Food Chemistry* **2020**, *310*, 125914.  
421 <https://doi.org/10.1016/J.FOODCHEM.2019.125914>.
- 422 (16) Silva, E.; Belinato, J. R.; Porto, C.; Nunes, E.; Guimarães, F.; Meyer, M. C.; Pilau, E. J.  
423 Soybean Metabolomics Based in Mass Spectrometry: Decoding the Plant's Signaling and  
424 Defense Responses under Biotic Stress. *Journal of Agricultural and Food Chemistry* **2021**, *69*  
425 (26), 7257–7267. <https://doi.org/10.1021/ACS.JAFC.0C07758>.
- 426 (17) Carneiro, A. M.; Moreira, E. A.; Bragagnolo, F. S.; Borges, M. S.; Pilon, A. C.; Rinaldo, D.;  
427 Funari, C. S. Soya Agricultural Waste as a Rich Source of Isoflavones. *Food Research*  
428 *International* **2020**, *130*. <https://doi.org/10.1016/j.foodres.2019.108949>.

- 429 (18) Hama, J. R.; Kolpin, D. W.; LeFevre, G. H.; Hubbard, L. E.; Powers, M. M.; Strobel, B. W.  
430 Exposure and Transport of Alkaloids and Phytoestrogens from Soybeans to Agricultural Soils  
431 and Streams in the Midwestern United States. *Environmental Science & Technology* **2021**, *55*  
432 (16), 11029–11039. <https://doi.org/10.1021/ACS.EST.1C01477>.
- 433 (19) Assirati, J.; Rinaldo, D.; Rabelo, S. C.; Bolzani, V. da S.; Hilder, E. F.; Funari, C. S. A Green,  
434 Simplified, and Efficient Experimental Setup for a High-Throughput Screening of Agri-Food  
435 by-Products – From Polar to Nonpolar Metabolites in Sugarcane Solid Residues. *Journal of*  
436 *Chromatography A* **2020**, *1634*. <https://doi.org/10.1016/j.chroma.2020.461693>.
- 437 (20) da Silva, M. R.; Bragagnolo, F. S.; Carneiro, R. L.; de Oliveira Carvalho Pereira, I.; Ribeiro, J.  
438 A. A.; Martins, C.; Jelley, R. E.; Fedrizzi, B.; Funari, C. S. Metabolite Characterization of  
439 Fifteen By-Products of the Coffee Production Chain: From Farm to Factory. *Food Chemistry*  
440 **2021**, 130753. <https://doi.org/10.1016/J.FOODCHEM.2021.130753>.
- 441 (21) Rodrigues, M. I.; Iemma, A. F. *Experimental Design and Process Optimization*; Crc Press,  
442 2014. <https://doi.org/10.1201/b17848>.
- 443 (22) Zimmerman, J. B.; Anastas, P. T.; Erythropel, H. C.; Leitner, W. Designing for a Green  
444 Chemistry Future. *Science (1979)* **2020**, *400* (January), 397–400.
- 445 (23) Lee, E. M.; Park, S. J.; Lee, J. E.; Lee, B. M.; Shin, B. K.; Kang, D. J.; Choi, H. K.; Kim, Y.  
446 S.; Lee, D. Y. Highly Geographical Specificity of Metabolomic Traits among Korean  
447 Domestic Soybeans (*Glycine Max*). *Food Research International* **2019**, *120*, 12–18.  
448 <https://doi.org/10.1016/j.foodres.2019.02.021>.
- 449 (24) Lu, Y.; Lam, H.; Pi, E.; Zhan, Q.; Tsai, S.; Wang, C.; Kwan, Y.; Ngai, S. Comparative  
450 Metabolomics in *Glycine Max* and *Glycine Soja* under Salt Stress to Reveal the Phenotypes of  
451 Their Offspring. *Journal of Agricultural and Food Chemistry* **2013**, *61* (36), 8711–8721.  
452 <https://doi.org/10.1021/jf402043m>.
- 453 (25) Song, H. H.; Ryu, H. W.; Lee, K. J.; Jeong, I. Y.; Kim, D. S.; Oh, S. R. Metabolomics  
454 Investigation of Flavonoid Synthesis in Soybean Leaves Depending on the Growth Stage.  
455 *Metabolomics* **2014**, *10* (5), 833–841. <https://doi.org/10.1007/s11306-014-0640-3>.
- 456 (26) Sisa, M.; Bonnet, S. L.; Ferreira, D.; van der Westhuizen, J. H. Photochemistry of Flavonoids.  
457 *Molecules* **2010**, *Vol. 15*, Pages 5196-5245 **2010**, *15* (8), 5196–5245.  
458 <https://doi.org/10.3390/MOLECULES15085196>.
- 459 (27) Pérez-Serradilla, J. A.; Japón-Luján, R.; Luque de Castro, M. D. Simultaneous Microwave-  
460 Assisted Solid-Liquid Extraction of Polar and Nonpolar Compounds from Alperujo. *Analytica*  
461 *Chimica Acta* **2007**, *602* (1), 82–88. <https://doi.org/10.1016/j.aca.2007.09.008>.
- 462 (28) Ballesteros-Vivas, D.; Álvarez-Rivera, G.; Morantes, S. J.; Sánchez-Camargo, A. del P.;  
463 Ibáñez, E.; Parada-Alfonso, F.; Cifuentes, A. An Integrated Approach for the Valorization of  
464 Mango Seed Kernel: Efficient Extraction Solvent Selection, Phytochemical Profiling and  
465 Antiproliferative Activity Assessment. *Food Research International* **2019**, *126*, 108616.  
466 <https://doi.org/10.1016/J.FOODRES.2019.108616>.
- 467 (29) Wang, M.; Carver, J. J.; Phelan, V. v.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.;  
468 Watrous, J.; Kapon, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. v.;

- 469 Meehan, M. J.; Liu, W. T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-  
470 Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D.  
471 J.; Gavilan, R. G.; Kleigrew, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle,  
472 B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel,  
473 J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.;  
474 Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.;  
475 Larson, C. B.; Boya, C. A. P.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L.  
476 M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E.  
477 A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J.  
478 A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.;  
479 Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.;  
480 Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.;  
481 Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B.  
482 M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.;  
483 Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.;  
484 Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.;  
485 Palsson, B.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.;  
486 Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass  
487 Spectrometry Data with Global Natural Products Social Molecular Networking. *Nature*  
488 *Biotechnology*. **2016**. <https://doi.org/10.1038/nbt.3597>.
- 489 (30) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for  
490 Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data.  
491 *BMC Bioinformatics* **2010**, *11*. <https://doi.org/10.1186/1471-2105-11-395>.
- 492 (31) Katajamaa, M.; Miettinen, J.; Orešič, M. MZmine: Toolbox for Processing and Visualization  
493 of Mass Spectrometry Based Molecular Profile Data. *Bioinformatics* **2006**, *22* (5).  
494 <https://doi.org/10.1093/bioinformatics/btk039>.
- 495 (32) kim, H.; Wang, M.; Leber, C.; Nothias, L.-F.; Reher, R.; Kang, K. bin; Hooft, J. J. J. van der;  
496 Dorrestein, P.; Gerwick, W.; Cottrell, G. NPClassifier: A Deep Neural Network-Based  
497 Structural Classification Tool for Natural Products. **2020**.  
498 <https://doi.org/10.26434/CHEMRXIV.12885494.V1>.
- 499 (33) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.;  
500 Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated  
501 Chemical Classification with a Comprehensive, Computable Taxonomy. *Journal of*  
502 *Cheminformatics* **2016**, *8* (1), 1–20. <https://doi.org/10.1186/s13321-016-0174-y>.
- 503 (34) Leme, G. M.; Coutinho, I. D.; Creste, S.; Hojo, O.; Carneiro, R. L.; Bolzani, V. D. S.;  
504 Cavalheiro, A. J. HPLC-DAD Method for Metabolic Fingerprinting of the Phenotyping of  
505 Sugarcane Genotypes. *Analytical Methods* **2014**, *6* (19), 7781–7788.  
506 <https://doi.org/10.1039/C4AY01750A>.
- 507 (35) Gałuszka, A.; Migaszewski, Z.; Namieśnik, J. The 12 Principles of Green Analytical  
508 Chemistry and the SIGNIFICANCE Mnemonic of Green Analytical Practices. *TrAC - Trends*  
509 *in Analytical Chemistry*. **2013**. <https://doi.org/10.1016/j.trac.2013.04.010>.

- 510 (36) Yun, D. Y.; Kang, Y. G.; Kim, E. H.; Kim, M.; Park, N. H.; Choi, H. T.; Go, G. H.; Lee, J. H.;  
511 Park, J. S.; Hong, Y. S. Metabolomics Approach for Understanding Geographical Dependence  
512 of Soybean Leaf Metabolome. *Food Research International* **2018**.  
513 <https://doi.org/10.1016/j.foodres.2018.01.061>.
- 514 (37) Liigand, J.; Laaniste, A.; Kruve, A. PH Effects on Electrospray Ionization Efficiency. *J Am*  
515 *Soc Mass Spectrom* **2017**, *28* (3), 461–469. <https://doi.org/10.1007/s13361-016-1563-1>.
- 516 (38) Wu, Z.; Gao, W.; Phelps, M. A.; Wu, D.; Miller, D. D.; Dalton, J. T. Favorable Effects of  
517 Weak Acids on Negative-Ion Electrospray Ionization Mass Spectrometry. *Analytical*  
518 *Chemistry* **2004**, *76* (3), 839–847. <https://doi.org/10.1021/ac0351670>.
- 519 (39) Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A.  
520 Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte.  
521 *Analytical Chemistry* **2017**, *89* (11), 5665–5668.  
522 <https://doi.org/10.1021/acs.analchem.7b00096>.
- 523 (40) Snyder, L. R.; Dolan, J. W. Optimizing Selectivity during Reversed-Phase High Performance  
524 Liquid Chromatography Method Development: Prioritizing Experimental Conditions. *Journal*  
525 *of Chromatography A* **2013**, *1302*, 45–54. <https://doi.org/10.1016/j.chroma.2013.05.082>.
- 526 (41) Prat, D.; Wells, A.; Hayler, J.; Sneddon, H.; McElroy, C. R.; Abou-Shehada, S.; Dunn, P. J.  
527 CHEM21 Selection Guide of Classical- and Less Classical-Solvents. *Green Chemistry* **2016**,  
528 *18* (1), 288–296. <https://doi.org/10.1039/c5gc01008j>.
- 529 (42) Fahy, E.; Subramaniam, S.; Brown, H. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz,  
530 C. R. H.; Russell, D. W.; Seyama, Y.; Shaw, W.; Shimizu, T.; Spener, F.; van Meer, G.;  
531 VanNieuwenhze, M. S.; White, S. H.; Witztum, J. L.; Dennis, E. A. A Comprehensive  
532 Classification System for Lipids. *Journal of Lipid Research* **2005**, *46* (5), 839–861.  
533 <https://doi.org/10.1194/JLR.E400004-JLR200>.
- 534 (43) Wang, J. F.; Liu, S. S.; Song, Z. Q.; Xu, T. C.; Liu, C. S.; Hou, Y. G.; Huang, R.; Wu, S. H.  
535 Naturally Occurring Flavonoids and Isoflavonoids and Their Microbial Transformation: A  
536 Review. *Molecules* **2020**, *Vol. 25*, Page 5112 **2020**, *25* (21), 5112.  
537 <https://doi.org/10.3390/MOLECULES25215112>.
- 538 (44) Sajid, M.; Stone, S. R.; Kaur, P. Recent Advances in Heterologous Synthesis Paving Way for  
539 Future Green-Modular Bioindustries: A Review With Special Reference to Isoflavonoids.  
540 *Frontiers in Bioengineering and Biotechnology* **2021**, *9*, 532.  
541 <https://doi.org/10.3389/FBIOE.2021.673270/BIBTEX>.
- 542 (45) Gong, G.; Zheng, Y.; Kong, X.; Wen, Z. Anti-Angiogenesis Function of Ononin via  
543 Suppressing the MEK/Erk Signaling Pathway. *Journal of Natural Products* **2021**.  
544 <https://doi.org/10.1021/ACS.JNATPROD.1C00008>.
- 545 (46) Chen, X.; Zhang, M.; Ahmed, M.; Surapaneni, K. M.; Veeraraghavan, V. P.; Arulselvan, P.  
546 Neuroprotective Effects of Ononin against the Aluminium Chloride-Induced Alzheimer's  
547 Disease in Rats. *Saudi Journal of Biological Sciences* **2021**, *28* (8), 4232–4239.  
548 <https://doi.org/10.1016/J.SJBS.2021.06.031>.

