

This is a post-peer-review, pre-copyedit version of an article published in ANALYTICAL AND BIOANALYTICAL CHEMISTRY. The final authenticated version is available online at:

<http://dx.doi.org/10.1007/s00216-020-02830-2>

Postprint of: Cruz M.B., Place B.J., Wood L.J. et al., A nontargeted approach to determine the authenticity of Ginkgo biloba L. plant materials and dried leaf extracts by liquid chromatography-high-resolution mass spectrometry (LC-HRMS) and chemometrics. *Anal Bioanal Chem* (2020)

A nontargeted approach to determine the authenticity of Ginkgo biloba L. plant materials and dried leaf extracts by liquid chromatography-high resolution mass spectrometry (LC-HRMS) and chemometrics.

Journal:	<i>Analytical and Bioanalytical Chemistry</i>
Manuscript ID	ABC-00888-2020.R1
Type of Paper:	Research Paper
Date Submitted by the Author:	15-Jul-2020
Complete List of Authors:	Cruz, Meryl; Gdańsk University of Technology; NIST Place, Benjamin; NIST, Wood, Laura; NIST Urbas, Aaron; National Institute of Standards and Technology, Wasik, Andrzej; Gdańsk University of Technology, Analytical Chemistry; Politechnika Gdanska, Analytical Chemistry Rocha, Werickson; National Institute of Metrology, Standardization and Industrial Quality, Metrology Chemistry
Keywords:	nontargeted analysis, adulteration, LC-HRMS, PCA, Ginkgo Biloba

SCHOLARONE™
Manuscripts

A nontargeted approach to determine the authenticity of *Ginkgo biloba* L. plant materials and dried leaf extracts by liquid chromatography-high resolution mass spectrometry (LC-HRMS) and chemometrics

Meryl B. Cruz^{1,2}, Benjamin J. Place^{1,*}, Laura J. Wood¹, Aaron Urbas¹, Andrzej Wasik², Werickson Fortunato de Carvalho Rocha³

¹ Chemical Sciences Division, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Gaithersburg, MD 20899, USA

² Department of Analytical Chemistry, Faculty of Chemistry, Gdańsk University of Technology, 11/12 Narutowicza Street, 80-233 Gdańsk, Poland

³ National Institute of Metrology, Quality and Technology (INMETRO), 25250-020, Xerém, Duque de Caxias, RJ, Brazil

*Corresponding Author, E-mail: benjamin.place@nist.gov

((Footnote))Certain commercial equipment, instruments or materials may be identified in this report to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

To obtain up-to-date official values for NIST reference materials, consult the NIST Standard Reference Material web site at <https://www.nist.gov/srm>.”

Abstract

The lack of stringent regulations regarding raw materials for herbal supplements used for medicinal purposes has been a constant challenge in the industry. *Ginkgo biloba* L. leaf extracts attract consumers because of the supposed positive effect on mental performance and memory. Supplements are produced using dried leaf materials and standardized leaf extracts such as EGb 761. Adulteration of *Ginkgo biloba* L. plants and extracts are becoming more and more common practice due to economically driven motivation from increasing demand in the market and the high cost of raw materials and production. Reinforcement in quality control (QC) to avoid adulterations is necessary to ensure the efficacy of the supplements. In this study, liquid chromatography-high resolution mass spectrometry (LC-HRMS) was used with principal component analysis (PCA) as an unsupervised exploratory method to analyze, identify, and evaluate the adulterated *Ginkgo biloba* L. plant materials and dried leaf extracts using the PCA scores and loadings obtained and compound identification.

Keywords: Nontargeted analysis, *Ginkgo biloba* L., Adulteration, LC-HRMS, PCA

Introduction

1 *Ginkgo biloba* L. (Ginkgoaceae) or the Maidenhair tree is the sole living specie of the
2 Ginkgophyta division and is considered a living fossil dating back to 180 million years based
3 on the fossil records where its genus was once a diverse taxon [1]. *Ginkgo biloba* is native to
4 China but, can be found as ornamental trees in countries with warm temperate climates such
5 as Japan, Korea, Australia, some parts of Europe and North America [1-3]. Most of the
6 scientific and medicinal research of *Ginkgo biloba* L. focuses on the leaves and the extracts,
7 because these contain the active constituents such as flavonoids and terpene trilactones
8 (TTLs), to which the supposed health benefits are attributed. The use of ginkgo leaf extracts
9 started in the 1960s in Germany for the improvement of the blood circulation, to fight fatigue,
10 as an aid for early dementia, for memory improvement, and as a cure for tinnitus [2, 4, 5].
11 The antioxidant effects of ginkgo leaves were attributed to flavonol glycosides, which are the
12 most prevalent group of flavonoids in ginkgo leaves especially the derivatives of quercetin,
13 kaempferol, and isorhamnetin [1, 4, 6, 7]. Terpene trilactones, including ginkgolide A (GA),
14 ginkgolide B (GB), ginkgolide C (GC), a minor TTL named ginkgolide J (GJ), and
15 bilobalide, are considered the true markers of a pure ginkgo leaves as they are uniquely
16 attributed to *Ginkgo biloba* [2, 4, 5, 7].

17 The demand for natural products has increased in the dietary supplement industry
18 resulting in the large-scale cultivation of ginkgo in several parts of the world. According to a
19 review by S. Gafner in 2018 [8], a consistent growth in the global demand in leaf extracts
20 was observed from US \$118 million in 2013 to US \$162 million in 2016. The estimated
21 global demand for dried ginkgo leaves was 60,000 metric tons in 2014. Manufacturers from
22 ConsumerLab.com reported that the cost for a *Ginkgo biloba* extract varied between US \$35
23 per kg and US \$90 per kg, while the price of rutin, a known adulterant isolated from
24 buckwheat, is approximately US \$10 per kg. Canadian Phytopharmaceutical Corp. also
25 reported that the ginkgo extracts from Chinese manufacturers in 2015 cost between US \$150
26 per kg and US \$240 per kg, while Japanese sophora flower extracts (another widely used
27 adulterant) were sold for US \$30 per kg [8]. For this reason, ginkgo products are susceptible
28 to adulteration specifically to boost the flavonol glycoside content using lower-cost sources.

29 The roles of quality assurance (QA) and quality control (QC) can be significant in
30 the industry in assuring proper plant parts and correct botanical taxon are used in
31 manufacturing the finished product, and consistent quality of ginkgo leaves and extracts,
32 despite their inherent natural variability and chemical complexity. With the numerous studies
33 conducted over the years, the established ratio for ginkgo leaf extracts is 6% or greater

1
2
3 1 terpene trilactones, 24% or greater flavonol glycosides, and less than 5 mg/kg of ginkgolic
4
5 2 acids [1, 2, 5, 9]. Pharmacopoeias, such as the Chinese Pharmacopoeia and USP, establish
6
7 3 testing methods to ensure standardization of raw herbal materials by providing monographs
8
9 4 [10]. Most manufacturers of ginkgo leaf extracts comply with the different pharmacopoeias
10
11 5 in their regions to improve quality control. Chinese Pharmacopoeia and USP directives
12
13 6 include monitoring a quercetin/kaempferol/isorhamnetin (Q/K/I) ratio of the hydrolyzed
14
15 7 extract based on the respective peak areas using HPLC methods with an acceptable range of
16
17 8 1/0.8-1.2/≥1 [8, 9]. Authentic composition is critical in quality control of herbal supplements.
18
19 9 However, adulteration of botanicals is still common due to economical motivation, and can
20
21 10 jeopardize not only the quality but also the safety of the finished products [4, 11].

21
22
23 11 Ginkgo leaf extracts can be adulterated in numerous ways. Spiking the original plant
24
25 12 extracts or product formulations with pure flavonols or flavonol glycosides is the most
26
27 13 common form of adulteration, manufacturers use less expensive materials to achieve the
28
29 14 target chemical specification of 24% flavonol glycosides. The pure flavonols found in ginkgo
30
31 15 products like rutin, quercetin and kaempferol are the typical compounds used in spiking as
32
33 16 they are highly effective in inflating the assay values of flavonol glycosides. However, it was
34
35 17 observed that as the total flavonol content increases, the authenticity of ginkgo decreases for
36
37 18 these adulterated samples [11]. The other parts of *Ginkgo biloba* (roots, stem, and seeds)
38
39 19 were also used to reduce the cost of manufacture but, since these plant parts contain a
40
41 20 different set of active components, may contribute to different physiological effects that
42
43 21 could be harmful to the consumers. Furthermore, fortifications using other flavonol
44
45 22 glycoside-rich extracts such as *Styphnolobium japonicum* (Fabaceae) and *Fagopyrum*
46
47 23 *esculentum* M. (Polygonaceae) of the original *Ginkgo biloba* L. plant extracts have been
48
49 24 reported [8]. It was also noted that with this method of adulteration, additional compounds
50
51 25 might be present as other plant extracts have their own set of active components [8, 10].

52
53 26 A useful tool for benchmark comparison in the prevention of adulteration of
54
55 27 botanicals is the use of Certified Reference Materials (CRMs). CRMs are homogeneous,
56
57 28 stable materials that have been well-characterized for one or more property values and
58
59 29 provide associated uncertainties and traceabilities using validated procedures [5]. Analytical
60
61 30 techniques such as chromatography and spectroscopy have been used extensively to detect,
62
63 31 characterize, and estimate both quantitatively and qualitatively the different bioactive
64
65 32 components in *Ginkgo biloba* L. leaves and leaf extracts to meet the required specifications
66
67 33 especially for the flavonol glycosides. Among the most commonly used techniques are high

1 performance liquid chromatography (HPLC) and gas chromatography (GC) coupled with
2 various detectors such as mass spectrometry (MS), thin-layer chromatography (TLC),
3 inductively coupled plasma mass spectrometry (ICP-MS), nuclear magnetic resonance
4 spectroscopy (NMR), and near infrared spectroscopy (NIR) [2, 7, 12-16]. Liquid
5 chromatography coupled with high resolution mass spectrometry (LC-HRMS) is a useful tool
6 for a nontargeted approach as the full scan acquisition mode allows retrospective analysis
7 without further injections and without limitations in the number of monitored compounds [7].
8 Nontargeted MS provides a holistic approach in which known and unknown compounds are
9 detected, quantified, and all the obtained variables are considered simultaneously as the
10 synergic or total effect between variables are not possible to be examined individually. This
11 type of approach requires multivariate techniques since univariate and classical statistical
12 approaches are unfeasible [17].