- 549 (47) Amerizadeh, A.; Asgary, S.; Vaseghi, G.; Farajzadegan, Z. Effect of Genistein Intake on  
550 Some Cardiovascular Risk Factors: An Updated Systematic Review and Meta-Analysis.  
551 *Current Problems in Cardiology* **2021**, 100902.  
552 <https://doi.org/10.1016/J.CPCARDIOL.2021.100902>.
- 553 (48) Lethaby, A.; Marjoribanks, J.; Kronenberg, F.; Roberts, H.; Eden, J.; Brown, J.  
554 Phytoestrogens for Menopausal Vasomotor Symptoms. *Cochrane Database of Systematic*  
555 *Reviews* **2013**, 2013 (12).  
556 [https://doi.org/10.1002/14651858.CD001395.PUB4/MEDIA/CDSR/CD001395/IMAGE\\_N/N](https://doi.org/10.1002/14651858.CD001395.PUB4/MEDIA/CDSR/CD001395/IMAGE_N/N)  
557 [CD001395-CMP-001-07.PNG](https://doi.org/10.1002/14651858.CD001395.PUB4/MEDIA/CDSR/CD001395/IMAGE_N/N).
- 558 (49) Yang, M.; Li, W.-Y.; Xie, J.; Wang, Z.-L.; Wen, Y.-L.; Zhao, C.-C.; Tao, L.; Li, L.-F.; Tian,  
559 Y.; Sheng, J. Astragalosin Inhibits the Proliferation and Migration of Human Colon Cancer  
560 HCT116 Cells by Regulating the NF- $\kappa$ B Signaling Pathway. *Frontiers in Pharmacology*  
561 **2021**, 0, 467. <https://doi.org/10.3389/FPHAR.2021.639256>.
- 562 (50) Kim, T. H.; Ku, S. K.; Bae, J. S. Inhibitory Effects of Kaempferol-3-O-Sophoroside on  
563 HMGB1-Mediated Proinflammatory Responses. *Food and Chemical Toxicology* **2012**, 50 (3–  
564 4), 1118–1123. <https://doi.org/10.1016/J.FCT.2011.12.004>.
- 565 (51) Szekalska, M.; Sosnowska, K.; Tomczykowa, M.; Winnicka, K.; Kasacka, I.; Tomczyk, M. In  
566 Vivo Anti-Inflammatory and Anti-Allergic Activities of Cynaroside Evaluated by Using  
567 Hydrogel Formulations. *Biomedicine & Pharmacotherapy* **2020**, 121, 109681.  
568 <https://doi.org/10.1016/J.BIOPHA.2019.109681>.
- 569 (52) Zhong, Y.; Zhao, Y. Chemical Composition and Functional Properties of Three Soy  
570 Processing By-Products (Soy Hull, Okara and Molasses). *Quality Assurance and Safety of*  
571 *Crops and Foods* **2015**, 7 (5), 651–660. <https://doi.org/10.3920/QAS2014.0481>.
- 572 (53) Nam, K. H.; Kim, D. Y.; Kim, H. J.; Park, I. S.; Kim, H. J.; Chung, Y. S.; Kim, S. Y.; Kim, C.  
573 G. Global Metabolite Profiling Based on GC–MS and LC–MS/MS Analyses in ABF3-  
574 Overexpressing Soybean with Enhanced Drought Tolerance. *Applied Biological Chemistry*  
575 **2019**, 62 (1), 15. <https://doi.org/10.1186/s13765-019-0425-5>.
- 576 (54) Silva, E.; da Graça, J. P.; Porto, C.; Martin do Prado, R.; Hoffmann-Campo, C. B.; Meyer, M.  
577 C.; de Oliveira Nunes, E.; Pilau, E. J. Unraveling Asian Soybean Rust Metabolomics Using  
578 Mass Spectrometry and Molecular Networking Approach. *Scientific Reports* **2020**, 10 (1).  
579 <https://doi.org/10.1038/s41598-019-56782-4>.
- 580 (55) Li, M.; Xu, J.; Guo, R.; Liu, Y.; Wang, S.; Wang, H.; Ullah, A.; Shi, L. Identifying the  
581 Metabolomics and Physiological Differences among Soja in the Early Flowering Stage. *Plant*  
582 *Physiology and Biochemistry* **2019**, 139, 82–91. <https://doi.org/10.1016/j.plaphy.2019.03.012>.
- 583 (56) Li, M.; Xu, J.; Wang, X.; Fu, H.; Zhao, M.; Wang, H.; Shi, L. Photosynthetic Characteristics  
584 and Metabolic Analyses of Two Soybean Genotypes Revealed Adaptive Strategies to Low-  
585 Nitrogen Stress. *Journal of Plant Physiology* **2018**, 229, 132–141.  
586 <https://doi.org/10.1016/j.jplph.2018.07.009>.
- 587 (57) Silva, E.; Perez Da Graça, J.; Porto, C.; Martin Do Prado, R.; Nunes, E.; Corrêa Marcelino-  
588 Guimarães, F.; Conrado Meyer, M.; Jorge Pilau, E. Untargeted Metabolomics Analysis by

- 589 UHPLC-MS/MS of Soybean Plant in a Compatible Response to Phakopsora Pachyrhizi  
590 Infection. *Metabolites* **2021**, *11* (3), 179. <https://doi.org/10.3390/metabo11030179>.
- 591 (58) Shibata, M.; Hirotsuka, M.; Mizutani, Y.; Takahashi, H.; Kawada, T.; Matsumiya, K.;  
592 Hayashi, Y.; Matsumura, Y. Diversity of  $\gamma$ - Glutamyl Peptides and Oligosaccharides, the  
593 “Kokumi” Taste Enhancers, in Seeds from Soybean Mini Core Collections. *Bioscience*,  
594 *Biotechnology, and Biochemistry* **2018**, *82* (3), 507–514.  
595 <https://doi.org/10.1080/09168451.2018.1436960>.
- 596 (59) Zhong, Z.; Kobayashi, T.; Zhu, W.; Imai, H.; Zhao, R.; Ohno, T.; Rehman, S. ur; Uemura, M.;  
597 Tian, J.; Komatsu, S. Plant-Derived Smoke Enhances Plant Growth through Ornithine-  
598 Synthesis Pathway and Ubiquitin-Proteasome Pathway in Soybean. *Journal of Proteomics*  
599 **2020**, *221*. <https://doi.org/10.1016/j.jprot.2020.103781>.
- 600 (60) Jiang, Z. feng; Liu, D. dan; Wang, T. qiong; Liang, X. long; Cui, Y. hai; Liu, Z. hua; Li, W.  
601 bin. Concentration Difference of Auxin Involved in Stem Development in Soybean. *Journal*  
602 *of Integrative Agriculture* **2020**, *19* (4), 953–964. <https://doi.org/10.1016/S2095->  
603 [3119\(19\)62676-6](https://doi.org/10.1016/S2095-3119(19)62676-6).
- 604 (61) Zanzarin, D. M.; Hernandez, C. P.; Leme, L. M.; Silva, E.; Porto, C.; Martin do Prado, R.;  
605 Meyer, M. C.; Favoreto, L.; Nunes, E. de O.; Pilau, E. J. Metabolomics of Soybean Green  
606 Stem and Foliar Retention (GSFR) Disease Using Mass Spectrometry and Molecular  
607 Networking. *Rapid Communications in Mass Spectrometry* **2020**, *34* (S3).  
608 <https://doi.org/10.1002/rcm.8655>.
- 609 (62) Ban, Y. J.; Song, Y. H.; Kim, J. Y.; Baiseitova, A.; Lee, K. W.; Kim, K. D.; Park, K. H.  
610 Comparative Investigation on Metabolites Changes in Soybean Leaves by Ethylene and  
611 Activation of Collagen Synthesis. *Industrial Crops and Products* **2020**, *154*.  
612 <https://doi.org/10.1016/j.indcrop.2020.112743>.
- 613 (63) Silva, F. A. C.; O-Panizzi, M. C. C.; Blassioli-Moraes, M. C.; Panizzi, A. R. Influence of  
614 Volatile and Nonvolatile Secondary Metabolites from Soybean Pods on Feeding and on  
615 Oviposition Behavior of Euschistus Heros (Hemiptera: Heteroptera: Pentatomidae).  
616 *Environmental Entomology* **2013**, *42* (6), 1375–1382. <https://doi.org/10.1603/EN13081>.
- 617 (64) Tsuno, Y.; Fujimatsu, T.; Endo, K.; Sugiyama, A.; Yazaki, K. Soyasaponins: A New Class of  
618 Root Exudates in Soybean (Glycine Max). *Plant and Cell Physiology* **2018**, *59* (2), 366–375.  
619 <https://doi.org/10.1093/pcp/pcx192>.
- 620 (65) Dresler, S.; Wójciak-Kosior, M.; Sowa, I.; Strzemski, M.; Sawicki, J.; Kováčik, J.; Blicharski,  
621 T. Effect of Long-Term Strontium Exposure on the Content of Phytoestrogens and Allantoin  
622 in Soybean. *International Journal of Molecular Sciences* **2018**, *19* (12), 3864.  
623 <https://doi.org/10.3390/ijms19123864>.
- 624 (66) Coutinho, I. D.; Henning, L. M. M.; Döpp, S. A.; Nepomuceno, A.; Moraes, L. A. C.;  
625 Marcolino-Gomes, J.; Richter, C.; Schwalbe, H.; Colnago, L. A. Identification of Primary and  
626 Secondary Metabolites and Transcriptome Profile of Soybean Tissues during Different Stages  
627 of Hypoxia. *Data in Brief* **2018**, *21*, 1089–1100. <https://doi.org/10.1016/j.dib.2018.09.122>.
- 628 (67) Dillon, F. M.; Chludil, H. D.; Mithöfer, A.; Zavala, J. A. Solar UVB-Inducible Ethylene  
629 Alone Induced Isoflavonoids in Pods of Field-Grown Soybean, an Important Defense against

- 630 Stink Bugs. *Environmental and Experimental Botany* **2020**, *178*, 104167.  
631 <https://doi.org/10.1016/j.envexpbot.2020.104167>.
- 632 (68) Seo, W. D.; Kang, J. E.; Choi, S. W.; Lee, K. S.; Lee, M. J.; Park, K. do; Lee, J. H.  
633 Comparison of Nutritional Components (Isoflavone, Protein, Oil, and Fatty Acid) and  
634 Antioxidant Properties at the Growth Stage of Different Parts of Soybean [Glycine Max (L.)  
635 Merrill]. *Food Science and Biotechnology* **2017**, *26* (2), 339–347.  
636 <https://doi.org/10.1007/s10068-017-0046-x>.
- 637 (69) Zou, J.; Yu, H.; Yu, Q.; Jin, X.; Cao, L.; Wang, M.; Wang, M.; Ren, C.; Zhang, Y.  
638 Physiological and UPLC-MS/MS Widely Targeted Metabolites Mechanisms of Alleviation of  
639 Drought Stress-Induced Soybean Growth Inhibition by Melatonin. *Industrial Crops and*  
640 *Products* **2021**, *163*, 113323. <https://doi.org/10.1016/J.INDCROP.2021.113323>.
- 641 (70) Zhao, B.; Zhang, S.; Yang, W.; Li, B.; Lan, C.; Zhang, J.; Yuan, L.; Wang, Y.; Xie, Q.; Han,  
642 J.; Mur, L. A. J.; Hao, X.; Roberts, J. A.; Miao, Y.; Yu, K.; Zhang, X. Multi-Omic Dissection  
643 of the Drought Resistance Traits of Soybean Landrace LX. *Plant, Cell & Environment* **2021**,  
644 *44* (5), 1379–1398. <https://doi.org/10.1111/PCE.14025>.
- 645 (71) Takahashi, H.; Ochiai, K.; Sasaki, K.; Izumi, A.; Shinyama, Y.; Mohri, S.; Nomura, W.;  
646 Jheng, H.-F.; Kawada, T.; Inoue, K.; Goto, T. Metabolome Analysis Revealed That Soybean–  
647 *Aspergillus Oryzae* Interaction Induced Dynamic Metabolic and Daidzein Prenylation  
648 Changes. *PLOS ONE* **2021**, *16* (7), e0254190.  
649 <https://doi.org/10.1371/JOURNAL.PONE.0254190>.
- 650 (72) Abeysekara, N. S.; Swaminathan, S.; Desai, N.; Guo, L.; Bhattacharyya, M. K. The Plant  
651 Immunity Inducer Pipecolic Acid Accumulates in the Xylem Sap and Leaves of Soybean  
652 Seedlings Following *Fusarium Virguliforme* Infection. *Plant Science* **2016**, *243*, 105–114.  
653 <https://doi.org/10.1016/J.PLANTSCI.2015.11.008>.
- 654 (73) Zocolo, G. J. Occurrence of Soy Isoflavones in the Environment and Correlation with  
655 Estrogenic Activities: A Case Study in the Region of Dourados (MS). Ph.D. Thesis, State  
656 University of São Paulo, Araraquara, Brazil, 2010.
- 657 (74) Mushtaq, M. Y.; Choi, Y. H.; Verpoorte, R.; Wilson, E. G. Extraction for Metabolomics:  
658 Access to The Metabolome. *Phytochemical Analysis* **2014**, *25* (4), 291–306.  
659 <https://doi.org/10.1002/PCA.2505>.
- 660 (75) Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K. Metabolomics: Back to Basics.  
661 *Phytochemistry Reviews* **2008**, *7* (3), 525–537. [https://doi.org/10.1007/S11101-008-9091-](https://doi.org/10.1007/S11101-008-9091-7/TABLES/2)  
662 [7/TABLES/2](https://doi.org/10.1007/S11101-008-9091-7/TABLES/2).
- 663 (76) Radzikowska, U.; Rinaldi, A. O.; Sözener, Z. Ç.; Karaguzel, D.; Wojcik, M.; Cypryk, K.;  
664 Akdis, M.; Akdis, C. A.; Sokolowska, M. The Influence of Dietary Fatty Acids on Immune  
665 Responses. *Nutrients* **2019**, *Vol. 11, Page 2990* **2019**, *11* (12), 2990.  
666 <https://doi.org/10.3390/NU11122990>.
- 667 (77) Takigawa, H.; Nakagawa, H.; Kuzukawa, M.; Mori, H.; Imokawa, G. Deficient Production of  
668 Hexadecenoic Acid in the Skin Is Associated in Part with the Vulnerability of Atopic  
669 Dermatitis Patients to Colonization by *Staphylococcus Aureus*. *Dermatology* **2005**, *211* (3),  
670 240–248. <https://doi.org/10.1159/000087018>.



- 671 (78) Szewczyk, K.; Chojnacka, A.; Górnicka, M. Tocopherols and Tocotrienols—Bioactive  
672 Dietary Compounds; What Is Certain, What Is Doubt? *International Journal of Molecular*  
673 *Sciences* 2021, Vol. 22, Page 6222 **2021**, 22 (12), 6222.  
674 <https://doi.org/10.3390/IJMS22126222>.
- 675 (79) Vincenzo, A. di; Tana, C.; Hadi, H. el; Pagano, C.; Vettor, R.; Rossato, M. Antioxidant, Anti-  
676 Inflammatory, and Metabolic Properties of Tocopherols and Tocotrienols: Clinical  
677 Implications for Vitamin E Supplementation in Diabetic Kidney Disease. *International*  
678 *Journal of Molecular Sciences* 2019, Vol. 20, Page 5101 **2019**, 20 (20), 5101.  
679 <https://doi.org/10.3390/IJMS20205101>.
- 680 (80) Bae, H.; Park, S.; Yang, C.; Song, G.; Lim, W. Disruption of Endoplasmic Reticulum and  
681 ROS Production in Human Ovarian Cancer by Campesterol. *Antioxidants* 2021, Vol. 10, Page  
682 379 **2021**, 10 (3), 379. <https://doi.org/10.3390/ANTIOX10030379>.
- 683 (81) Pratiwi, R.; Nantasenamat, C.; Ruankham, W.; Suwanjang, W.; Prachayasittikul, V.;  
684 Prachayasittikul, S.; Phopin, K. Mechanisms and Neuroprotective Activities of Stigmasterol  
685 Against Oxidative Stress-Induced Neuronal Cell Death via Sirtuin Family. *Frontiers in*  
686 *Nutrition* **2021**, 0, 184. <https://doi.org/10.3389/FNUT.2021.648995>.
- 687 (82) Lee, J.-H.; Lee, J. H.; Lee, K.-T. Chemical and Volatile Characterization of Structured Lipid  
688 from Soybean Oil Containing Conjugated Linoleic Acid. *Journal of Food Science and*  
689 *Nutrition* **2003**, 8 (3), 219–224.
- 690 (83) Hu, W.; Zhang, L.; Li, P.; Wang, X.; Zhang, Q.; Xu, B.; Sun, X.; Ma, F.; Ding, X.  
691 Characterization of Volatile Components in Four Vegetable Oils by Headspace Two-  
692 Dimensional Comprehensive Chromatography Time-of-Flight Mass Spectrometry. *Talanta*  
693 **2014**, 129, 629–635. <https://doi.org/10.1016/J.TALANTA.2014.06.010>.
- 694 (84) Hu, B.; Yang, C.; Iqbal, N.; Deng, J.; Zhang, J.; Yang, W.; Liu, J. Development and  
695 Validation of a GC–MS Method for Soybean Organ-Specific Metabolomics. *Plant Production*  
696 *Science* **2018**, 21 (3), 215–224. <https://doi.org/10.1080/1343943X.2018.1488539>.
- 697 (85) Zhao, M.; Guo, R.; Li, M.; Liu, Y.; Wang, X.; Fu, H.; Wang, S.; Liu, X.; Shi, L. Physiological  
698 Characteristics and Metabolomics Reveal the Tolerance Mechanism to Low Nitrogen in  
699 Glycine Soja Leaves. *Physiologia Plantarum* **2020**, 168 (4), 819–834.  
700 <https://doi.org/10.1111/PPL.13022>.
- 701 (86) Wang, X.; Guo, R.; Li, M.; Liu, Y.; Zhao, M.; Fu, H.; Liu, X.; Wang, S.; Shi, L.  
702 Metabolomics Reveals the Drought-Tolerance Mechanism in Wild Soybean (Glycine Soja).  
703 *Acta Physiologiae Plantarum* **2019**, 41 (9). <https://doi.org/10.1007/s11738-019-2939-1>.
- 704 (87) Gupta, S.; Lee, J. J. L.; Chen, W. N. Analysis of Improved Nutritional Composition of  
705 Potential Functional Food (Okara) after Probiotic Solid-State Fermentation. *Journal of*  
706 *Agricultural and Food Chemistry* **2018**, 66 (21), 5373–5381.  
707 <https://doi.org/10.1021/ACS.JAFC.8B00971>.
- 708 (88) Xue, J.-A.; Mao, X.; Yang, Z.-R.; Wu, Y.-M.; Jia, X.-Y.; Zhang, L.; Yue, A.-Q.; Wang, J.-P.;  
709 Li, R.-Z. Expression of Yeast Acyl-CoA- $\Delta 9$  Desaturase Leads to Accumulation of Unusual  
710 Monounsaturated Fatty Acids in Soybean Seeds. *Biotechnology Letters* 2013 35:6 **2013**, 35  
711 (6), 951–959. <https://doi.org/10.1007/S10529-013-1149-Y>.

- 712 (89) Kongprawes, G.; Wongsawaeng, D.; Hosemann, P.; Ngaosuwan, K.; Kiatkittipong, W.;  
713 Assabumrungrat, S. Improvement of Oxidation Stability of Fatty Acid Methyl Esters Derived  
714 from Soybean Oil via Partial Hydrogenation Using Dielectric Barrier Discharge Plasma.  
715 *International Journal of Energy Research* **2021**, *45* (3), 4519–4533.  
716 <https://doi.org/10.1002/ER.6121>.
- 717 (90) Zhao, Y.; Cao, P.; Cui, Y.; Liu, D.; Li, J.; Zhao, Y.; Yang, S.; Zhang, B.; Zhou, R.; Sun, M.;  
718 Guo, X.; Yang, M.; Xin, D.; Zhang, Z.; Li, X.; Lv, C.; Liu, C.; Qi, Z.; Xu, J.; Wu, X.; Chen,  
719 Q. Enhanced Production of Seed Oil with Improved Fatty Acid Composition by  
720 Overexpressing NAD<sup>+</sup>-Dependent Glycerol-3-Phosphate Dehydrogenase in Soybean. *Journal*  
721 *of Integrative Plant Biology* **2021**, *63* (6), 1036–1053. <https://doi.org/10.1111/JIPB.13094>.
- 722 (91) Alberdi-Cedeño, J.; Ibargoitia, M. L.; Cristillo, G.; Sopolana, P.; Guillén, M. D. A New  
723 Methodology Capable of Characterizing Most Volatile and Less Volatile Minor Edible Oils  
724 Components in a Single Chromatographic Run without Solvents or Reagents. Detection of  
725 New Components. *Food Chemistry* **2017**, *221*, 1135–1144.  
726 <https://doi.org/10.1016/J.FOODCHEM.2016.11.046>.
- 727 (92) Jing, G.; Tang, D.; Yao, Y.; Su, Y.; Shen, Y.; Bai, Y.; Jing, W.; Zhang, Q.; Lin, F.; Guo, D.;  
728 Zhang, W. Seed Specifically Over-Expressing DGAT2A Enhances Oil and Linoleic Acid  
729 Contents in Soybean Seeds. *Biochemical and Biophysical Research Communications* **2021**,  
730 *568*, 143–150. <https://doi.org/10.1016/J.BBRC.2021.06.087>.
- 731 (93) Oliveira, T. J. P.; Cardoso, C. R.; Ataíde, C. H. Fast Pyrolysis of Soybean Hulls: Analysis of  
732 Bio-Oil Produced in a Fluidized Bed Reactor and of Vapor Obtained in Analytical Pyrolysis.  
733 *Journal of Thermal Analysis and Calorimetry 2015 120:1* **2015**, *120* (1), 427–438.  
734 <https://doi.org/10.1007/S10973-015-4600-6>.
- 735 (94) Jiménez, J. J.; Bernal, J. L.; Nozal, M. J.; Toribio, L.; Bernal, J. Profile and Relative  
736 Concentrations of Fatty Acids in Corn and Soybean Seeds from Transgenic and Isogenic  
737 Crops. *Journal of Chromatography A* **2009**, *1216* (43), 7288–7295.  
738 <https://doi.org/10.1016/J.CHROMA.2009.08.015>.
- 739 (95) Sipeiece, E.; Mišina, I.; Qian, Y.; Grygier, A.; Sobieszczńska, N.; Sahu, P. K.; Rudzińska,  
740 M.; Patel, K. S.; Górnaś, P. Fatty Acid Profile and Squalene, Tocopherol, Carotenoid, Sterol  
741 Content of Seven Selected Consumed Legumes. *Plant Foods for Human Nutrition 2021 76:1*  
742 **2021**, *76* (1), 53–59. <https://doi.org/10.1007/S11130-020-00875-3>.
- 743 (96) Ramos, A. H.; Timm, N. da S.; Ferreira, C. D.; Antunes, A. C.; Hoffmann, J. F.; Rios, A. de  
744 O.; Oliveira, M. de. Effects of the Intensification of Soybean Defects: Degradation  
745 Metabolism of Carbohydrates, Organic Acids, Proteins, Lipids, and Phenolics. *Journal of*  
746 *Food Processing and Preservation* **2021**, *45* (6), e15516. <https://doi.org/10.1111/JFPP.15516>.
- 747 (97) Ghosh, S.; Zhang, S.; Azam, M.; Qi, J.; Abdelghany, A. M.; Shaibu, A. S.; Gebregziabher, B.  
748 S.; Feng, Y.; Huai, Y.; Htway, H. T. P.; Agyenim-Boateng, K. G.; Liu, Y.; Feng, H.; Li, J.;  
749 Song, W.; Li, B.; Sun, J. Seed Tocopherol Assessment and Geographical Distribution of 1151  
750 Chinese Soybean Accessions from Diverse Ecoregions. *Journal of Food Composition and*  
751 *Analysis* **2021**, *100*, 103932. <https://doi.org/10.1016/J.JFCA.2021.103932>.

- 752 (98) Naegeli, H.; Bresson, J. L.; Dalmay, T.; Dewhurst, I. C.; Epstein, M. M.; Firbank, L. G.;  
753 Guerche, P.; Hejatko, J.; Moreno, F. J.; Mullins, E.; Nogué, F.; Rostoks, N.; Sánchez Serrano,  
754 J. J.; Savoini, G.; Veromann, E.; Veronesi, F.; Álvarez, F.; Ardizzone, M.; Fernandez, A.;  
755 Gennaro, A.; Gómez Ruiz, J. Á.; Kagkli, D. M.; Lanzoni, A.; Neri, F. M.; Papadopoulou, N.;  
756 Paraskevopoulos, K.; Raffaello, T.; Streissl, F.; de Sanctis, G. Assessment of Genetically  
757 Modified Soybean GMB151 for Food and Feed Uses, under Regulation (EC) No 1829/2003  
758 (Application EFSA-GMO-NL-2018-153). *EFSA Journal* **2021**, *19* (4).  
759 <https://doi.org/10.2903/j.efsa.2021.6424>.
- 760 (99) Damato, T. C.; Carrasco, L. D. M.; Carmona-Ribeiro, A. M.; Luiz, R. V.; Godoy, R.; Petri, D.  
761 F. S. The Interactions between Surfactants and the Epicuticular Wax on Soybean or Weed  
762 Leaves: Maximal Crop Protection with Minimal Wax Solubilization. *Crop Protection* **2017**,  
763 *91*, 57–65. <https://doi.org/10.1016/J.CROPRO.2016.09.019>.
- 764 (100) Kim, K. S.; Park, S. H.; Kim, D. K.; Jenks, M. A. Influence of Water Deficit on Leaf  
765 Cuticular Waxes of Soybean (*Glycine Max* [L.] Merr.). <https://doi.org/10.1086/510496> **2015**,  
766 *168* (3), 307–316. <https://doi.org/10.1086/510496>.
- 767 (101) Claux, O.; Rapinel, V.; Goupy, P.; Patouillard, N.; Vian, M. A.; Jacques, L.; Chemat, F. Dry  
768 and Aqueous 2-Methyloxolane as Green Solvents for Simultaneous Production of Soybean Oil  
769 and Defatted Meal. *ACS Sustainable Chemistry & Engineering* **2021**, *9* (21), 7211–7223.  
770 <https://doi.org/10.1021/ACSSUSCHEMENG.0C09252>.
- 771 (102) Zhu, H.; Chen, J.; He, Z.; Hao, W.; Liu, J.; Kwek, E.; Ma, K. Y.; Bi, Y. Plasma Cholesterol-  
772 Lowering Activity of Soybean Germ Phytosterols. *Nutrients* **2019**, *Vol. 11*, Page 2784 **2019**,  
773 *11* (11), 2784. <https://doi.org/10.3390/NU11112784>.
- 774 (103) Ranjan, A.; Westrick, N. M.; Jain, S.; Piotrowski, J. S.; Ranjan, M.; Kessens, R.; Stiegman,  
775 L.; Grau, C. R.; Conley, S. P.; Smith, D. L.; Kabbage, M. Resistance against *Sclerotinia*  
776 *Sclerotiorum* in Soybean Involves a Reprogramming of the Phenylpropanoid Pathway and  
777 Up-Regulation of Antifungal Activity Targeting Ergosterol Biosynthesis. *Plant Biotechnology*  
778 *Journal* **2019**, *17* (8), 1567–1581. <https://doi.org/10.1111/pbi.13082>.
- 779 (104) JA, M.; AL, D.; T, K.; AM, H.; WD, N. Soybean Sterol Composition and Utilization by  
780 *Phytophthora Sojae*. *Phytochemistry* **2001**, *58* (3), 423–428. [https://doi.org/10.1016/S0031-](https://doi.org/10.1016/S0031-9422(01)00219-9)  
781 [9422\(01\)00219-9](https://doi.org/10.1016/S0031-9422(01)00219-9).
- 782 (105) Xu, B.; Zhang, L.; Ma, F.; Zhang, W.; Wang, X.; Zhang, Q.; Luo, D.; Ma, H.; Li, P.  
783 Determination of Free Steroidal Compounds in Vegetable Oils by Comprehensive Two-  
784 Dimensional Gas Chromatography Coupled to Time-of-Flight Mass Spectrometry. *Food*  
785 *Chemistry* **2018**, *245*, 415–425. <https://doi.org/10.1016/J.FOODCHEM.2017.10.114>.
- 786 (106) Chen, J.; Tang, G.; Zhou, J.; Liu, W.; Bi, Y. The Characterization of Soybean Germ Oil and  
787 the Antioxidative Activity of Its Phytosterols. *RSC Advances* **2019**, *9* (68), 40109–40117.  
788 <https://doi.org/10.1039/C9RA08771K>.
- 789 (107) Zhu, H.; Chen, J.; He, Z.; Hao, W.; Liu, J.; Kwek, E.; Zhao, Y.; Ma, K. Y.; He, W.-S.; Chen,  
790 Z.-Y. Soybean Germ Oil Reduces Blood Cholesterol by Inhibiting Cholesterol Absorption and  
791 Enhancing Bile Acid Excretion. *Food & Function* **2019**, *10* (4), 1836–1845.  
792 <https://doi.org/10.1039/C8FO02585A>.