13 In this study, an LC-HRMS was utilized as a tool to investigate the authenticity of
14 *Ginkgo biloba* L. samples including dried plant material parts (leaves and stems) and a
15 variety of dried leaf extracts (different water-solvent preparations) from different
16 manufacturers, by a nontargeted approach with the aid of NIST Standard Reference Materials
17 and subsequent data analysis. Principal component analysis (PCA), an unsupervised
18 exploratory technique for multivariate analysis, was used to discriminate and discern patterns
19 in each resulting large dataset to create models that will aid effective detection and
20 identification of adulterated ginkgo samples for quality control purposes.

21 **Materials and Methods**

22 **Ginkgo Samples**

23 A total of 32 samples of *Ginkgo biloba* L. plant materials and dried leaf extracts were
24 used in the study and labelled as datasets A (plant materials) and B (dried leaf extracts).
25 Ginkgo leaves were classified into two types, untreated and steam-treated, and were obtained
26 from the same supplier (source A) while the stem samples were from a different source
27 (source B). Dried leaf extracts were collected from different commercial manufacturers and
28 were prepared using a variety of water-solvent ratios. These ginkgo samples were then
29 adulterated at NIST, randomly labeled A1 through A16 for plant materials or B1 through B16
30 for extract materials and are shown according to the adulteration scheme in Table 1. This
31 table shows the summary of the classification of samples by adulteration and by material
32 source. The samples that were duplicated in the study served as a blind check for the

1 reproducibility of the chemometric analysis. During the LC-HRMS analysis, the samples'
2 identities were not used for the nontargeted analysis and were only used to aid data analysis.

3 The SRMs used in this study were NIST SRM 3246 (*Ginkgo biloba* (Leaves)) for the
4 leaf samples and NIST SRM 3247 (*Ginkgo biloba* (Extract)) for the commercial raw leaf
5 extract samples. The NIST SRM 3247 was prepared according to the German Pharmacopoeia
6 (non-clinical) and was acquired from the manufacturer. Further storage preparations were
7 done at ChromaDex Inc. as stated in the certificate of analysis. The SRMs served as
8 analytical quality control materials to aid in the evaluation of the authenticity of these
9 samples.

10 **Chemicals**

11 All solvents used for LC-HRMS analysis were Optima™ LC-MS grade and were
12 purchased from Fisher Chemical, Fisher Scientific Company L.L.C, Pittsburgh, PA, USA.
13 The extraction solvent was prepared by mixing methanol, water, and formic acid to achieve a
14 concentration of 90:9:1 (volume fraction). Mobile phases A and B for gradient elution were
15 prepared using 0.1% (v/v) formic acid in water and 0.1% (v/v) formic acid in acetonitrile,
16 respectively. Previous studies had reported poor peak shape for terpenoid (-)-bilobalide, a
17 main component of *Ginkgo biloba*, with the use of formic acid in the mobile phase, however,
18 the extraction procedure performed in this study was not meant to identify specific
19 compounds (e.g. bilobalide and other terpene trilactones), but rather to broadly profile the
20 compounds in the sample extracts [7].

21 **LC-HRMS Analysis**

22 A 0.3 g to 0.6 g sample was weighed into pre-weighed 15 mL polypropylene (PP)
23 centrifuge tubes. Approximately 5 mL of extraction solvent was added, the tubes re-weighed,
24 and the mixtures were vortexed to ensure there were no dry sample at the bottom. The
25 samples, including the SRMs, were sonicated for 15 minutes and were centrifuged at 50 Hz
26 for 15 minutes. The supernatant was collected and filtered through a 0.45 µm nylon filter
27 (Phenomenex, Torrance, CA, USA) into a new set of centrifuge tubes. All samples were
28 extracted in duplicate on different days and refrigerated until ready for analysis. The
29 duplicate extracts were not combined subsequently but were run as individual samples. Blank
30 samples were also prepared in duplicate for both sample sets. Separate pooled samples for
31 plant material and leaf extract samples were prepared in a similar way. Using a micropipette,
32 100 µL of each plant or leaf extract sample was placed into a vial then mixed thoroughly

1 using a vortex. All sample extracts were placed in HPLC vials and consequently positioned in
2 the autosampler for LC-MS analysis.

3 The chromatographic separation was performed using a Thermo Ultimate 3000 Liquid
4 Chromatograph coupled with Q-Exactive Hybrid Orbitrap Mass Spectrometer which was
5 controlled with Thermo Scientific Chromeleon Chromatography Data System version 6.80
6 SR11 (Thermo Fisher Scientific, Waltham, MA, USA). The analyses were conducted in
7 reversed phase using a Halo C18 column (2.1 mm x 100 mm, 2.7 μ m particle size, MAC-
8 MOD Analytical Inc., Chadds Ford, PA, USA). Gradient elution was used in the LC
9 separation, because the polarity of the main components present in *Ginkgo biloba* varies. The
10 mass spectrometer was operated using electrospray ionization (ESI) in full scan mode for
11 positive and negative ionization modes, independently. Table 2 shows the detailed
12 chromatographic and mass spectrometer conditions used in the analysis.

13 **Data Analysis for PCA**

14 Experimental data were collected in Microsoft Excel™ 2016 (Microsoft Corporation,
15 Redmond, WA, USA) and processed using the PLS_Toolbox 8.6.2 (Eigenvector Research,
16 Inc., Manson, WA, USA) running in MATLAB R2018a (The Mathworks Inc, Natick, MA,
17 USA). MZmine 2 Version 2.36 software (<http://mzmine.github.io/>), a Java-based open source
18 software used for data processing, feature extraction, and differential profiling, was also used
19 to pre-process the MS/MS data before importing it to MATLAB [18, 19].

20 Preprocessing using MZmine 2 software was performed based on approaches for
21 nontargeted metabolomics and lipidomics datasets optimized internally at NIST as shown in
22 Figure 1. The workflow is composed of several data processing stages and requires different
23 sets of criteria to be optimized. The LC-MS1 data of the instrument full scan raw data were
24 converted to an open source format (.mzxml file) using ProteoWizard MS convert tool before
25 importing into the MZmine. Datasets for LC-HRMS were divided into four groups: A
26 negative, A positive, B negative and B positive, with A and B describing the plant materials
27 and leaf extracts, respectively, while the terms positive and negative designate the ionization
28 polarity modes used in the analysis. Each dataset was processed and analyzed separately.

29 The nontargeted batch file steps described in Figure 1 were performed first to create
30 the feature peak list that will be used for the samples. Pooled, blank, and SRM samples were
31 imported and a mass list was built using the mass detection step. An appropriate noise level
32 setting was used based on the sensitivity of the instrument and on the original chromatograms

1 and mass spectra of the samples. Ion chromatograms were constructed for each of the masses
2 in the mass list using the chromatogram builder to produce a peak list containing the
3 extracted ion chromatograms (EICs or XICs) for masses that have been detected by mass
4 spectrometer continuously over a certain duration of time. After the EICs were built, peak
5 detection by chromatogram deconvolution was performed using a local minimum search
6 algorithm which aims to find the local minima in the chromatogram as border points between
7 individual peaks and can set restrictions on minimum absolute and relative intensities, or a
8 minimum ratio of peak top or edge [19]. Construction of EICs and detection of
9 chromatographic peaks from the EICs are considered important steps as an ion chromatogram
10 may contain multiple peaks and these functions are useful for the identification and relative
11 quantitation of compounds. Also, errors produced at this stage can spread throughout the data
12 preprocessing and succeeding statistical analysis to be performed [20]. Isotopic peaks grouper
13 was then used to combine the features corresponding to the same analyte with different
14 charge states and isotopomers. Once the data were deisotoped, join aligner was performed to
15 align and combine the peaks based on the retention time and m/z tolerance settings. The final
16 step was filling the gaps by using two functions, the peak finder and the same t_R and m/z
17 range gap filler. Areas without peaks in some scans will be filled in and the peaks with the
18 same t_R and m/z range that were not detected due to being close to the detection limit in the
19 original peak window can be identified.

20 The peak list extracted from MZmine consisting of the column features of row ID,
21 row retention time, row m/z , and the peak areas of the blank, pooled samples, and SRM
22 samples were exported into a comma-separated value (.csv) file. The row ID is defined as the
23 number that identifies the peak list row and this peak list row can have one or several peaks
24 that have the same mass range and retention time range but originating from a different raw
25 data. The row retention time is the representative retention time value (average retention time
26 of all peaks) and the row m/z is the representative m/z value (average m/z value of all peaks)
27 for a row peak. The retention time value or m/z value of each peak is dependent on peak
28 detection method [21]. This peak list was the transformed data matrix after preprocessing
29 using the nontargeted batch file steps. Peak areas of EICs that were higher in the blank than
30 the pooled samples or SRMs were removed manually using Excel. The feature list (1) was
31 created from this peak list in a new .csv file containing only the selected data with the
32 following features in the sequence of row m/z , row retention time, and row ID. The targeted
33 batch file steps in Figure 1 were then performed by importing the samples and SRMs and by

1 using the feature list (1) as the “targeted peak list” in the targeted peak detection in
 2 constructing the EICs for the samples. Additional steps including peak list rows filter and
 3 duplicate filter were done apart from the similar steps in the nontargeted batch file procedure.
 4 The final feature list (2) was then extracted and saved in a similar manner as the first feature
 5 list. This is the final preprocessed dataset that was used for multivariate analysis.