793

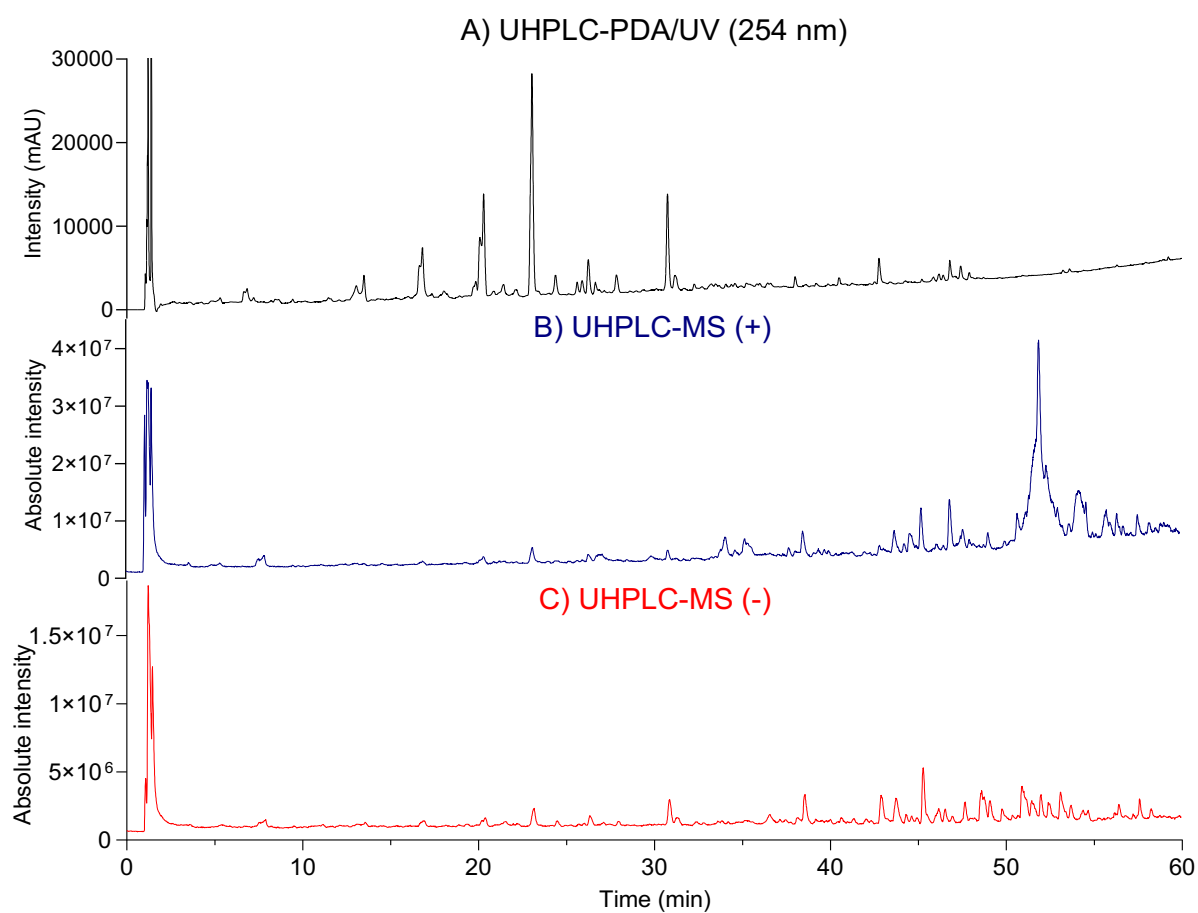
794

795 **FUNDING**

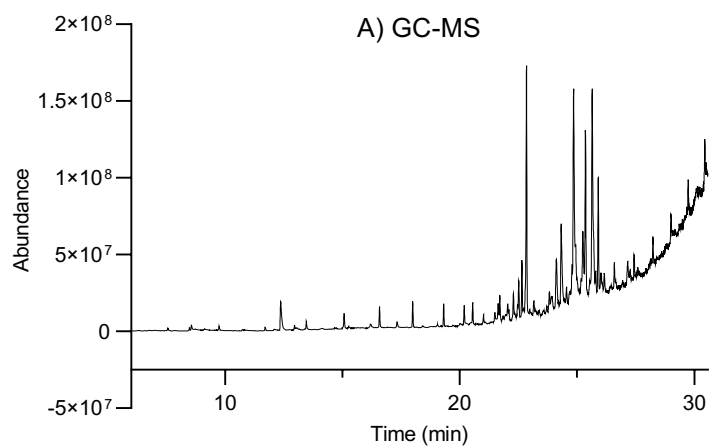
796 The authors would like to thank the São Paulo Research Foundation - FAPESP (2020/09500-0,  
797 2018/21128-9, 2018/01786-1, 2017/06216-6, and 2014/50926-0) and the Coordenação de  
798 Aperfeiçoamento de Pessoal de Nível Superior - Brasil - CAPES (Finance Code 001). G.A.R. would  
799 like to acknowledge the Ministry of Science and Innovation (MICINN, Spain) for a “Juan de la Cierva-  
800 Incorporación” postdoctoral grant IJC2019-041482-I.

801

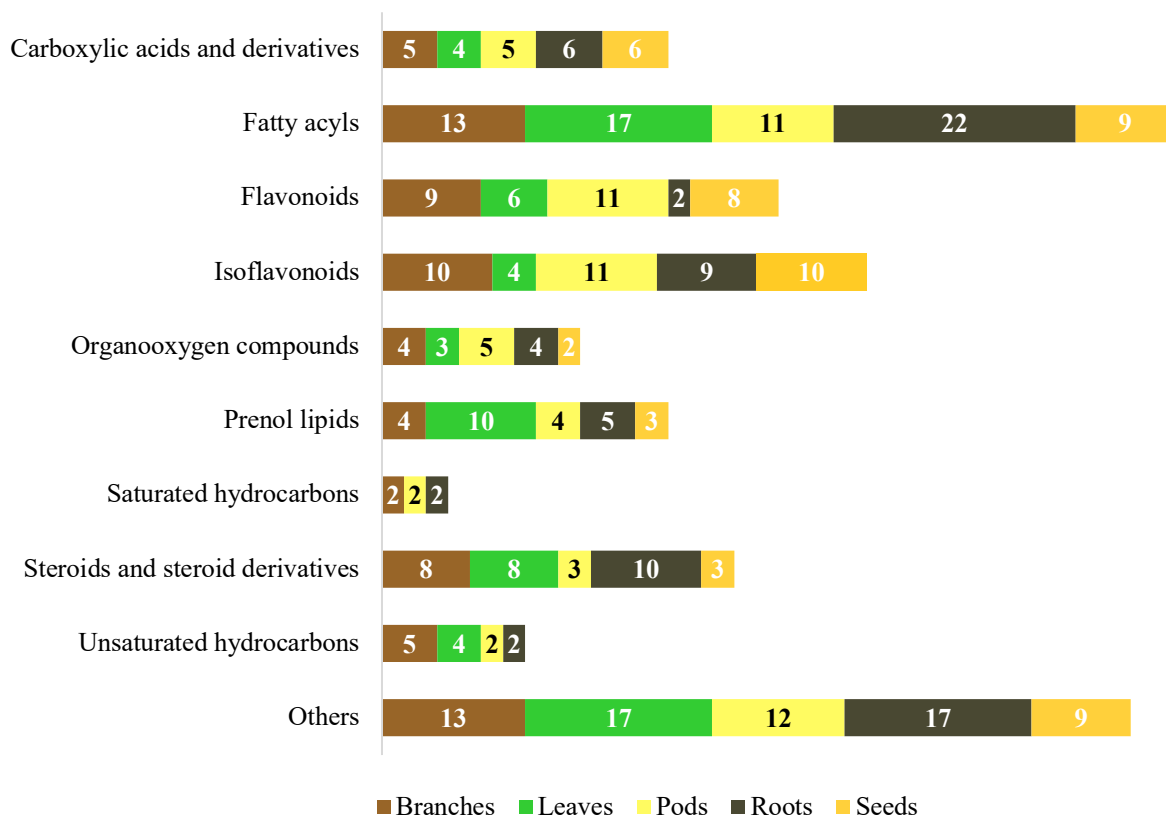
## FIGURE CAPTIONS



**Figure 1.** Representative UHPLC-PDA/UV-MS chromatograms of a mix of soy by-products (branches, leaves, pods, and roots) with beans. Column: Agilent Zorbax Eclipse Plus C18 column (150 mm  $\times$  2.1 mm  $\times$  1.8  $\mu$ m). Mobile-phase: 0.1 % HCOOH in H<sub>2</sub>O and MeOH from 15 to 100 % of MeOH in 60 min. Flow rate: 0.3 mL/min. Analysis temperature: 30 °C. Injection volume: 2  $\mu$ L of non-concentrated EtOH-H<sub>2</sub>O 7:3 (v/v) extract.



**Figure 1.** Representative optimized GC-MS chromatogram of the mix of soy by-products (branches, leaves, pods, and roots) with beans. Column: PerkinElmer Elite-5MS column (30 m  $\times$  0.25 mm, 0.25  $\mu$ m). Injector and MS transfer line temperature: 330 and 310  $^{\circ}$ C, respectively. Helium flow rate: 1.6 mL/min. Oven temperature: 150  $^{\circ}$ C (2 min); followed by 350  $^{\circ}$ C at the rate of 7  $^{\circ}$ C/min. Injection volume: 2  $\mu$ L of the non-concentrated n-heptane extract with a split flow of 5 mL/min.



**Figure 3.** Number of compounds identified in the different extracts of by-products and soybeans. They were grouped according to Classyfire<sup>33</sup>. For annotation of the compound, please refer to Tables 4 and 5.

## TABLES

**Table 1.** Experimental conditions determined by the two-factor central composite rotatable design for optimization of an UHPLC-PDA/UV-MS condition for EtOH-H<sub>2</sub>O 7:3 (v/v) extracts

Experiment	Initial % of B ( $x_1$ ) <sup>a</sup>	% of HCOOH in the M.P. ( $x_2$ ) <sup>a</sup>	Total numbers of peaks		
			254 nm ( $y_1$ )	ESI(+) ( $y_2$ )	ESI(-) ( $y_3$ )
1	-1 (7.90)	-1 (0.16)	41	72	37
2	1 (22.10)	-1 (0.16)	36	75	39
3	-1 (7.90)	1 (0.44)	29	70	30
4	1 (22.10)	1 (0.44)	27	70	31
5	-1.41 (5)	0 (0.30)	32	73	33
6	1.41 (25)	0 (0.30)	31	70	33
7	0 (15)	-1.41 (0.10)	41	81	42
8	0 (15)	1.41 (0.50)	32	72	36
9 (CP) <sup>b</sup>	0 (15)	0 (0.30)	40	73	35
10 (CP) <sup>b</sup>	0 (15)	0 (0.30)	41	73	36
11 (CP) <sup>b</sup>	0 (15)	0 (0.30)	40	74	36
12 (CP) <sup>b</sup>	0 (15)	0 (0.30)	40	73	40
13 (CP) <sup>b</sup>	0 (15)	0 (0.30)	42	77	39

<sup>a</sup>Codified values are given without brackets, whereas the corresponding real values are indicated in Brackets;

<sup>b</sup>Central point.