6 Preprocessing prior to the use of a chemometric technique is necessary to transform
 7 the measured data into a more suitable form for the data analysis as variables measured can
 8 have different units and systematic effects and interferences may be present which can make
 9 the data analysis difficult. Each data processing step was performed multiple times with
 10 different values to obtain the optimized parameters. The final parameters used in data
 11 processing are summarized in Table 3 for LC-HRMS. The .CSV files of the final feature list
 12 from MZmine 2.0 were imported to MATLAB R2018a for further multivariate analysis. With
 13 the PLS Toolbox, unsupervised exploratory analysis using the PCA with some preprocessing
 14 methods was performed on the extracted data from the *Ginkgo biloba* samples and SRMs.
 15 Using a preprocessing step to transform the data into a suitable form for data analysis can
 16 make data analysis less difficult.

17 All nontargeted results were normalized by sample and extraction solvent masses (for
 18 a relative sample concentration) using the Equation 1 below which is further elaborated in the
 19 Electronic Supplementary Material (ESM).

$$\begin{bmatrix} A_{i,1} \\ A_{i,2} \\ A_{i,3} \\ \vdots \\ A_{i,n-1} \\ A_{i,n} \end{bmatrix} \times \frac{m_{sol}}{m_{sample}} = \begin{bmatrix} C_{i,1} \\ C_{i,2} \\ C_{i,3} \\ \vdots \\ C_{i,n-1} \\ C_{i,n} \end{bmatrix} \quad \text{Equation 1}$$

21 Normalizing the data by reducing the peak area to relative sample concentration can
 22 minimize the within-replicate variability and incorporate the discrepancy from sample
 23 preparation into the concentration values. This calculation assumes that the extraction
 24 efficiency (i.e. recovery) for each individual compound is equal across all samples given the
 25 similar nature of the sample matrices. Peak identification was done using R scripts [22]
 26 linked to the NIST MS Search program (v2.3; <https://chemdata.nist.gov>) by scanning the
 27 final feature list (2) for both positive and negative ion modes obtained from MZMine as these
 28 lists were assumed to contain all the detected peaks in the samples and SRMs.

29 Results and Discussion

1 LC-HRMS analysis was performed using the sample set of 16 plant materials, 16
2 dried leaf extracts, 2 SRMs of leaves and dried leaf extracts, and a pooled sample for each set
3 (plant material and leaf extracts). The nontargeted approach for LC-HRMS was carried out
4 using a full scan mode for both negative and positive ion modes creating four datasets: A
5 negative, A positive, B negative, and B positive, with A and B describing the plant materials
6 and leaf extracts, respectively [23]. These four final data matrices were analyzed as some
7 compounds only appear in one mode or another due to their pH.

8 For the adulteration classification, the data analysis using PCA showed that there
9 were no significant differences in the adulterated samples of groups 3% and 7%. The original
10 PCA results of the entire dataset showed that the groups 3% and 7% were clustered together
11 which may be assumed that the adulteration was significant enough to separate the samples.
12 Thus, the adulteration levels were grouped as 0 % adulteration, 3 % to 7 % adulteration, and
13 15 % adulteration for both plant materials and dried leaf extract samples to give a more visual
14 presentation of the adulteration screening in the PCA score plots.

15 **PCA of Plant Material Samples**

16 The final matrices for dataset A, plant material samples, are summarized in Table 4.
17 Using the adulteration level classification, the PCA score plots of the plant material samples
18 in Fig. 2 shows a separation trend among three levels of adulteration (0 % adulteration, 3 %
19 to 7 % adulteration, and 15 % adulteration) with only mean-centering as the preprocessing
20 method. Loadings have information about variables, in this case, m/z , peak area, and retention
21 time values. Analyzing the results without using a strong preprocessing method can be useful
22 to examine the raw loadings that will enable identification of the significant peaks
23 responsible for the separation. The mean-centered results of the plant material (dataset A) for
24 both positive and negative ion modes using the first four principal components had total
25 variation explained of 98.56 % and 97.02 %, respectively. The best separation of samples by
26 adulteration level was obtained using a combination of PCs 1 and 3 for the A positive ion
27 mode and PCs 2 and 3 for the A negative ion mode as shown in Figure 2. The number of
28 principal components for all the score plots created was selected based on the variance
29 captured (%) plot with the principal component containing a percent variance greater than 1
30 %.

31 Using NIST-MS Search, a summary of the identified compounds present in the
32 positive ion mode and the negative ion mode are presented in Table 5. For high-resolution

1 mass spectrometry, any compound with a match factor (MF) higher than 500 is considered a
2 “Good Match”, which is a tentative identification but not definitive. As mentioned on the
3 Data Analysis for PCA under the Materials and Methods Section, the feature list (2),
4 containing all the sample features, was scanned on the database instead of the 68 individual
5 sample results. It was a more efficient way to identify the compounds for all the samples as
6 the data tool used was PCA and the same feature list was used to build the PCA plots. The
7 only disadvantage of scanning the feature list was that samples containing the identified
8 compounds cannot be presented in this study. However, the score plots (samples) and the
9 loading plots (variables) were found to have a strong correlation on the adulterated samples
10 and their corresponding loadings which were the variables (m/z , peak area, and retention time
11 values) as shown in Figure 2. This also shows how LC-HRMS plays a role in terms of its
12 high sensitivity by detecting high mass accuracies, in this case, it detected up to 4 decimal
13 places for the identified compounds especially for sophoricoside and genistein, which can be
14 strong evidences of adulteration.

15 Figures 2A and 2C show that the plant material samples were separated along PC 3
16 for both modes and the encircled loadings (Figs. 2B and 2D) on the loadings plot suggest a
17 correlation on the adulterated samples. The separations were not distinct however, the score
18 plots exhibited a clear trend with respect to the different adulteration levels that did not
19 appear from other methods using the same samples in the master thesis study conducted by
20 Cruz, M., including ICP-MS, NIR, and GC-MS [23]. For the positive ion mode, loadings ID
21 47 on the positive quadrant along PC3 (Fig. 2B), identified as sophoricoside, was one of the
22 variables with the highest loadings with respect to differentiating the adulterated samples. For
23 the negative ion mode, loading IDs 22 and 29 correspond to the sophoricoside and genistein,
24 respectively, which contribute significantly to the discrimination of the adulterated samples
25 along PC2 (Figs. 2C and 2D). For the positive and negative ion modes (Figs. 2A and 2C), the
26 repeatability of the SRM 3246 was observed to be slightly different in the MS1 normalized
27 dataset and PCA results. The differences of the position of two SRMs may be attributed to
28 the sample preparation as it was done on different days. It was also observed that one of the
29 SRM samples had a different behavior, SRM3246-1 was clustered with the other
30 unadulterated leaves samples, while SRM3246-2 and sample A6 behaved in a similar
31 manner. The differences between SRM 3246 leaves and the ginkgo plant material samples
32 may be attributed to the provenance of the leaves, sample heterogeneity, and storage
33 preparation.

1
2
3 1 Based on the literature, sophoricoside is not an innate compound in *Ginkgo biloba* L.
4 and is specifically found in the dried fruits and flower buds of *Styphnolobium japonicum* (L.)
5 2 Schott (syn. *Sophora japonica* L., Fabaceae) or the Japanese Pagoda tree. This tree is a
6 3 known medicinal plant and one of the commonly alleged adulterants of ginkgo extracts used
7 4 to boost the flavonol glycoside contents due to its lower cost compared to the authentic
8 5 *Ginkgo biloba* extracts [8]. Glycitein, a common compound found from several plants from
9 6 the family Fabaceae including Japanese sophora, was also detected, but was not reported in
10 7 the study [8]. Upon closer examination of the loadings plot of the positive ion of dataset A,
11 8 the loading ID 80, identified as glycitein, was not clearly separated and was clustered with
12 9 the other loadings that were positioned just below the reported loadings (encircled loadings in
13 10 Fig. 2B). This might be due to the differences on the concentration levels of sophoricoside
14 11 and glycitein present in the adulterated samples for this study.
15 12

16 13 In the case of genistein, there was a question over whether genistein is a component of
17 14 *Ginkgo biloba* L. since according to the review and studies of H. Wohlmuth et al. and S.
18 15 Gafner [8, 11], few reports were published concerning authors claiming that genistein was a
19 16 genuine constituent of *G. biloba* even though only low concentrations were detected. In one
20 17 paper, genistein was purified from a commercial leaf extract however, the authenticity of the
21 18 raw material used to manufacture the ginkgo extract in that study was not demonstrated [24].
22 19 In another publication, quantification of flavonoids using ginkgo plant parts such as leaf,
23 20 stem, and fruit from three authentic ginkgo trees in India was detailed and the authors noted
24 21 that genistein was absent in female ginkgo tree leaves but, was identified in the leaf and stem
25 22 of male ginkgo trees [25]. The reported genistein by Yao et al. [26] had concentrations
26 23 between 5-28 µg/g dry leaf using a validated HPLC-UV method with detection at 350 nm.
27 24

28 25 However, from the data in this study, it suggested that at the levels that genistein was
29 26 detected, it was an indicator of the adulteration. This compound was also directly correlated
30 27 to the adulterated samples based on the PCA results. If genistein was present in the
31 28 unadulterated samples, then it was below the detection limit of the qualitative technique. In
32 29 the study of López-Gutierrez [7], the isoflavone genistein was detected in low concentrations
33 30 (between 0.02 and 2.41 mg/g) together with the remarkably high concentrations of rutin
34 31 (27.2-38.2 mg/g) in three products. They also reported the presence of glycitein in two
35 32 products which clearly an indicator of adulteration [7, 8]. Genistein has been reported to be
36 33 native to the pericarp of fruits and flowers of *Styphnolobium japonicum* L., and consequently,
37 34 researchers have proposed that genistein can be used as a marker to detect adulteration with
38 39 extracts of Japanese sophora [8, 11, 24].
39 40
40 41
41 42
42 43
43 44
44 45
45 46
46 47
47 48
48 49
49 50
50 51
51 52
52 53
53 54
54 55
55 56
56 57
57 58
58 59
59 60
60 61

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

1 Figure 3 shows that autoscaling was the best fit for both positive and negative dataset
2 A. Autoscaling compares the variables based on correlations and these variables become
3 equally important. A disadvantage of this approach is that measurement errors will increase
4 as noise and interferences are also adjusted at the same level as those of relatively large
5 variables [27]. Figs. 3A and 3B show the score plots of both ion modes and showed
6 consistent separations between PC2 and PC3. Using the first three principal components, a
7 total of 94.21 % cumulative variance for positive ion mode and 90.26 % for negative ion
8 mode using adulteration level classifications were obtained for LC-HRMS method and these
9 % cumulative variances were not observed from other analytical techniques [23].
10 Concatenated plant material (dataset A) of the normalized positive and negative ion modes
11 data extracted from MZmine and the concatenated principal components of both ion modes
12 were also observed and are shown in Figs. 3C and 3D with the separations between PC2 and
13 PC3, and PC1 and PC3, respectively. LC-HRMS provided adequate results in comparison
14 with other aforementioned methods based on the same plant part material samples especially
15 for small adulterations [23]. However, a model based on the LC-HRMS results still needs
16 additional resources, such as additional authentic samples, to improve the robustness of the
17 PCA models obtained.

33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

18 The PCA score plots for the plant part materials by source classification were found to
19 have a distinct separation among samples for both ion modes with and without preprocessing,
20 as well as the combined results as shown in the Figure, Supplemental Information 2.

21 **PCA of Dried Leaf Extract Samples**

22
23
24
25
26
27
28
29
30
31
32

22 For the dried leaf extracts (dataset B), 34 samples by 58 variables yielded dataset B
23 negative, and 34 samples by 163 variables yielded dataset B positive. Results for dataset B
24 were also combined however, are not reported as trends or separation for adulteration level
25 were not observed. All score plots of dataset B used class centroid centering and scaling, a
26 class-aware type of autoscaling that is useful for samples in subsets identified by a row class
27 set, as the preprocessing method. In this approach, the data are centered by class centroid
28 method to avoid the mean being dominated by the most populous subset, and then scaled by
29 the pooled standard deviation of the classes. Samples that belong to an unknown class are not
30 used in the calculation of centroid or pooled variance [28].

31
32

31 The dried leaf extract samples (dataset B) were examined based on adulteration level
32 per material source as the material source variation appeared to overshadow the adulteration

1
2
3 1 level which may be attributed to the origin of leaves, solvent extracts and composition used
4 during the manufacturing processes of the dried leaf. Table 6 shows the different extraction
5 2 solvents and ratios based on the COA of dried leaf extracts. Similar to SRM 3246 leaves and
6 3 plant material samples in dataset A, SRM 3247 also behaved differently with respect to the
7 4 rest of the ginkgo leaf extract samples as observed in the PCA score plots. This might be
8 5 attributed to the differences in the provenance of the leaves, dried leaf extract preparation and
9 6 sample heterogeneity. Score plots in Fig. 4A and 5A represent the complete dataset identified
10 7 by material source for both positive and negative ion modes which were separated between
11 8 PC 1 and PC2 with a total cumulative variance of 87.84 %, and PC1 and PC3 with 89.24 %,
12 9 respectively. The remaining score plots for dataset B show the adulteration level for each
13 10 material source.
14 11

12 **Conclusions**

13 The use of LC-HRMS and PCA to determine the authenticity of *Ginkgo biloba* L.
14 samples enabled transformation of the results into a new set of data containing principal
15 components and projection into PCA models aided the visualization and evaluation of the
16 data. Determination of adulterated samples among the plant materials (dataset A) and the
17 dried leaf extracts (dataset B) was possible using the score plots obtained in PCA. The ginkgo
18 results of LC-HRMS were transformed into datasets which were divided in two separate
19 ionization modes, negative and positive, creating a total of four datasets: A positive, A
20 negative, B positive, and B negative. The obtained score plots and loadings plots for dried
21 leaf materials on both negative and positive ion modes showed promising results using the
22 adulteration level classification as separation of adulterated samples from unadulterated ones
23 were visible in the score plots even only using mean centering as the preprocessing method.
24 Furthermore, a clear correlation between the adulterated samples and the variables that
25 influences the sample behavior was also observed using the loadings. Consequently, these
26 loadings were inspected showing significant variables that were selected and identified using
27 NIST-MS Search which include the presence of sophoricoside, for both negative and positive
28 ion modes, and genistein for positive ion mode. Based on the phytochemical investigations
29 and literature searches, these compounds are not known to be native in *Ginkgo biloba* L. and
30 in the case of genistein, if it is a genuine component, it will be only be detected in very low
31 concentrations. This could be an indication that these ginkgo plant samples were indeed
32 contaminated or adulterated, possibly with the extracts from *Styphnolobium japonicum* L. or
33 *Sophora japonica* L. plant which is known to have the said compounds.

1 The LC-HRMS results for the ginkgo extract samples did not obtain initial separations
2 based on adulteration level but by observing the trends for each individual source using the
3 adulteration level classification, it showed a possibility of using the method for investigation
4 purposes as the unadulterated samples were separated from the adulterated extracts. With the
5 LC-HRMS results, a different tool such as NMR or NIR may be used to further explore these
6 extract samples and it can be an easier and more efficient method especially for the
7 manufacturers. Compared to the leaf samples, extract samples are initially processed and
8 thus, the origins and manufacturing processes may have contributed to its complexity. The
9 extract samples may have become too similar with each other that the nontargeted approach
10 using LC-HRMS and the type of sample preparation used may not be appropriate and enough
11 to fully discriminate adulterated from unadulterated samples. It must also be considered that
12 the composition of *Ginkgo biloba* L. leaves and dried leaf extracts may vary due to
13 provenance, heterogeneity, and manufacturing processes. For the SRMs, the results are not
14 very useful in this study since their chemical compositions were either too similar or too
15 different with the samples. However, if SRMs are used repeatedly, as in QC purposes for
16 example, consistent results from assay performance can be monitored and consequently the
17 method can be evaluated to be performing as expected.

18 Overall, LC-HRMS method was capable of detecting small adulterations for the plant
19 part material (dataset A). However, from the quality control point of view, LC-HRMS may
20 be a difficult instrument to maintain and to handle for routine purposes. Further improvement
21 of the study can be done such as development of other instrumental techniques as screening
22 methods and the addition of more authentic samples to evaluate and validate the robustness of
23 the PCA models obtained using other chemometric techniques.

24 **Acknowledgements**

25 This project was part of a research master thesis supported by the Education, Audiovisual,
26 and Culture Executive Agency (EACEA) under the program Erasmus Mundus Masters in
27 Quality in Analytical Laboratories (EMQAL 10th edition), Gdansk University of Technology
28 (GUT), and its collaboration with the National Institute of Standards and Technology (NIST
29 Gaithersburg, USA) and the National Institute of Metrology, Quality and Technology
30 (INMETRO Brazil). This project would not be possible without the help and overwhelming
31 support of NIST and GUT supervisors, colleagues and program coordinators.

- 1
2
3 1 emerging approach to determine the composition of herbal products. *Comput Struct*
4 2 *Biotechnol. J.* 2013;4(5):1–7.
- 5
6 3 [18] Katajamaa M, Miettinen J, Orešič M. Processing methods for differential analysis of LC/MS
7 4 profile data. *BMC Bioinformatics.* 2006;22(5):634–636.
- 8
9 5 [19] Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for
10 6 processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC*
11 7 *Bioinformatics.* 2010;11:395.
- 12
13 8 [20] Myers OD, Sumner SJ, Li S, Barnes S, and Du X. One Step Forward for Reducing False
14 9 Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics
15 10 Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting
16 11 Chromatographic Peaks. *Anal Chem.* 2017;89:2.
- 17
18 12 [21] MZmine Development Team. MZmine 2.3 Manual. 2005–2011.
19 13 <http://mzmine.sourceforge.net/manual.pdf>. Accessed 28 Nov 2019.
- 20
21 14 [22] R Core Development Team. R: A language and environment for statistical computing. In: R
22 15 Foundation for Statistical Computing, Vienna, Austria. 2013. <http://www.R-project.org/>.
23 16 Accessed 30 July 2019.
- 24
25 17 [23] Cruz MB. Determination of the authenticity of *Ginkgo biloba* L. plant part materials and dry
26 18 leaf extracts using different analytical methods and chemometric techniques [master's thesis].
27 19 Gdansk, Poland: Gdansk University of Technology; 2019.
- 28
29 20 [24] Wang F, Jiang K, Li Z, Purification and Identification of Genistein in *Ginkgo biloba* Leaf
30 21 Extract. *Chinese J Chromatogr.* 2007;25(4):509–513.
- 31
32 22 [25] Pandey R, Chandra P, Arya KR, Kumar B. Development and validation of an ultra high
33 23 performance liquid chromatography electrospray ionization tandem mass spectrometry method
34 24 for the simultaneous determination of selected flavonoids in *Ginkgo biloba*. *J Sep Sci.*
35 25 2014;37(24):3610–3618.
- 36
37 26 [26] Yao JB et al. Seasonal variability of genistein and 6-hydroxykynurenic acid contents in
38 27 *Ginkgo biloba* leaves from different areas of China. *Nat Prod Commun.* 2017;12(8):1241–
39 28 1244.
- 40
41 29 [27] van den Berg RA, Hoefsloot H-CJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering,
42 30 scaling, and transformations: improving the biological information content of metabolomics
43 31 data. *BMC Genomics.* 2006;7:142.
- 44
45 32 [28] Eigenvector Research Documentation. Advanced Preprocessing: Variable Centering -
46 33 Eigenvector Documentation Wiki.
47 34 http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Variable_Centering.
48 35 Accessed 18 Apr 2019.
- 49
50 36
51 37
52 38
53 39
54 40
55 41
56 42
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