**Table 2.** Two-factor central composite rotatable design the optimization of a GC-MS condition for *n*-heptane extracts

Experiment	Ramp rate ( $x_1$ ) <sup>a</sup> (°C/min)	Carrier gas flow ( $x_2$ ) <sup>a</sup> (mL/min)	Total numbers of peaks (GC-MS)
1	-1 (4.60)	-1 (1.20)	56
2	1 (22.10)	-1 (1.20)	51
3	-1 (4.60)	1 (1.80)	44
4	1 (22.10)	1 (1.80)	61
5	-1.41 (4)	0 (1.50)	41
6	1.41 (8)	0 (1.50)	65
7	0 (6)	-1.41 (1.00)	50
8	0 (6)	1.41 (2.00)	56
9 (CP) <sup>b</sup>	0 (6)	0 (1.50)	67
10 (CP) <sup>b</sup>	0 (6)	0 (1.50)	65
11 (CP) <sup>b</sup>	0 (6)	0 (1.50)	66
12 (CP) <sup>b</sup>	0 (6)	0 (1.50)	59
13 (CP) <sup>b</sup>	0 (6)	0 (1.50)	61

<sup>a</sup>Codified values are given without brackets, whereas the corresponding real values are indicated in Brackets;

<sup>b</sup>Central point.

**Table 3.** Doehlert design with four variables ( $x$ ) normalized to one, used for the optimization of a two-liquid phase extraction by dynamic maceration, and the result for each experiment ( $y$ ).

Experiment	Time (min) ( $x_1$ ) <sup>a</sup>	Temperature (°C) ( $x_2$ ) <sup>a</sup>	Stirring (rpm) ( $x_3$ ) <sup>a</sup>	Granulometry (mesh) ( $x_4$ ) <sup>a</sup>	$\Sigma$ number of peaks ( $y$ ) <sup>b</sup>
1	0 (20)	0 (35)	-0,75 (700)	1 (48)	268
2	0 (20)	1 (45)	0 (1000)	0 (31)	277
3	1 (35)	0,5 (40)	0 (1000)	0 (31)	274
4	0,33 (25)	0,5 (40)	1 (1400)	0 (31)	254
5	0,33 (25)	0,5 (40)	0,25 (1100)	1 (48)	293
6	0 (20)	-1 (25)	0 (1000)	0 (31)	212
7	-1 (5)	-0,5 (30)	0 (1000)	0 (31)	199
8	-0,33 (15)	-0,5 (30)	-1 (600)	0 (31)	255
9	-0,33 (15)	-0,5 (30)	-0,25 (900)	-1 (14)	183
10	-1 (5)	0,5 (40)	0 (1000)	0 (31)	244
11	-0,33 (15)	0,5 (40)	-1 (600)	0 (31)	239
12	-0,33 (15)	0,5 (40)	-0,25 (900)	-1 (14)	189
13	1 (35)	-0,5 (30)	0 (1000)	0 (31)	249
14	0,67 (30)	0 (35)	-1 (600)	0 (31)	235
15	0,67 (30)	0 (35)	-0,25 (900)	-1 (14)	222
16	0,33 (25)	-0,5 (30)	1 (1400)	0 (31)	236
17	-0,67 (10)	0 (35)	1 (1400)	0 (31)	233
18	0 (20)	0 (35)	0,75 (1300)	-1 (14)	206
19	0,33 (25)	-0,5 (30)	0,25 (1100)	1 (48)	279
20	-0,67 (10)	0 (35)	0,25 (1100)	1 (48)	244
21 (CP) <sup>c</sup>	0 (20)	0 (35)	0 (1000)	0 (31)	228
22 (CP) <sup>c</sup>	0 (20)	0 (35)	0 (1000)	0 (31)	268
23 (CP) <sup>c</sup>	0 (20)	0 (35)	0 (1000)	0 (31)	264
24 (CP) <sup>c</sup>	0 (20)	0 (35)	0 (1000)	0 (31)	238
25 (CP) <sup>c</sup>	0 (20)	0 (35)	0 (1000)	0 (31)	258

<sup>a</sup>Codified values are given without brackets, whereas the corresponding real values are indicated in Brackets;

<sup>b</sup>The sum of number of peaks obtained by UHPLC-UV/MS (UV at 254 nm, and MS at positive and negative modes) and GC-MS for EtOH-H<sub>2</sub>O 7:3 (v/v) and n-heptane extract, respectively;

<sup>c</sup>Central point.

**Table 4.** List of tentatively identified compounds in the EtOH-H<sub>2</sub>O 7:3 (v/v) extracts of the soy by-products, branches (B), leaves (L), pods (P), and roots (R), and seeds (S) by UHPLC-ESI-QToF-MS.

Peak Number	Rt (min)	Tentative identification	Classification	Molecular formula	Molecular Ion	Measured mass ( $\Delta$ ppm)	MS/MS fragments (relative abundance)	B	L	P	R	S	Ref.
1	1.54	Acetyl-L-Carnitine	Fatty Acyls	C <sub>9</sub> H <sub>17</sub> NO <sub>4</sub>	[M+H] <sup>+</sup>	204.123(-6.5)	85.0289 (100), 43.0190 (41.9), 87.0419 (29.8)		X		X		*
2	1.57	Adenine	Imidazopyrimidines	C <sub>5</sub> H <sub>5</sub> N <sub>5</sub>	[M+H] <sup>+</sup>	136.0618(0.9)	119.0360 (100), 94.0520 (34.8), 82.0490 (17.1)	X	X	X	X		52
3	1.79	5'-Deoxy-5'-(methylsulfinyl)adenosine	5'-deoxyribonucleosides	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>4</sub> S	[M+H] <sup>+</sup>	314.0918(2.7)	136.0639 (100), 97.03099 (61.2), 164.0570 (10.7)	X			X		52-54
4	1.86	Meglutol	Fatty Acyls	C <sub>6</sub> H <sub>10</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	161.0455(8.4)	57.0359 (100), 59.0149 (33.1), 41.0040 (32.1)	X	X	X	X	X	55
5	1.88	Adenosine	Purine nucleosides	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	[M+H] <sup>+</sup>	268.104(-3.5)	136.0619 (100), 137.0610 (11), 119.0350 (8.8)	X	X	X	X	X	52-54

6	1.93	Tyrosine	Carboxylic acids and derivatives	C <sub>9</sub> H <sub>11</sub> NO <sub>3</sub>	[M+H] <sup>+</sup>	182.0812(0.2)	91.0530 (100), 56.9430 (59.8), 136.0740 (49.6)					X	X	52,53
7	2.32	Pantothenic acid	Carboxylic acids and derivatives	C <sub>9</sub> H <sub>17</sub> NO <sub>5</sub>	[M-H] <sup>-</sup>	218.1034(7.8)	71.0459 (100), 44.0149 (72.1), 88.0459 (64.6)	X	X	X	X	X	X	52
8	2.37	N-Fructosyl isoleucine	Carboxylic acids and derivatives	C <sub>12</sub> H <sub>23</sub> NO <sub>7</sub>	[M+H] <sup>+</sup>	294.1547(-4.2)	230.1269 (100), 258.1340 (99.9), 86.0940 (88.1)	X	X	X	X			*
9	2.58	3-[(2S,3R,4S,5S,6R)-6-[[[(2R,3R,4R)-3,4-dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxymethyl]-3,4,5-trihydroxyoxan-2-yl]oxy-2-methylpyran-4-one	Organooxygen compounds	C <sub>17</sub> H <sub>24</sub> O <sub>12</sub>	[M+H] <sup>+</sup>	421.1341(2.5)	127.0390 (100), 69.0329 (11.3), 128.0410 (7.4)	X	X	X	X			*
	[M+Na] <sup>+</sup>				443.116(-4.9)	317.0830 (100), 149.0189 (29.5), 214.0879 (21.2)								
10	2.61	Leucine	Carboxylic acids and derivatives	C <sub>6</sub> H <sub>13</sub> NO <sub>2</sub>	[M+H] <sup>+</sup>	132.1019(6)	72.9369 (100), 43.0499 (98.7), 44.0489 (71)	X	X	X	X	X		52,56,57
11	2.77	Pelargonin	Flavonoids	C <sub>27</sub> H <sub>31</sub> O <sub>15</sub>	[M]	595.1657(5.6)	271.0639 (199), 433.1099 (77.9), 272.0669 (55.2)				X	X		*
12	2.81	Dianthoside	Organooxygen compounds	C <sub>12</sub> H <sub>16</sub> O <sub>8</sub>	[M+H] <sup>+</sup>	289.0918(1.1)	127.0390 (100), 256.0559 (11.1), 128.0449 (10.6)	X	X	X	X			*

13	2.92	Benzoic acid + 2O, O-Hex	Organooxygen compounds	C <sub>13</sub> H <sub>16</sub> O <sub>9</sub>	[M-H] <sup>-</sup>	315.0722(4.9)	109.0299 (100), 153.0209 (41.3)	X	X	X	X	X	*
14	3.43	Glutamyltyrosine	Carboxylic acids and derivatives	C <sub>14</sub> H <sub>18</sub> N <sub>2</sub> O <sub>6</sub>	[M+H] <sup>+</sup>	311.1238(0.1)	136.0740 (100), 84.0439 (98), 165.0540 (77.3)			X		X	58
15	3.47	D-Pantothenic acid	Organooxygen compounds	C <sub>9</sub> H <sub>17</sub> NO <sub>5</sub>	[M+H] <sup>+</sup>	220.1179(-3.4)	90.0540 (100), 58.0639 (27.8), 98.0220 (22.8)			X	X	X	52
16	3.64	Coatline B	Linear 1,3-diarylpropanoids	C <sub>21</sub> H <sub>24</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	451.1246(5.4)	149.0489 (100), 176.0359 (79.8), 207.0429 (77.8)				X		*
17	4.15	2,6-Dihydroxy-2-[(4-hydroxyphenyl)methyl]-4-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[[[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxymethyl]oxan-2-yl]oxy-1-benzofuran-3-one	Flavonoids	C <sub>27</sub> H <sub>32</sub> O <sub>15</sub>	[M-H] <sup>-</sup>	595.1668(1.3)	269.0469 (100), 259.0639 (82.9), 125.0250 (47.1)					X	*
18	4.22	1-[2,4-Dihydroxy-3-[(2S,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]phenyl]-2-hydroxy-3-(4-hydroxyphenyl)propan-1-one	Linear 1,3-diarylpropanoids	C <sub>21</sub> H <sub>24</sub> O <sub>10</sub>	[M-H] <sup>-</sup>	435.1297(4.9)	315.089 (100), 163.0410 (81.2), 190.0290 (76.5)	X		X	X	X	*
19	4.25	3,5-Dihydroxy-2-(4-hydroxyphenyl)-7-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy-2,3-dihydrochromen-4-one	Flavonoids	C <sub>21</sub> H <sub>22</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	449.1089(5.5)	125.0260 (100), 259.0639 (46.8), 269.0480 (46)	X		X	X	X	*

20	4.36	12-Hydroxyjasmonate sulfate	Fatty Acyls	C <sub>12</sub> H <sub>18</sub> O <sub>7</sub> S	[M-H] <sup>-</sup>	305.07(9.4)	96.9629 (100), 59.0139 (43.7), 79.9580 (20.1)	X	X	X	X	X	*
21	4.39	Salicylamide	Phenols	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub>	[M-H] <sup>-</sup>	136.0404(7.3)	41.9990 (100), 93.0350 (99.2), 91.0199 (33.2)	X		X	X		*
22	4.64	3-[(1-Carboxyvinyl)oxy]benzoic acid	Benzene and substituted derivatives	C <sub>10</sub> H <sub>8</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	207.0299(9.2)	93.0329 (100), 79.9560 (40.4), 135.0379 (25.3)				X		*
23	4.69	4-Hydroxycinnamic acid	Cinnamic acids and derivatives	C <sub>9</sub> H <sub>8</sub> O <sub>3</sub>	[M-H] <sup>-</sup>	163.0401(8.8)	119.0490 (100), 93.0339 (53.6), 117.0189 (19.2)	X	X		X	X	56,57
24	4.79	Isoflavone base + 2O, O-Hex	Isoflavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>9</sub>	[M+HCOO] <sup>-</sup>	461.1089(1.9)	253.0529 (100), 44.9980 (48.6), 252.0449 (48.3)	X		X	X	X	*
25	5.00	3-(4-hydroxyphenyl)-7-methoxy-5-[(3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxychromen-4-one	Isoflavonoids	C <sub>22</sub> H <sub>22</sub> O <sub>10</sub>	[M+FA-H] <sup>-</sup>	491.1195(4.3)	283.0620 (100), 282.0580 (41.1), 445.1170 (36.5)					X	*
26	5.08	N-Acetyltryptophan	Carboxylic acids and derivatives	C <sub>13</sub> H <sub>14</sub> N <sub>2</sub> O <sub>3</sub>	[M-H] <sup>-</sup>	245.0932(9.1)	74.0250 (100), 58.0330 (63.3), 116.0500 (40.4)	X		X	X	X	59