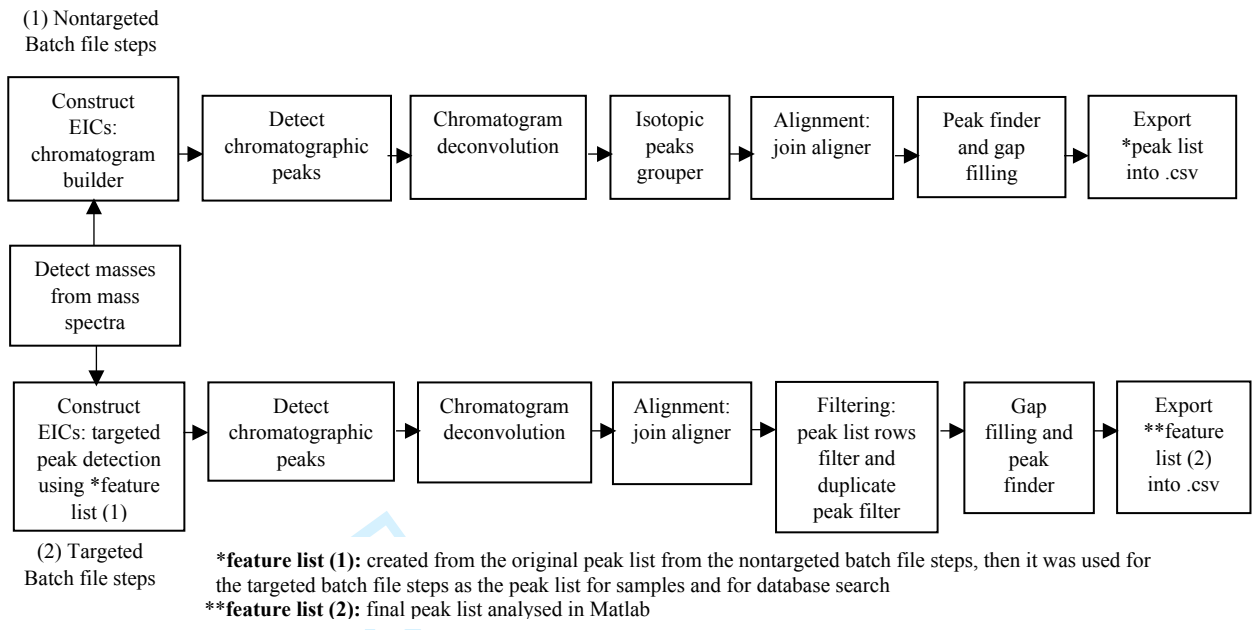


Fig. 1 Workflow of LC-MS feature extraction based on MZmine 2.0

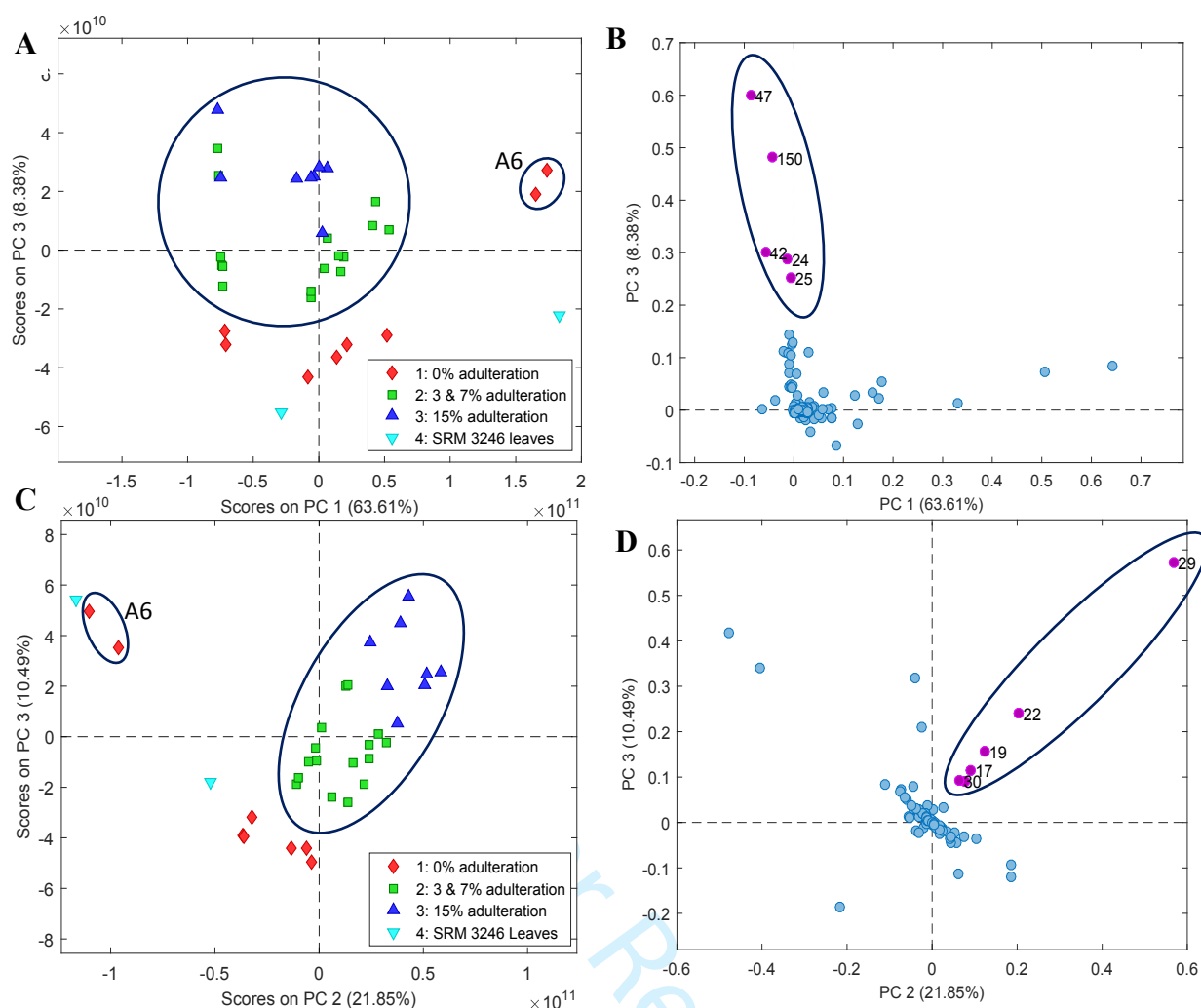


Fig. 2 PCA scores and loadings of plant materials (dataset A) using mean centering and classification by adulteration level: (A) score plot of dataset A positive ion mode, encircled samples: adulterated and A6 samples (B) loadings plot of dataset A positive ion mode, (C) score plot of dataset A negative ion mode, encircled samples: adulterated and A6 samples, and (D) loadings plot of dataset A negative ion mode; the encircled scores in 2A and 2C plots pertain to the adulterated samples and the encircled loadings in 2B and 2D plots are the variables correlated with the adulterated samples

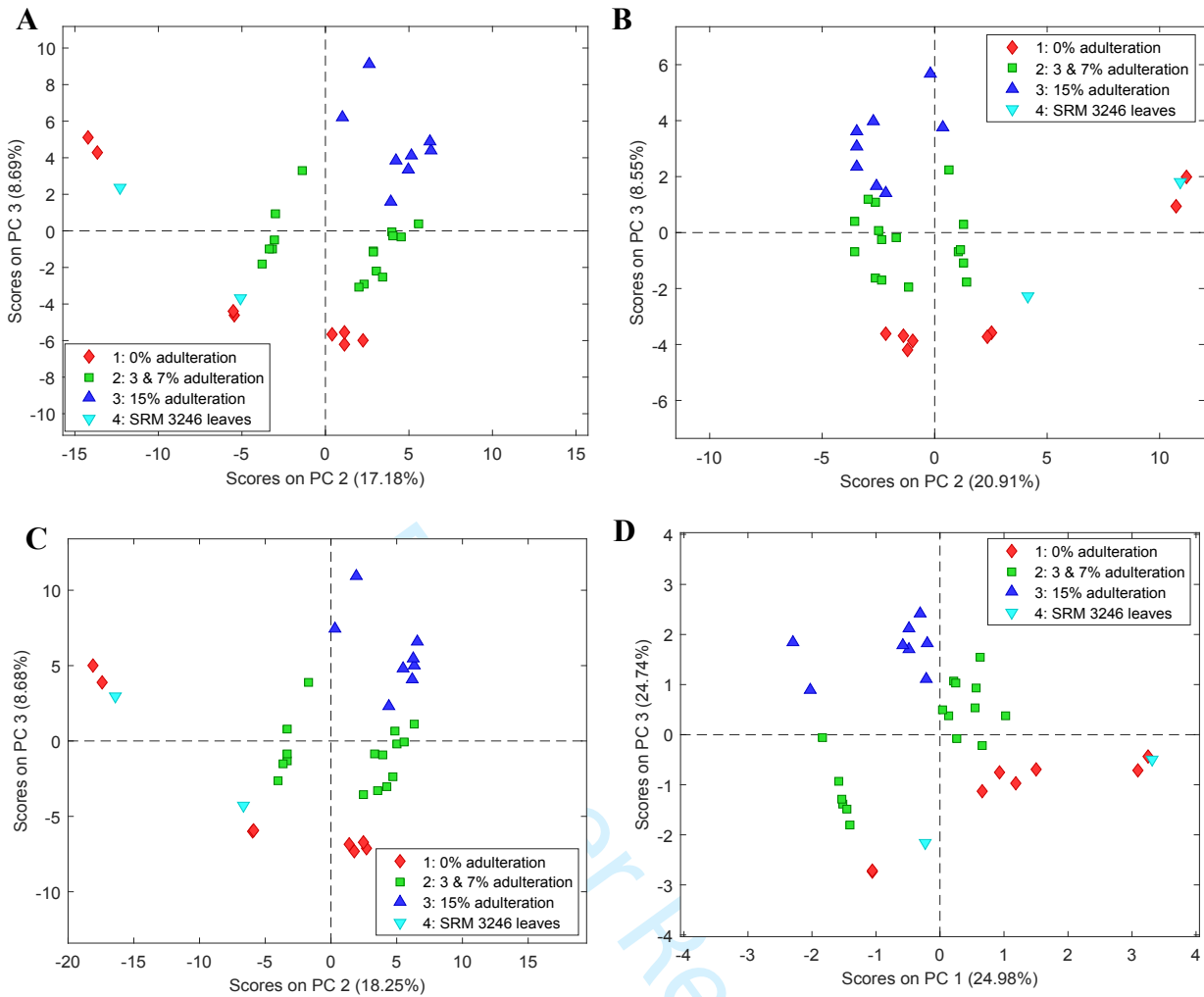


Fig. 3 PCA scores of plant materials (dataset A) and the combined dataset A results using autoscaling and classification by adulteration level: (A) score plot of dataset A positive ion mode (B) score plot of dataset A negative ion mode (C) Score plot of total dataset A (positive and negative ion modes) using the normalized data extracted from MZmine; and (D) score plot of total dataset A using the concatenated principal components