27	5.15	Kaempferol 3-O-sophoroside	Flavonoids	C <sub>27</sub> H <sub>30</sub> O <sub>16</sub>	[M-H] <sup>-</sup>	609.1461(3.3)	284.0350 (100), 285.0419 (60.5), 255.0319 (28)	X	X	X	X	54	
28	5.16	Prunin	Flavonoids	C <sub>21</sub> H <sub>22</sub> O <sub>10</sub>	[M-H] <sup>-</sup>	433.114(6.6)	151.0040 (100), 119.0490 (58.1), 271.0640 (53.4)				X	60	
29	5.32	Daidzin	Isoflavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>9</sub>	[M-H] <sup>-</sup>	415.1035(6.1)	252.0460 (100), 253.0520 (58.5), 223.0429 (35.6)	X		X	X	X	17,54,57,6 1-68
30	5.52	D-Tryptophan	Indoles and derivatives	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	[M+H] <sup>+</sup>	205.0972(-0.8)	146.0579 (100), 118.0660 (36.9), 144.0800 (25.6)	X	X	X	X	X	54,57,62
31	5.52	L-Tryptophan	Indoles and derivatives	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	[M+H] <sup>+</sup>	205.0972(-0.8)	118.0650 (100), 91.0530 (97.5), 115.0540 (65.7)	X	X	X	X	X	54,57,62
32	5.54	Indole-3-carboxyaldehyde	Indoles and derivatives	C <sub>9</sub> H <sub>7</sub> NO	[M-H] <sup>-</sup>	144.0455(6.3)	115.0429 (100), 99.9260 (59.9), 65.9990 (44.1)	X		X	X	*	
33	5.62	Luteolin 7-O-glucoside	Flavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	447.0933(6.3)	285.0429 (100), 286.0469 (34.9), 284.0360 (30.4)			X		57	

34	5.65	5,7-Dihydroxy-2-(4-hydroxy-3-methoxyphenyl)-3-[3,4,5-trihydroxy-6-[[[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxymethyl]oxan-2-yl]oxychromen-4-one	Flavonoids	C <sub>28</sub> H <sub>32</sub> O <sub>16</sub>	[M-H] <sup>-</sup>	623.1618(3.6)	299.0230 (100), 314.0469 (56.4), 300.0419 (35.5)	X						*
35	5.78	Methylthioadenosine	5'-deoxyribonucleosides	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>3</sub> S	[M+H] <sup>+</sup>	298.0968(0.9)	136.0619 (100), 84.9589 (22.4), 61.0099 (11.2)					X		53,54
36	17.91 6.19	Apigenin 7-O-glucoside <sup>‡</sup>	Flavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>	[M+H] <sup>+</sup> [M-H] <sup>-</sup>	433.1129(1.1) 431.0984(5.6)	271.0580 (100), 272.0660 (20.7), 273.0650 (2.1) 268.0400 (100), 269.0459 (25.4), 211.0420 (8)	X	X	X		X		69
37	6.05	Kaempferol 3-O-glucoside	Flavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	447.0933(3.6)	284.0350 (100), 255.0319 (88.3), 227.0379 (79.2)	X	X	X				70
38	6.26	Isoleucyl-glutamic acid	Carboxylic acids and derivatives	C <sub>11</sub> H <sub>20</sub> N <sub>2</sub> O <sub>5</sub>	[M+H] <sup>+</sup>	261.1445(3.5)	84.0439 (100), 86.0960 (87.5), 44.0499 (15)	X	X		X	X		*
39	6.37	Isoflavone base + 4O, O-MalonylHex	Flavonoids	C <sub>24</sub> H <sub>22</sub> O <sub>14</sub>	[M-H] <sup>-</sup>	533.0937(4.5)	285.0429 (100), 489.1090 (75), 284.0350 (58.6)	X			X			*
40	6.38	Benzenesulfonimidic acid	Benzene and substituted derivatives	C <sub>6</sub> H <sub>7</sub> NO <sub>2</sub> S	[M-H] <sup>-</sup>	156.0125(7.2)	63.9630 (100), 79.9639 (97.6), 92.0500 (73)				X	X	X	*



41	6.89	Formononetin 7-O-glucoside	Isoflavonoids	C <sub>22</sub> H <sub>22</sub> O <sub>9</sub>	[M+HCOO] <sup>-</sup>	475.1246(3.4)	44.9980 (100), 267.0690 (91.4), 252.0460 (54.8)	X	X	X	X	54,61
	22.10				[M+H] <sup>+</sup>	431.1337(-2.2)	269.0790 (100), 270.0820 (20.6), 213.0899 (10.9)					
42	6.97	Genistein	Isoflavonoids	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	269.0455(8)	117.0350 (100), 63.0239 (64.8), 65.029 (57.1)	X	X	X	X	17,61- 64,68
43	6.98	Kynurenic acid	Quinolines and derivatives	C <sub>10</sub> H <sub>7</sub> NO <sub>3</sub>	[M+H] <sup>+</sup>	190.0499(4.9)	116.0479 (100), 89.0390 (83.8), 144.0420 (29.9)		X	X		60
44	7.01	Isoliquiritigenin	Linear 1,3-diarylpropanoids	C <sub>15</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	255.0663(9.1)	91.0189 (100), 119.0490 (81.7), 44.9990 (63.5)	X	X	X	X	71
45	7.36	Biochanin A	Isoflavonoids	C <sub>16</sub> H <sub>12</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	283.0612(5.3)	91.0189 (100), 211.0429 (91.1), 135.0099 (83.4)	X	X		X	54
46	7.57	4-O-Beta-D-glucosyl-trans-4-coumaric acid	Organooxygen compounds	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub>	[M+NH <sub>4</sub> ] <sup>+</sup>	344.134(-1.7)	147.0440 (100), 165.0540 (84.8), 85.0299 (60.3)			X		*
47	8.15	Naringenin	Flavonoids	C <sub>15</sub> H <sub>12</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	271.0612(9.2)	119.0500 (100), 65.0029 (54.2), 107.0139 (47.2)			X	X	59

48	8.62	Retusin 7-methyl ether	Isoflavonoids	C <sub>17</sub> H <sub>14</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	297.0768(8.9)	239.0380 (199), 195.0470 (85.5), 117.0320 (80.5)	X	X	X	X	*	
49	8.95	Neobavaisoflavone	Isoflavonoids	C <sub>20</sub> H <sub>18</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	321.1132(9.6)	265.0530 (100), 277.0540 (58.3), 91.0190 (38.3)	X			X	57,61	
50	9.07	(2S,3S,4S,5R,6R)-6-[[[(3S,6aR,6bS,8aS,14bR)-8a-carboxy-4,4,6a,6b,11,11,14b-heptamethyl-1,2,3,4a,5,6,7,8,9,10,12,12a,14,14a-tetradecahydropicen-3-yl]oxy]-3,4,5-trihydroxyoxane-2-carboxylic acid	Prenol lipids	C <sub>36</sub> H <sub>56</sub> O <sub>9</sub>	[M-H] <sup>-</sup>	631.3852(4)	75.0090 (100, 85.0289 (86.9), 113.0210 (78.3)		X			*	
51	9.09	2-Methyl-4-oxo-4H-pyran-3-yl 6-O-(4-carboxy-3-hydroxy-3-methylbutanoyl)-beta-D-glucopyranoside	Saccharolipids	C <sub>18</sub> H <sub>24</sub> O <sub>12</sub>	[M+H] <sup>+</sup>	433.1341(-5.4)	127.0390 (100), 128.0410 (6.8), 85.0279 (6.6)	X	X	X	X	X	*
52	9.60	9-Hydroxy-10,12-octadecadienoic acid	Fatty Acyls	C <sub>18</sub> H <sub>32</sub> O <sub>3</sub>	[M-H] <sup>-</sup>	295.2279(9.6)	277.2210 (100), 195.1410 (68.4), 59.0149 (56.9)		X	X	X	X	72
53	10.69 52.56	1-(9Z-Octadecenoyl)-sn-glycero-3-phosphoethanolamine	Glycerophospholipids	C <sub>26</sub> H <sub>50</sub> NO <sub>7</sub> P	[M+HCOO] <sup>-</sup> [M+H] <sup>+</sup>	564.3307(3.2) 520.3398(-2.4)	279.2359 (100), 78.9589 (32.3), 44.9980 (24.1) 184.0690 (100), 104.1029 (30.4), 89.9280 (19.4)	X	X		X	X	*
54	11.23	2,4,6-Trihydroxy-2-[(4-hydroxyphenyl)methyl]-1-benzofuran-3-one	Aurone flavonoids	C <sub>15</sub> H <sub>12</sub> O <sub>6</sub>	[M+H] <sup>+</sup>	289.0707(-3)	243.0630 (100), 215.0690 (68.8), 149.0240 (63.7)				X	*	

55	12.28	Loliolide	Benzofurans	C <sub>11</sub> H <sub>16</sub> O <sub>3</sub>	[M+H] <sup>+</sup>	197.1172(-6.2)	133.1000 (100), 107.0859 (81.2), 179.1049 (44)		X					*
56	14.08	Biochanin A 7-O-glucoside	Isoflavonoids	C <sub>22</sub> H <sub>22</sub> O <sub>10</sub>	[M+H] <sup>+</sup>	447.1286(-4.6)	285.0729 (100), 286.0780 (19.6), 270.0509 (9.3)			X		X		54
57	16.56	Undecaethylene glycol	Organooxygen compounds	C <sub>22</sub> H <sub>46</sub> O <sub>12</sub>	[M+H] <sup>+</sup>	503.3062(-6)	45.0330 (100), 89.0589 (96.6), 133.0839 (34.9)	X						*
58	17.03	3-[(2S,4S,6R)-6-[[[(2R,3R,4R,5S,6S)-3,5-dihydroxy-6-methyl-4-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxymethyl]-3,4,5-trihydroxyoxan-2-yl]oxy-5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one	Flavonoids	C <sub>33</sub> H <sub>40</sub> O <sub>20</sub>	[M+Na] <sup>+</sup>	779.2005(-3.7)	493.1530 (100), 494.1539 (29.7), 347.0929 (8.7)	X	X	X				*
59	17.25	Genistin <sup>§</sup>	Isoflavonoids	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>	[M+H] <sup>+</sup>	433.1129(-2.8)	271.0589 (100), 272.0629 (22.8), 153.0180 (7.5)	X		X	X	X		17,57,61- 65,67
60	17.28	Salipurposid	Flavonoids	C <sub>21</sub> H <sub>22</sub> O <sub>10</sub>	[M+Na] <sup>+</sup>	457.1105(-8.1)	185.0429 (100), 295.0549 (91.8), 337.0520 (74.8)					X		*
61	17.78	Kaempferol 3-O-(2,6-di-O-rhamnosyl)galactoside	Flavonoids	C <sub>33</sub> H <sub>40</sub> O <sub>19</sub>	[M+H] <sup>+</sup>	741.2237(-2.6)	287.0549 (100), 129.0540 (25.4), 288.0570 (19)	X	X	X				62
	18.54				[M+Na] <sup>+</sup>	763.2056(-3.4)	477.1579 (100), 478.1549 (23.6), 331.1000 (22)							

62	18.02	Kaempferol 3-O-beta-sophoroside	Flavonoids	C <sub>27</sub> H <sub>30</sub> O <sub>16</sub>	[M+Na] <sup>+</sup>	633.1426(1.9)	347.0939 (100), 348.0989 (14.8), 308.0270 (7.2)	X	X	X	X	54	
63	20.57	6"-O-Malonylglycitin	Isoflavonoids	C <sub>25</sub> H <sub>24</sub> O <sub>13</sub>	[M+H] <sup>+</sup>	533.129(-4.8)	118.0640 (100), 91.0530 (64.8), 146.0580 (55.4)			X		54,62-64	
64	20.76	6"-O-Malonyldaidzin	Isoflavonoids	C <sub>24</sub> H <sub>22</sub> O <sub>12</sub>	[M+H] <sup>+</sup>	503.1184(-5.4)	255.0650 (100), 256.0660 (21.7), 199.0750 (2.36)	X		X	X	17,54,57,6 1-67	
65	20.99	6"-O-Malonylgenistin	Isoflavonoids	C <sub>24</sub> H <sub>22</sub> O <sub>13</sub>	[M+H] <sup>+</sup>	519.1133(-4.1)	271.0590 (100), 272.0650 (24.8), 153.0170 (3.0)				X	54,57,61- 65,67	
66	25.33	Daidzein	Isoflavonoids	C <sub>15</sub> H <sub>10</sub> O <sub>4</sub>	[M+H] <sup>+</sup>	255.0652(-2.3)	137.0220 (100), 91.0530 (89.9), 199.0749 (73.9)	X		X	X	X	17,54,56,5 7,61-68
67	27.22	Formononetin 7-O-glucoside-6"-O-malonate	Isoflavonoids	C <sub>25</sub> H <sub>24</sub> O <sub>12</sub>	[M+H] <sup>+</sup>	517.1341(-4.4)	269.0799 (100), 270.0840 (17.5), 271.0880 (1.3)			X	X	57	

\*Means that no corresponding compound was reported in "soybean" or "Glycine max" on SciFinder.

‡Apigenin 7-O-glucoside and genistin present the same molecular formula and similar fragment ions in MS

**Table 5.** List of tentatively identified compounds in the *n*-heptane extracts of the soy by-products, branches (B), leaves (L), pods (P), and roots (R), and seeds by GC-QToF-MS.