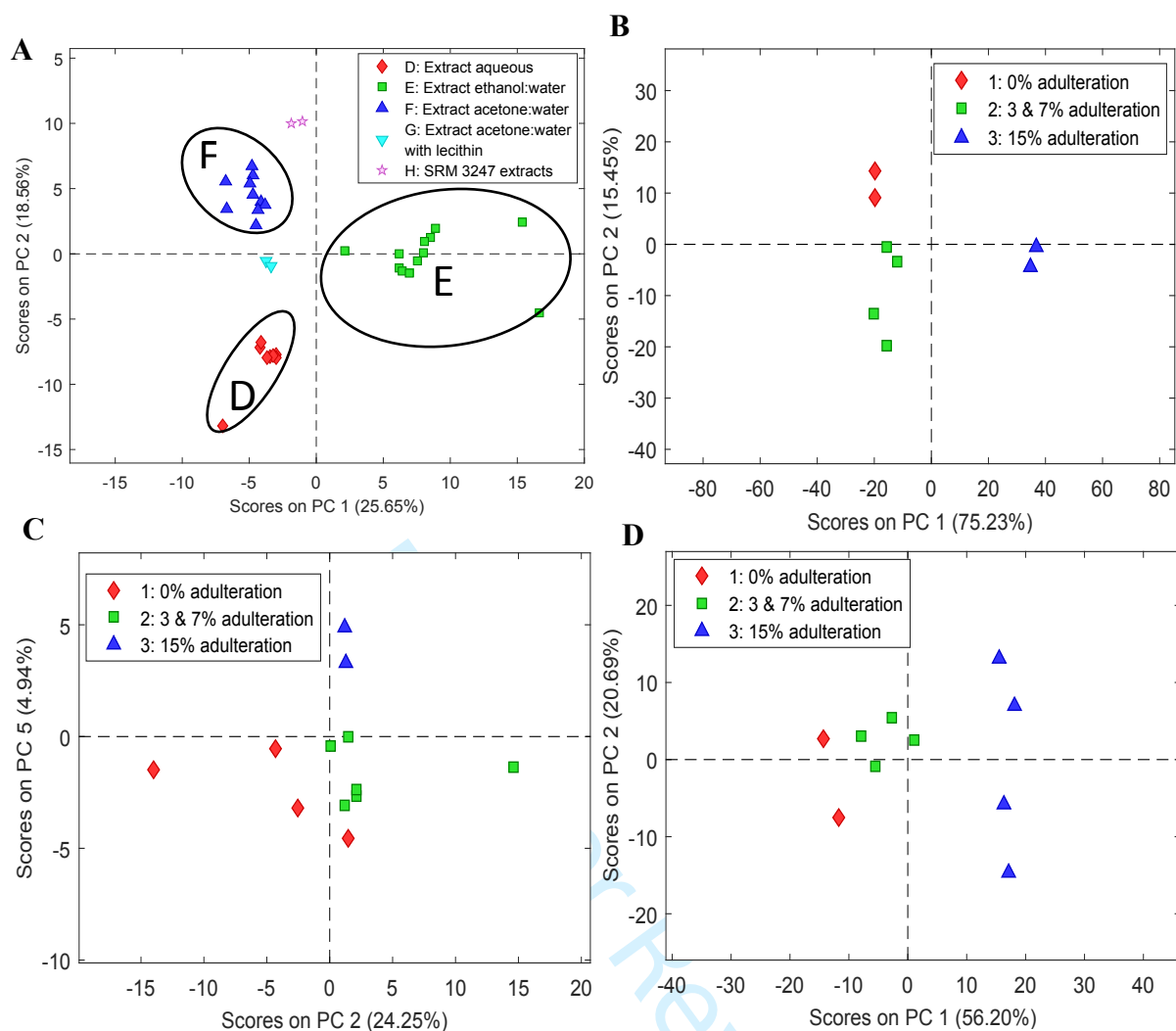


Fig. 4 PCA scores for dried leaf extracts (dataset B) using class centroid scaling and centering, and classification by adulteration level for each material source in positive ion mode: (A) score plot of whole dataset based on the material source, encircled groups represent the leaf extract types analyzed individually (B) score plot of the source D (extract aqueous) samples, (C) score plot of the source E (extract ethanol:water) samples and (D) score plot of the source F (extract acetone:water) samples

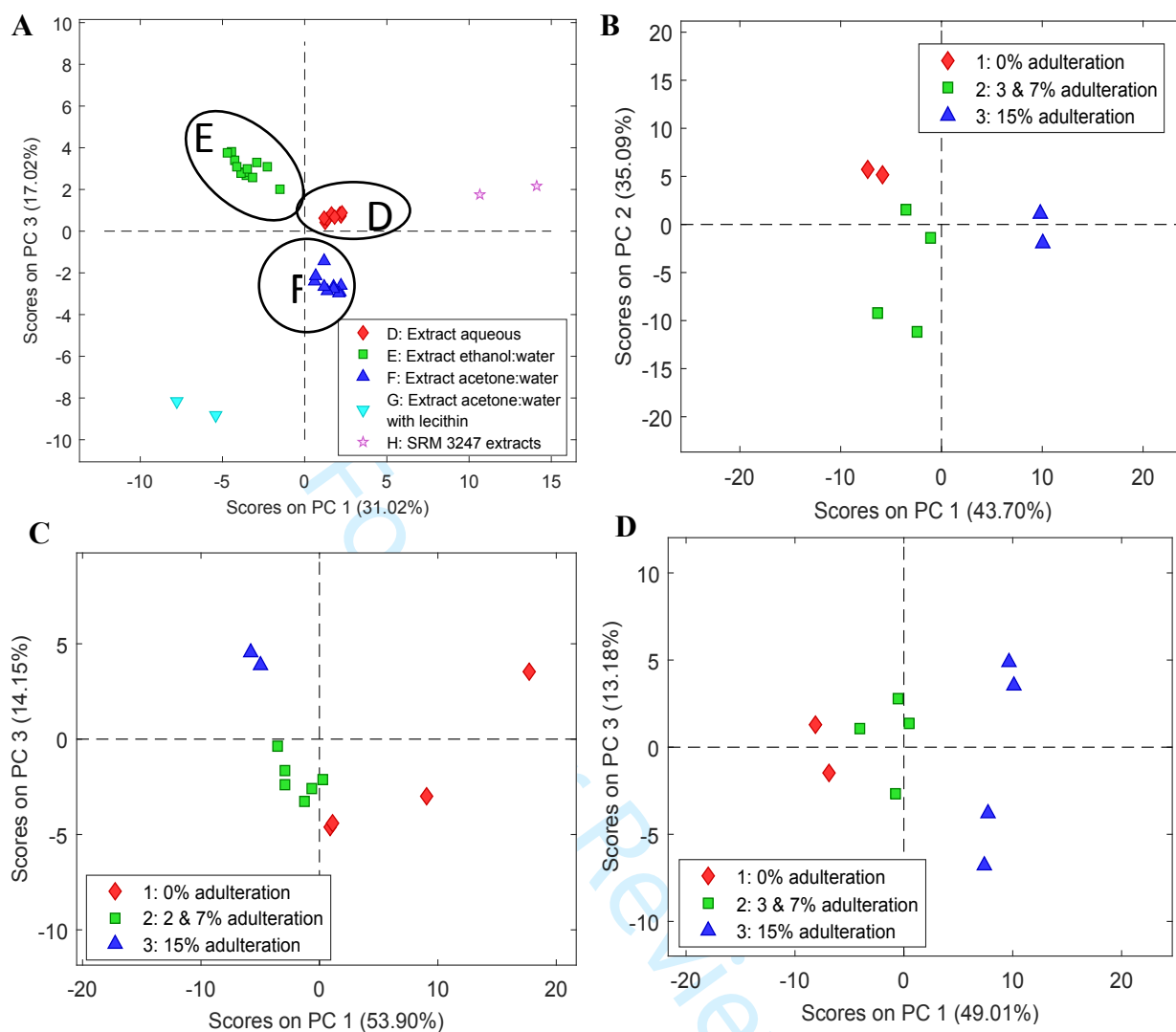


Fig. 5 PCA scores for dried leaf extracts (dataset B) using class centroid scaling and centering, and classification by adulteration level for each material source in negative ion mode: (A) score plot of whole dataset based on the material source, encircled groups represent the leaf extract types analyzed individually (B) score plot of the source D (extract aqueous) samples, (C) score plot of the source E (extract ethanol:water) samples, and (D) score plot of the source F (extract acetone:water) samples

Table 1 Summary of *Ginkgo biloba* L. sample classification used in the study

Source of material	Adulteration Level			
	1: 0%	2: 3%	2: 7%	3: 15%
Dataset A – Plant parts				
Source A: Ginkgo leaves				
A1 - untreated	A9	A3	A16	A12, A4
A2 – steam-treated	A5	A14	A7, A13	A8
Source B: Ginkgo stem	A15	A2, A10	A1	A11
Source C: SRM 3246 <i>Ginkgo biloba</i> L. leaves	A6, SRM 3246			
Dataset B – Dried leaf extracts				
Source D: Ginkgo extract aqueous	B10	B5	B13	B7
Source E: Ginkgo extract ethanol:water	B3, B8	B12, B16	B1	B9
Source F: Ginkgo extract acetone:water	B6	B14	B11	B4, B15
Source G: Ginkgo extract acetone:water with lecithin	B2			
Source H: SRM 3247 <i>Ginkgo biloba</i> L. extract	SRM 3247			

For Peer Review

Instrument Conditions

LC Conditions

Injection volume	5 μ L	
Flow Rate	0.3 mL/min	
Mobile Phase	A: 0.1% formic acid in water, B: 0.1% formic acid in acetonitrile (As 100% water for A, 100% acetonitrile for B, and 50:50 water:acetonitrile with 0.1% formic acid for C)	
Gradient	Time	%B
	0	5
	1	5
	15	95
25	95	
Equilibration time	10 min	
Column Temperature	25 $^{\circ}$ C	

MS Conditions

Ionization	Electrospray
Polarity Ionization	Positive/Negative (separately)
Voltage	3000(+)/2500(-)
Capillary Temperature	350 $^{\circ}$ C
Sheath Gas	35
Auxiliary Gas	10
Probe Heater Temperature	300 $^{\circ}$ C
MS1 Scan Range	100 – 1500 m/z
MS1 Resolution	70,000
MS1 AGC Target	3e6
MS1 Maximum IT	100 ms
MS2 Experiment:	TopN (5)
MS2 Resolution	17,500
MS2 AGC Target:	1e5
MS2 Maximum IT:	50 ms
Dynamic exclusion:	5 s
Collision Energy:	30 with 50% stepped NCE (15, 30, 45)

Table 2 Chromatographic parameters used in LC-HRMS for *Ginkgo biloba* L. samples

Table 3 Parameters used in MZmine for *Ginkgo biloba* LC-HRMS datasets A and B

MZmine Parameters	LC-MS1 setting
Mass Detection	
Noise level	*1x10 ⁷
Chromatogram Builder	
Type of scans	MS level 1
Minimum time span	0.15 min
Minimum height	1x10 ⁷
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10 ppm
Chromatogram Deconvolution	
Algorithm	Local minima search
Chromatographic threshold	10%
Minimum retention time range	0.1 Min
Minimum relative height	10%
Minimum absolute height	*1x10 ⁷
Minimum ratio of peak top/edge	1
Peak duration range	0-10 min
Isotopic Peaks Grouper	
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10.0 ppm
Retention time tolerance	0.1 absolute min
Maximum charge	1
Representative Isotope	Most intense
Join Aligner	
<i>m/z</i> tolerance	0.005 to 10.0 ppm
Weight for <i>m/z</i>	20
Retention time tolerance	0.1 absolute min
Weight for RT	20
Peak finder	
Intensity tolerance	100.0%
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10.0 ppm
Retention time tolerance	0.1 absolute min
Same RT and <i>m/z</i> range gap filler	
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10.0 ppm
Targeted Peak Detection	
Peak list file	Select Targeted peak list created
Intensity tolerance	100.0%
Noise level	*1x10 ⁷
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10.0 ppm
Retention time tolerance	0.1 absolute min
Peak List Rows Filter	
<i>m/z</i> range	0.0000 to 80.0000
Keep or Remove rows	Remove rows that match all criteria
Duplicate Peak Filter	
Algorithm	NEW AVERAGE
<i>m/z</i> tolerance	0.005 <i>m/z</i> or 10.0 ppm
RT Tolerance	0.1 absolute min

* dataset A both modes: 1x10⁷; dataset B negative ion mode: 5x10⁷; and dataset B positive ion mode: 2x10⁷

Table 4 Data matrices for plant material samples

Data matrix	samples x variables
A negative ion mode	34 x 77
A positive ion mode	34 x 175
Combined negative and positive ion modes of A (normalized results)	34 x 252
Combined negative and positive ion modes of A (principal component results)	34 x 8

For Peer Review

Ginkgo dried leaf extract samples	Process	Solvent Ratio	Native Extract Ratio	Excipients
Ginkgo extract aqueous	Water extraction; spray dry	Not specified	Not specified	Not specified
Ginkgo extract ethanol:water	Not specified	Ethanol (60-80%)/ Water (20-40%)	Not specified	Not specified
Ginkgo extract acetone:water	Not specified		35-67:1	Syrup, Corn, Dehydrated
Ginkgo extract acetone:water with lecithin	Not specified		35-67:1	Lecithin (origin: Soy)

Table 5 List of peaks identified for normalized unprocessed dataset A (plant materials)

Table 6 Solvent extraction ratios and preparations of dried leaf samples based on COA

Positive ion mode					
Loadings ID from PCA	<i>m/z</i>	RT	Compound Name	MF	Prob.
42	579.1694	6.8944	Isorhoifolin	718	97.35
47	433.1119	7.0029	Sophoricoside	658	96.98
24	611.1590	6.3266	Rutin	460	93.13
25	611.1589	6.3550	Rutin	460	93.13
150	282.2783	17.2641	Oleamide	872	92.49
Negative ion mode					
19	609.1468	6.3919	Luteolin-7,3'-di-o-glucoside	345	96.77
20	593.1521	6.7255	Tiliroside	801	94.80
29	269.0456	8.9310	Genistein	706	83.89
22	431.0987	7.0047	Sophoricoside	908	83.40
30	285.0406	9.0730	Kaempferol	742	74.47
17	755.2057	6.0245	-	-	-

1
2
3 **Analytical and Bioanalytical Chemistry**
4
5

6
7 **Electronic Supplementary Material**
8
9

10
11
12 **A nontargeted approach to determine the authenticity of *Ginkgo biloba* L.**
13 **plant materials and dried leaf extracts by liquid chromatography-high**
14 **resolution mass spectrometry (LC-HRMS) and chemometrics**
15
16
17
18

19
20 Meryl B. Cruz, Benjamin J. Place, Laura J. Wood, Aaron Urbas, Andrzej Wasik,
21
22 Werickson Fortunato de Carvalho Rocha
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Normalization used in LC-HRMS data: Reduction of Peak Area to Sample Concentration

By reducing the peak area to relative sample concentration, the variability from sample preparation can be incorporated into concentrations and the within-replicate variability can be minimized.

It is assumed that the mass spectrometer detector response is linearly related to the amount of substance introduced onto the instrument, therefore peak area (A_A) is directly proportional to the amount of mass delivered to the detector ($m_{A,det}$)

$$A_A \propto m_{A,det}$$

Based on a 5 μL injection volume (V_{inj}) and density of the injection solvent (ρ_{sol}) and assuming the density of the solution is equivalent to the injection solvent, we can calculate the relative concentration of the analyte in the solution ($C_{A,sol}$) in mass analyte per mass solution units

$$\frac{m_{A,det}}{V_{inj} \times \rho_{sol}} \propto C_{A,sol}$$

Using $C_{A,sol}$, we can calculate the mass of analyte extracted into the solution ($m_{A,sol}$) using the mass of the solution (m_{sol})

$$C_{A,sol} \times m_{sol} \propto m_{A,sol}$$

Using the mass of the analyte extracted into the solution and assuming 100 % extraction efficiency, we can calculate the relative concentration of the analytes in the sample ($C_{A,sample}$) using the mass of the sample from which the analytes were extracted (m_{sample})

$$\frac{m_{A,sol}}{m_{sample}} \propto C_{A,sample}$$

Therefore:

$$C_{A,sample} \propto \frac{A_A}{V_{inj} \times \rho_{sol}} \times \frac{m_{sol}}{m_{sample}}$$

We can reduce $\frac{1}{V_{inj} \times \rho_{sol}}$ to a constant B.

$$C_{A,sample} \propto B \times \frac{A_A \times m_{sol}}{m_{sample}}$$

And since the samples are of constant composition and the injection size is constant, for comparison reasons we can drop the B term:

$$C_{A,sample} \propto \frac{A_A \times m_{sol}}{m_{sample}}$$

So, for each column of peak areas for each individual sample (sample i), multiply the peak area counts by the mass of the solvent used to extract the samples and divide it by the mass of the solid sample used to extract:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

$$\begin{bmatrix} A_{i,1} \\ A_{i,2} \\ A_{i,3} \\ \vdots \\ \vdots \\ A_{i,n-1} \\ A_{i,n} \end{bmatrix} \times \frac{m_{sol}}{m_{sample}} = \begin{bmatrix} C_{i,1} \\ C_{i,2} \\ C_{i,3} \\ \vdots \\ \vdots \\ C_{i,n-1} \\ C_{i,n} \end{bmatrix}$$