Peak Number	Rt (min)	Tentative identification	Classification	Match Factor	Molecular formula	B	L	P	R	S	Ref.
1	7.29	4-Nonene, 3-methyl-, (Z)-	Unsaturated hydrocarbons	89.5	C <sub>10</sub> H <sub>20</sub>		X				*
2	7.30	1-Pentene, 3,3-dimethyl-	Unsaturated hydrocarbons	74.3	C <sub>7</sub> H <sub>14</sub>			X			*
3	7.42	(S)-3-Ethyl-4-methylpentanol	Organooxygen compounds	73.7	C <sub>8</sub> H <sub>18</sub> O		X				*
4	7.61	1,3-Dioxolane, 4-methyl-2-propyl-	Dioxolanes	72.3	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>			X			*
5	7.62	Decanoic acid, 3-methyl-	Fatty Acyls	79	C <sub>11</sub> H <sub>22</sub> O <sub>2</sub>		X				<sup>82</sup>
6	7.68	Methyl 2-hydroxydecanoate	Fatty Acyls	76.9	C <sub>11</sub> H <sub>22</sub> O <sub>3</sub>		X				*
7	7.78	Cyclohexane, (1,2-dimethylbutyl)-	Saturated hydrocarbons	72.8	C <sub>12</sub> H <sub>24</sub>			X			*
8	7.95	Ethanone, 1-cyclopropyl-	Organooxygen compounds	81	C <sub>5</sub> H <sub>8</sub> O	X					*
9	8.09	4-Octene, 2,3,6,7-tetramethyl-	Unsaturated hydrocarbons	85.2	C <sub>12</sub> H <sub>24</sub>	X	X	X			*
10	8.87	2,4-Dimethylcyclopentanone	Organooxygen compounds	83.5	C <sub>7</sub> H <sub>12</sub> O		X				*
11	9.05	5-Undecene, 7-ethenyl-	Unsaturated hydrocarbons	72.7	C <sub>13</sub> H <sub>24</sub>		X				*
12	10.02	3-Hepten-1-ol	Fatty Acyls	71	C <sub>7</sub> H <sub>14</sub> O			X			<sup>83</sup>
13	10.08	<i>cis</i> -7,10-hexadecadienal	Fatty Acyls	90.6	C <sub>16</sub> H <sub>28</sub> O				X	X	*
14	10.17	1,6-Octadiene, 5,7-dimethyl-, (R)-	Unsaturated hydrocarbons	84.8	C <sub>10</sub> H <sub>18</sub>				X		*

15	10.41	Hexadecanoic acid	Fatty Acyls	89.5	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	X	X	X	X	X	84
16	10.50	17-Octadecynoic acid	Fatty Acyls	79.5	C <sub>18</sub> H <sub>32</sub> O <sub>2</sub>				X		*
17	11.22	1,4,9-Decatriene, (E)-	Unsaturated hydrocarbons	82.1	C <sub>10</sub> H <sub>16</sub>				X		*
18	11.99	E-2-Octadecadecen-1-ol	Organooxygen compounds	72.7	C <sub>18</sub> H <sub>36</sub> O		X				*
19	12.03	Cyclopentane	Unsaturated hydrocarbons	76.1	C <sub>5</sub> H <sub>8</sub>	X					*
20	12.30	Phytol	Prenol lipids	90.5	C <sub>20</sub> H <sub>40</sub> O		X				85,86
21	12.62	<i>trans</i> -13-Octadecenoic acid	Fatty Acyls	75.9	C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>	X	X	X		X	87
22	12.64	9,12,15-Octadecatrienoic acid, (Z,Z,Z)-	Fatty Acyls	90.2	C <sub>18</sub> H <sub>30</sub> O <sub>2</sub>	X	X	X	X	X	54,56,57
23	12.68	Oxalic acid, cyclohexyl hexyl ester	Carboxylic acids and derivatives	72.3	C <sub>14</sub> H <sub>24</sub> O <sub>4</sub>		X				*
24	12.75	<i>cis</i> -Vaccenic acid	Fatty Acyls	89.7	C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>				X		88
25	12.94	9,12-Octadecadienoic acid, methyl ester, (E,E)-	Fatty Acyls	80.7	C <sub>19</sub> H <sub>34</sub> O <sub>2</sub>				X		89
26	13.00	Octadecanoic acid	Fatty Acyls	87	C <sub>18</sub> H <sub>36</sub> O <sub>2</sub>	X	X	X	X	X	90
27	13.13	Nonanamide	Fatty Acyls	70.4	C <sub>9</sub> H <sub>19</sub> NO		X				*
28	13.17	Hexadecanamide	Fatty Acyls	81.3	C <sub>16</sub> H <sub>33</sub> NO	X			X		70
29	14.56	2-(2,3-dimethylcyclopentyl)propan-1-ol	Prenol lipids	82.4	C <sub>10</sub> H <sub>20</sub> O		X				*
30	14.57	Cyclopentane, 1-methyl-3-(1-methylethyl)-	Saturated hydrocarbons	77.8	C <sub>9</sub> H <sub>18</sub>				X		*
31	14.59	4-Undecene, 6-methyl-	Unsaturated hydrocarbons	82.1	C <sub>12</sub> H <sub>24</sub>	X					*
32	15.00	E,Z-2,13-Octadecadien-1-ol	Fatty Acyls	71.2	C <sub>18</sub> H <sub>34</sub> O				X		*
33	15.00	Cyclohexanemethanol	Organooxygen compounds	81.5	C <sub>7</sub> H <sub>14</sub> O		X				*

34	15.35	9-Octadecenamide, (Z)-	Fatty Acyls	88.7	C <sub>18</sub> H <sub>35</sub> NO				X		91
35	15.39	Eicosanoic acid	Fatty Acyls	84.3	C <sub>20</sub> H <sub>40</sub> O <sub>2</sub>	X	X	X	X		92
36	15.52	Bicyclo[5.2.0]nonane, 4-methylene-2,8,8-trimethyl-2-vinyl-	Unsaturated hydrocarbons	79	C <sub>15</sub> H <sub>24</sub>		X				*
37	15.65	Tetradecanamide	Fatty Acyls	78.5	C <sub>14</sub> H <sub>29</sub> NO	X			X		93
38	16.60	Methyl (8Z,11Z,14Z,17Z)-icosa-8,11,14,17-tetraenoate	Fatty Acyls	70.3	C <sub>21</sub> H <sub>34</sub> O <sub>2</sub>					X	*
39	16.61	1,Z-5,E-7-Dodecatriene	Unsaturated hydrocarbons	74.3	C <sub>12</sub> H <sub>20</sub>	X					*
40	16.96	5-Undecene, 4-methyl-	Unsaturated hydrocarbons	77.6	C <sub>12</sub> H <sub>24</sub>	X					*
41	17.89	Heptanoic acid	Fatty Acyls	91.6	C <sub>7</sub> H <sub>14</sub> O <sub>2</sub>				X		94
42	19.16	5-Octen-4-one, 7-methyl-	Organooxygen compounds	71.3	C <sub>9</sub> H <sub>16</sub> O				X		*
43	19.92	13-Docosenamide, (Z)-	Fatty Acyls	82.5	C <sub>22</sub> H <sub>43</sub> NO	X	X		X		*
44	20.30	Squalene	Prenol lipids	88.2	C <sub>30</sub> H <sub>50</sub>		X				95
45	20.30	(3E,7E)-4,8,12-Trimethyltrideca-1,3,7,11-tetraene	Prenol lipids	76.6	C <sub>16</sub> H <sub>26</sub>	X					*
46	20.87	Docosanoic acid	Fatty Acyls	70.3	C <sub>22</sub> H <sub>44</sub> O <sub>2</sub>				X		92
47	21.18	Octane, 3-ethyl-2,7-dimethyl-	Saturated hydrocarbons	76.6	C <sub>12</sub> H <sub>26</sub>				X		*
48	21.19	Nonacosane	Saturated hydrocarbons	70.8	C <sub>29</sub> H <sub>60</sub>			X			*
49	21.35	1,3-Bis-(2-cyclopropyl,2-methylcyclopropyl)-but-2-en-1-one	Prenol lipids	85.4	C <sub>18</sub> H <sub>26</sub> O		X		X		*
50	21.55	<i>delta</i> -Tocopherol	Prenol lipids	90.6	C <sub>27</sub> H <sub>46</sub> O <sub>2</sub>		X	X	X	X	96
51	22.10	Ergost-5-en-3-ol, acetate, (3.β.,24R)-	Steroids and steroid derivatives	77.9	C <sub>30</sub> H <sub>50</sub> O <sub>2</sub>	X	X	X			*

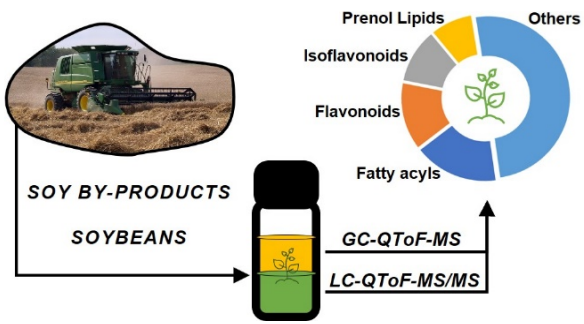
52	22.35	Cholesta-6,22,24-triene, 4,4-dimethyl-	Steroids and steroid derivatives	78.5	C <sub>29</sub> H <sub>46</sub>	X	X		X		*
53	22.40	.beta.-Tocopherol	Prenol lipids	84.1	C <sub>28</sub> H <sub>48</sub> O <sub>2</sub>		X				97
54	22.54	.gamma.-Tocopherol	Prenol lipids	90.3	C <sub>28</sub> H <sub>48</sub> O <sub>2</sub>	X	X		X	X	96-98
55	22.59	4,7,10,13,16,19-Docosahexaenoic acid, methyl ester, (all-Z)-	Fatty Acyls	73.8	C <sub>23</sub> H <sub>34</sub> O <sub>2</sub>				X		*
56	22.72	Cholesta-4,6-dien-3-ol, (3.beta.)-	Steroids and steroid derivatives	70.5	C <sub>27</sub> H <sub>44</sub> O		X				*
57	23.08	(S)-4-methyloctadecanal	Fatty Acyls	96.1	C <sub>19</sub> H <sub>38</sub> O	X	X	X	X		*
58	23.09	Heptacosane	Saturated hydrocarbons	79.8	C <sub>27</sub> H <sub>56</sub>	X					99,100
59	23.09	(R)-4-methyloctadecanal	Fatty Acyls	94.1	C <sub>19</sub> H <sub>38</sub> O		X	X	X		*
60	23.29	alpha-Tocopherol	Prenol lipids	92.7	C <sub>29</sub> H <sub>50</sub> O <sub>2</sub>	X	X	X	X	X	97,98,101
61	23.98	Ergosterol	Steroids and steroid derivatives	85.2	C <sub>28</sub> H <sub>44</sub> O	X	X		X		102,103
62	24.17	Campesterol	Steroids and steroid derivatives	83.7	C <sub>28</sub> H <sub>48</sub> O	X	X	X	X	X	95,101
63	24.40	Stigmasterol	Steroids and steroid derivatives	91.2	C <sub>29</sub> H <sub>48</sub> O	X	X	X	X	X	103
64	24.76	Obtusifoliol	Steroids and steroid derivatives	79.1	C <sub>30</sub> H <sub>50</sub> O		X				91,104,105
65	24.84	Cholesta-22,24-dien-5-ol, 4,4-dimethyl-	Steroids and steroid derivatives	84.3	C <sub>29</sub> H <sub>48</sub> O				X		*
66	24.86	Vitamin A	Prenol lipids	72	C <sub>20</sub> H <sub>30</sub> O			X			*



67	24.86	Pentacosane	Saturated hydrocarbons	88	C <sub>25</sub> H <sub>52</sub>	X					96,100
68	24.96	.beta.-Sitosterol	Steroids and steroid derivatives	93.1	C <sub>29</sub> H <sub>50</sub> O	X	X		X	X	103
69	25.06	Stigmastanol	Steroids and steroid derivatives	77.9	C <sub>29</sub> H <sub>52</sub> O				X		106,107
70	25.29	Cholesta-8,24-dien-3-ol, 4-methyl-, (3.beta.,4.alpha.)-	Steroids and steroid derivatives	72.4	C <sub>28</sub> H <sub>46</sub> O	X					*
71	25.38	Stigmastane, 3-oxo-	Steroids and steroid derivatives	79.8	C <sub>29</sub> H <sub>50</sub> O				X		*
72	25.46	2H-Cyclopropa[g]benzofuran, 4,5,5a,6,6a,6b-hexahydro-4,4,6b-trimethyl-2-(1-methylethenyl)-	Dihydrofurans	85.6	C <sub>15</sub> H <sub>22</sub> O		X				*
73	25.46	Ergosta-4,6,8(14),22-tetraen-3-one	Steroids and steroid derivatives	79.9	C <sub>28</sub> H <sub>40</sub> O				X		*
74	25.52	4,22-Stigmastadiene-3-one	Steroids and steroid derivatives	86.9	C <sub>29</sub> H <sub>46</sub> O	X			X		*
75	25.72	Lupeol	Prenol lipids	85.3	C <sub>30</sub> H <sub>50</sub> O	X		X	X		105
76	26.93	7,11,15-trimethyl-3-methylidenehexadec-1-ene	Prenol lipids	82.9	C <sub>20</sub> H <sub>38</sub>		X				*
77	28.36	Isopropyl linoleate	Fatty Acyls	81.7	C <sub>21</sub> H <sub>38</sub> O <sub>2</sub>	X	X				*
78	30.11	Eicosanoic acid, octadecyl ester	Fatty Acyls	73.2	C <sub>38</sub> H <sub>76</sub> O <sub>2</sub>	X	X				*
79	32.11	1,6,10,14-Hexadecatetraen-3-ol, 3,7,11,15-tetramethyl-, (E,E)-	Prenol lipids	77.5	C <sub>20</sub> H <sub>34</sub> O		X				*

\*Means that no corresponding compound was reported in "soybean" or "Glycine max" on SciFinder.

# GRAPHIC FOR TABLE OF CONTENTS



“For Table of Contents Only”

## METABOLITE PROFILING OF SOY BY-PRODUCTS: A COMPREHENSIVE APPROACH

Felipe Sanchez Bragagnolo<sup>1,2</sup>, Gerardo Álvarez-Rivera<sup>2</sup>, Márcia Cristina Breitreitz<sup>3</sup>, Elena Ibáñez<sup>2</sup>,  
Alejandro Cifuentes<sup>2</sup>, Cristiano Soleo Funari<sup>1\*</sup>

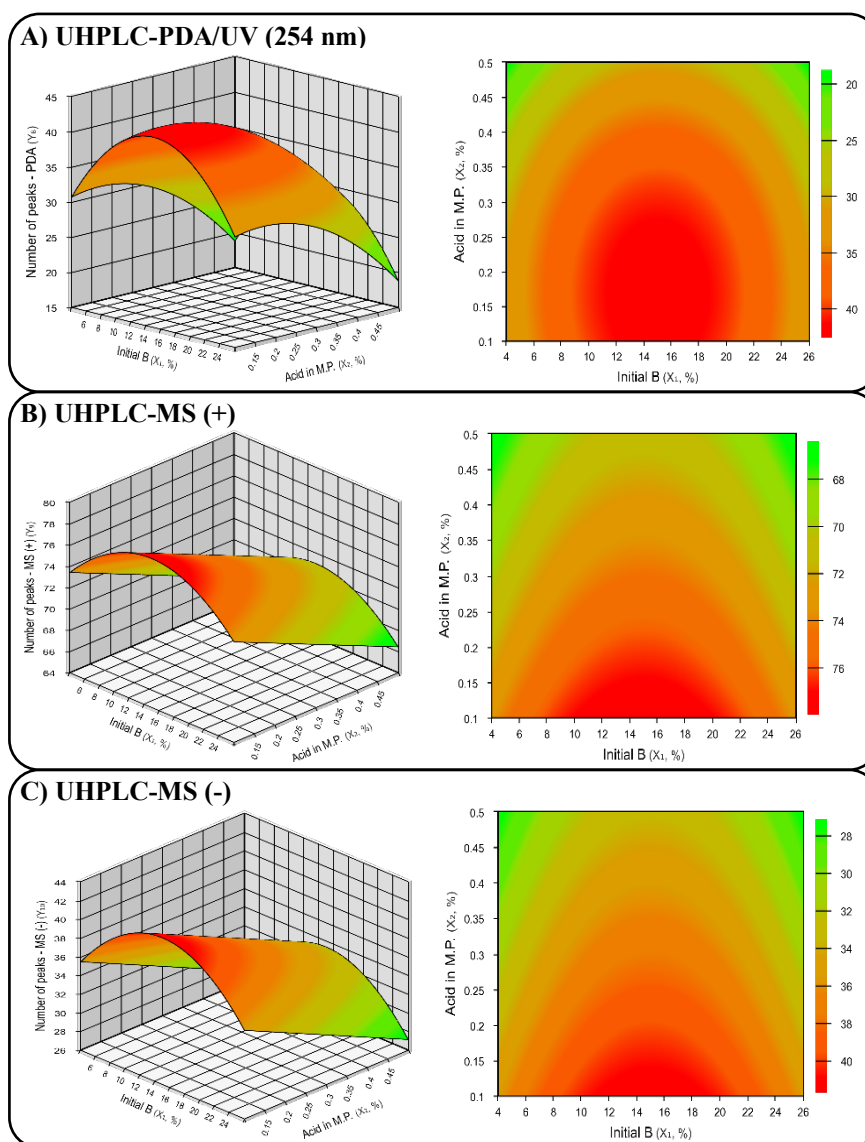
<sup>1</sup> Green Biotech Network, School of Agricultural Sciences, São Paulo State University (UNESP),  
Botucatu, São Paulo - 18610-034, Brazil

<sup>2</sup> Laboratory of Foodomics, Institute of Food Science Research (CIAL-CSIC), Madrid - 28049, Spain

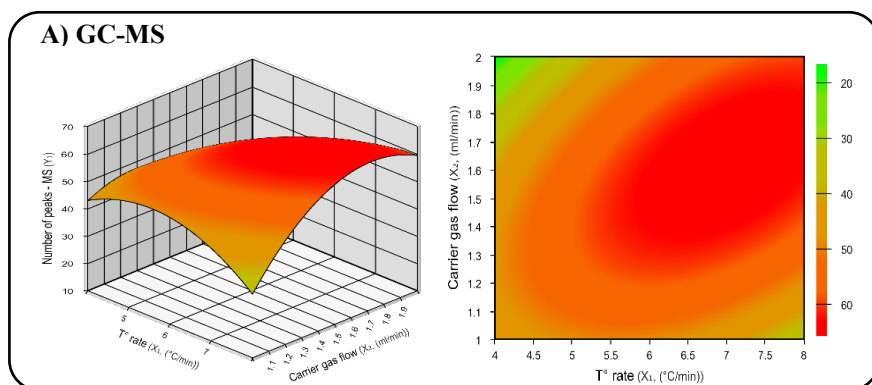
<sup>3</sup> Institute of Chemistry, University of Campinas (UNICAMP), Campinas, São Paulo - 13083-971,  
Brazil

\*E-mail: cristiano.funari@unesp.br

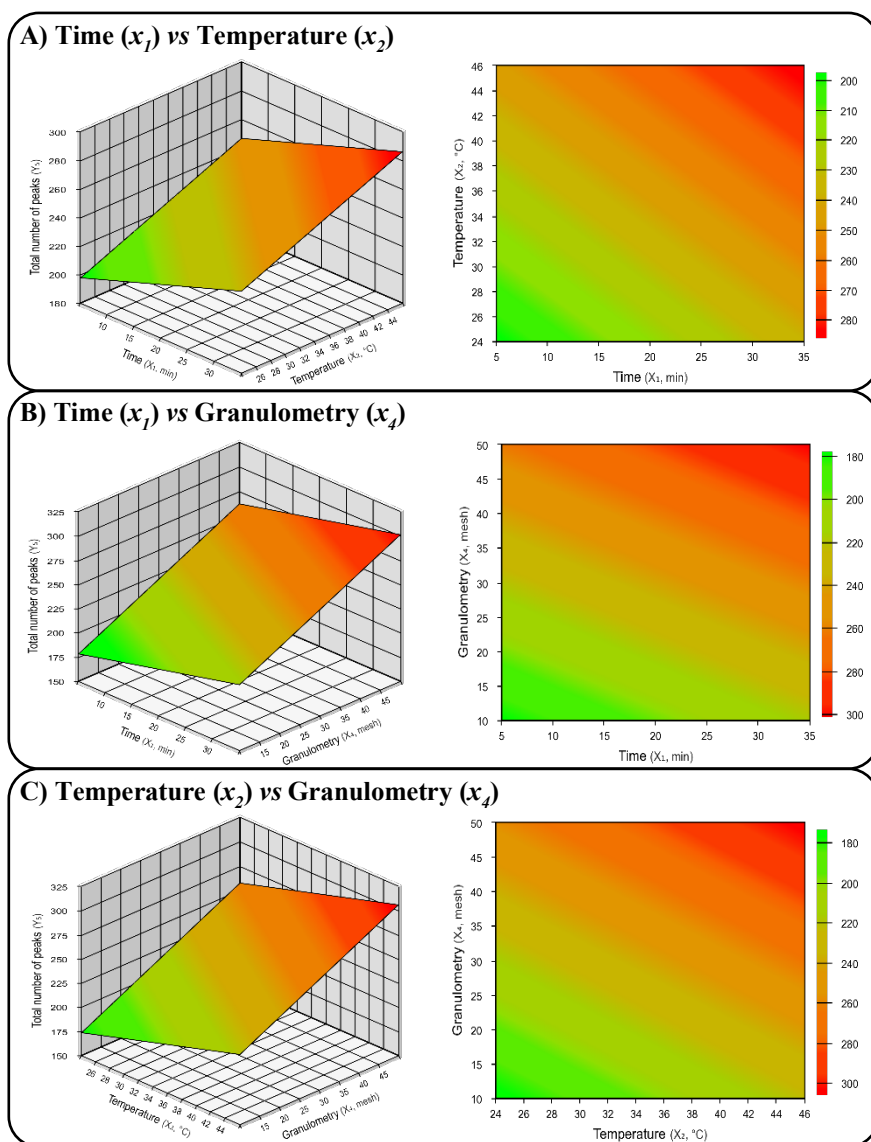
## SUPPORTING INFORMATION



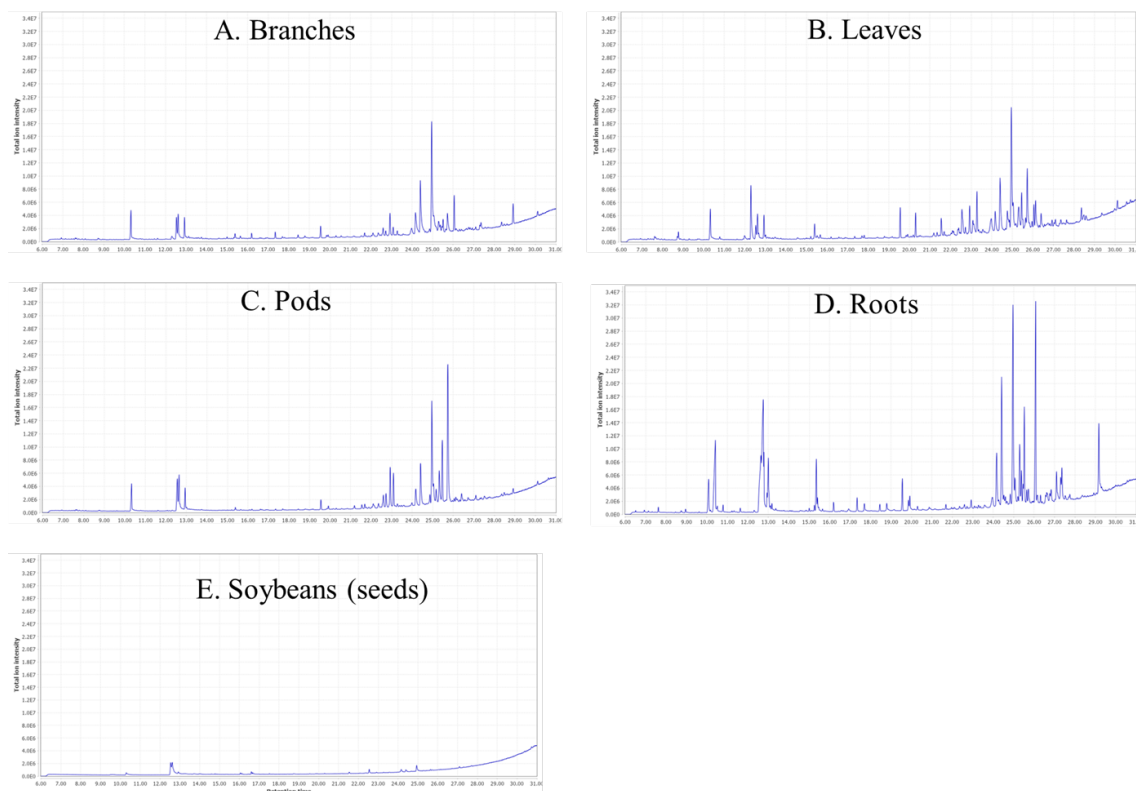
**Figure S1.** Response surface and contour plot for the total number of peaks in UHPLC-PDA/UV at 254 nm (A), and UHPLC-MS in positive (B) and negative mode (C), respectively, as a function of the variables  $x_1$  (initial % of B) and  $x_2$  (% of HCOOH in the M.P.)



**Figure S2.** Response surface and contour plot for the total number of peaks GC-MS as a function of the variables  $x_1$  (Ramp rate - °C/min) and  $x_2$  (carrier gas flow – mL/min)



**Figure S3.** Response surface and contour plot for the 2-Mac using the sum of total number of peaks in UHPLC-PDA/UV-MS and GC-MS as a function of the variables  $x_1$  (time - min),  $x_2$  (temperature – °C), and  $x_4$  (granulometry – mesh size)



**Figure S4.** Representative optimized GC-QToF-MS chromatogram of the individual soy by-products, i.e., branches (A), leaves (B), pods (C), and roots(D), and soybeans (E). Column: Agilent Zorbax DB5- MS+10 m – Duragard Capillary Column (30 m × 250 μm × 0.25 μm). Injector and MS transfer line temperature: 330 and 310 °C, respectively. Helium flow rate: 1.6 mL/min. Oven temperature: 150 °C (2 min); followed by 350 °C at the rate of 7 °C/min. Injection volume: 1 μL of *n*-heptane extract (10 mg/mL) with a split flow of 8.4 mL/min.