For Peer Review

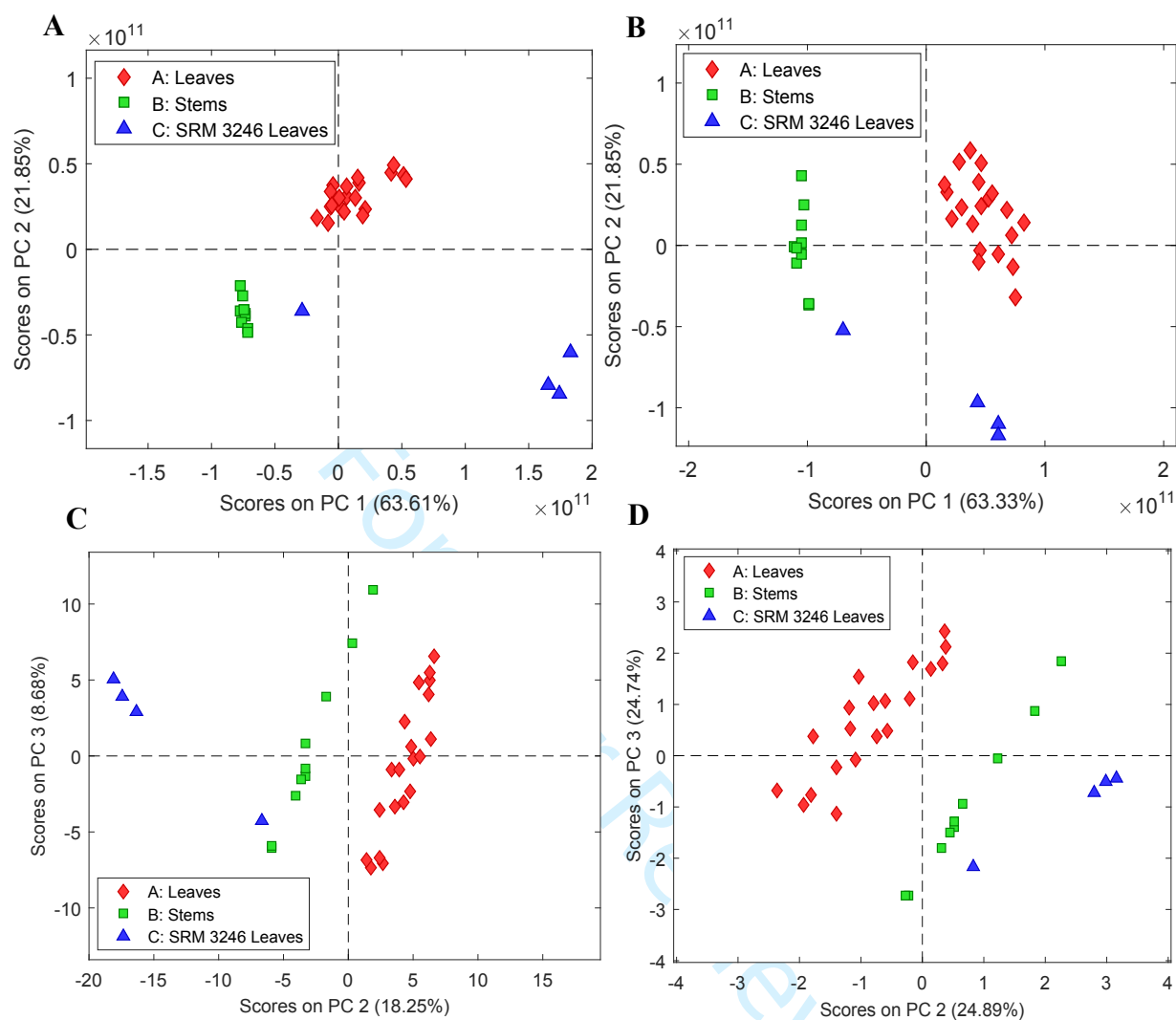


Fig 1 PCA scores of plant materials (dataset A) and the combined dataset A results using classification by material source: (A) score plot of dataset A positive ion mode using mean centering, (B) score plot of dataset A negative ion mode using mean centering, (C) score plot of total dataset A (positive and negative ion modes) using the normalized data extracted from MZmine using autoscaling; and (D) score plot of total dataset A using the concatenated principal components using autoscaling

1
2
3
4
5
6 **Analytical and Bioanalytical Chemistry**
7

8
9 **Electronic Supplementary Material**
10
11

12
13
14
15 **A nontargeted approach to determine the authenticity of *Ginkgo biloba* L.**
16 **plant materials and dried leaf extracts by liquid chromatography-high**
17 **resolution mass spectrometry (LC-HRMS) and chemometrics**
18
19
20

21
22 Meryl B. Cruz, Benjamin J. Place, Laura J. Wood, Aaron Urbas, Andrzej Wasik,
23
24 Werickson Fortunato de Carvalho Rocha
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Normalization used in LC-HRMS data: Reduction of Peak Area to Sample Concentration

By reducing the peak area to relative sample concentration, the variability from sample preparation can be incorporated into concentrations and the within-replicate variability can be minimized.

It is assumed that the mass spectrometer detector response is linearly related to the amount of substance introduced onto the instrument, therefore peak area (A_A) is directly proportional to the amount of mass delivered to the detector ($m_{A,det}$)

$$A_A \propto m_{A,det}$$

Based on a 5 μ L injection volume (V_{inj}) and density of the injection solvent (ρ_{sol}) and assuming the density of the solution is equivalent to the injection solvent, we can calculate the relative concentration of the analyte in the solution ($C_{A,sol}$) in mass analyte per mass solution units

$$\frac{m_{A,det}}{V_{inj} \times \rho_{sol}} \propto C_{A,sol}$$

Using $C_{A,sol}$, we can calculate the mass of analyte extracted into the solution ($m_{A,sol}$) using the mass of the solution (m_{sol})

$$C_{A,sol} \times m_{sol} \propto m_{A,sol}$$

Using the mass of the analyte extracted into the solution and assuming 100 % extraction efficiency, we can calculate the relative concentration of the analytes in the sample ($C_{A,sample}$) using the mass of the sample from which the analytes were extracted (m_{sample})

$$\frac{m_{A,sol}}{m_{sample}} \propto C_{A,sample}$$

Therefore:

$$C_{A,sample} \propto \frac{\frac{A_A}{V_{inj} \times \rho_{sol}} \times m_{sol}}{m_{sample}}$$

We can reduce $\frac{1}{V_{inj} \times \rho_{sol}}$ to a constant B.

$$C_{A,sample} \propto B \times \frac{A_A \times m_{sol}}{m_{sample}}$$

And since the samples are of constant composition and the injection size is constant, for comparison reasons we can drop the B term:

$$C_{A,sample} \propto \frac{A_A \times m_{sol}}{m_{sample}}$$

So, for each column of peak areas for each individual sample (sample i), multiply the peak area counts by the mass of the solvent used to extract the samples and divide it by the mass of the solid sample used to extract:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

$$\begin{bmatrix} A_{i,1} \\ A_{i,2} \\ A_{i,3} \\ \vdots \\ \vdots \\ A_{i,n-1} \\ A_{i,n} \end{bmatrix} \times \frac{m_{sol}}{m_{sample}} = \begin{bmatrix} C_{i,1} \\ C_{i,2} \\ C_{i,3} \\ \vdots \\ \vdots \\ C_{i,n-1} \\ C_{i,n} \end{bmatrix}$$

For Peer Review

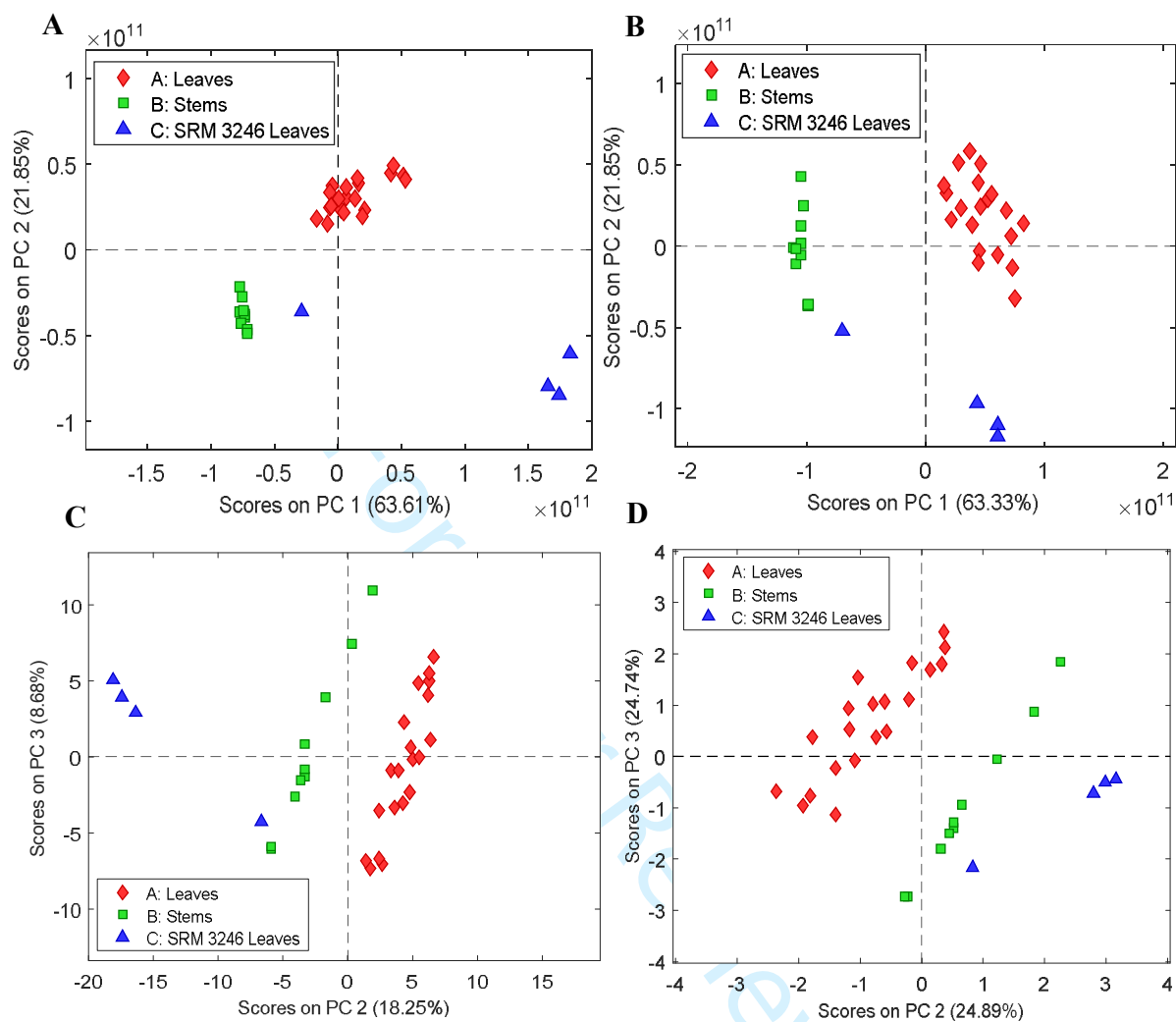


Fig 1 PCA scores of plant materials (dataset A) and the combined dataset A results using classification by material source: (A) score plot of dataset A positive ion mode using mean centering, (B) score plot of dataset A negative ion mode using mean centering, (C) score plot of total dataset A (positive and negative ion modes) using the normalized data extracted from MZmine using autoscaling; and (D) score plot of total dataset A using the concatenated principal components using autoscaling