Heliyon 8 (2022) e12044

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Classification of three types of ginseng samples based on ginsenoside profiles: appropriate data normalization improves the efficiency of multivariate analysis

Yahui Li [a], Bingkun Yang [a,b], Wei Guo [b], Panpan Zhang [a], Jianghua Zhang [a], Jing Zhao [a], Qiao Wang [b], Wei Zhang [a], Xiaowei Zhang [c,**,1], Dezhi Kong [a,*,1]

[a] School of Chinese Integrative Medicine, Hebei Medical University, Shijiazhuang, China
[b] School of Pharmacy, Hebei Medical University, Shijiazhuang, China
[c] The Second Hospital of Hebei Medical University, Shijiazhuang, China

ARTICLE INFO

Keywords:
Panax ginseng
Ginsenosides profiles
Data normalization
Multivariate analysis
High-resolution mass spectrometry

ABSTRACT

Background: It is well known that ginsenosides are the main active ingredients in ginseng, and they have also been important indexes for assessing the quality of ginseng. However, the absolute contents of ginsenosides in ginseng were shown to be varied with the origin, cultivated type, cultivated year and climate. It is a great challenge to distinguish the commercial types of ginsengs according to the content of one or several ginsenosides.

Methods: The common commercial types of ginsengs are white ginseng (WG), red ginseng (RG), American ginseng (AG). To clearly illustrate the differences among WG, RG and AG at the ginsenosides level, we established a strategy for the detection and identification of ginsenosides based on an optimized LC-Q-Orbitrap MS/MS method coupled with an in-house database of ginsenosides. Before and after the normalization, the ginsenosides datasheet was analyzed and compared using several state-of-the-art multivariate statistical analysis methods.

Results: Here, 81 ginsenosides were identified in different ginseng samples. The majority of the ginsenosides (59 in 81) were all shared by WG, RG and AG. When the shared ginsenosides datasheet was normalized by the level of ginsenoside Ro, our analysis strategy clearly divided the ginseng samples into three groups (i.e., WG, RG and AG groups). We found that the ginsenoside profiles in RG and WG were significantly different from those in AG. The potential markers and multivariate diagnostic models differentiating the three types of ginsengs were also indicated.

Conclusion: Our novel methodology based on ginsenoside profiles is more robust than existing methods, and data normalization is required to improve the efficiency of multivariate statistical analysis.

1. Introduction

Ginseng is mainly planted in China, Korea, and America, as a perennial plant belonging to the genus *Panax* of the *Araliaceae* family, and is widely used in over 35 countries throughout the world, reported in 2017 [1]. "Panax" as a botanical name means "heal everything" in Greek [2]. The history of using ginseng for its medical properties in China dates back approximately 5000 years [3]. It has been proven that ginseng exhibits antioxidant, anti-inflammatory, anti-aging, anticancer, anti-apoptotic, and immune-stimulatory pharmacological activities [2,4–8]. In recent years, ginseng has been widely sold as a dietary supplement and alternative medicine and has received great favor from users [9, 10].

The most common commercial types of ginseng are white ginseng (WG), red ginseng (RG) and American ginseng (AG), all under the *Panax* family. Each type of ginseng has slightly different uses [11], but all contain ginsenosides—steroidal saponins that contain the 4 trans-ring rigid steroid skeleton—and the type, number and location of their sugar moieties as their different [12]. It is well known that ginsenosides are the main active ingredients in ginseng, and they have also been important indexes for assessing the quality of ginseng [13].

* Corresponding author.
** Corresponding author.
E-mail addresses: Xiaowei_zhang@163.com (X. Zhang), kongdezhi@hebmu.edu.cn (D. Kong).
[1] Present address: No. 361, Zhongshan East Road, Chang'an District, Shijiazhuang, Hebei Province, China.

https://doi.org/10.1016/j.heliyon.2022.e12044
Received 24 June 2022; Received in revised form 20 October 2022; Accepted 24 November 2022

Some reports have shown that AG has little ginsenoside Rf and has a low ratio of ginsenoside Rg1 to Rb1 compared to WG and RG, while ginsenoside F11 is exclusively found in AG. The Rf/F11 ratio can be used as a phytochemical marker to distinguish AG among the three types of ginseng [14]. Red ginseng, which is heat-processed WG (steamed at 98–100 °C for 2–3 h), was found to have decreased contents of some common ginsenosides (Rb1, Rc, Rd, etc.) but contained an array of rare ginsenosides including Rg5, Rk1, Rk2, etc. [15]. Compared to WG, RG was thought to have an increased anti-cancer property [2]. However, the detailed ginsenoside profiles among the three types of ginseng have not been reported.

Additionally, the absolute contents of different ginsenosides in ginseng were shown to be varied with the origin, cultivated type (garden ginseng, forest ginseng or transplanted wild ginseng), cultivated year and climate [16, 17, 18]. Some ginsenoside concentrations varied by 15-fold (0.288–4.266% by *wt*) in *Panax* powders [19]. Therefore, it is a great challenge to distinguish WG, RG, and AG based on the absolute content of ginsenosides.

Many analytical ways have been developed to quantify ginsenosides, including TLC [20], HPLC coupled with a UV detector or an evaporative light scattering detector (ELSD) [21], and LC-MS [15, 22]. Ginsenosides have such complex properties, such as diversity, similarity and complexity, that their analysis is a formidable challenge. Nowadays, the best option for the simultaneous and accurate quantification of multiple ginsenosides in complex matrices is liquid chromatography coupled with electrospray tandem mass spectrometry (LC-ESI-MS/MS).

In the paper, the ginsenosides in AG, WG and RG were analyzed using an LC-ESI-Orbitrap Fusion system and identified based the strategies shown in Figure 1. To classify the three types of ginsengs based on the ginsenosides profiles, we introduced multivariate statistical techniques, and proposed an appropriate data normalization method to improve the efficiency of multivariate statistical analysis. The potential markers and multivariate diagnostic models differentiating the three types of ginsengs were also indicated by the principal component analysis (PCA) and logistic regression analysis.

2. Material and methods

2.1. Materials and reagents

Authentic standards of 15 ginsenosides—Rb1, Rb2, Rb3, Rc, Rd, Re, F1, F2, Ra3, Rg1, Rg3, Rg5, Rh1, Ro, and F11—were purchased from Chengdu Mansite Biotechnology Co., Ltd (Chengdu, China). These ginsenosides were determined to have a purity >98% by LC-UV, and their chemical structures are given in Figure S1. LC/MS-grade methanol (MeOH) and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Waltham, MA, USA). HPLC-grade ammonium acetate was obtained from Sigma-Aldrich, Co (St. Louis, MO, USA), and deionized water was prepared by a Thermo Nanopure water purification system (Waltham, MA, USA).

All ginseng samples were purchased from different herbal manufacturing enterprises in China, and the sample information is listed in Supplemental Table 1.

2.2. Ginsenosides database for screening

More than 100 ginsenosides have been identified since their first description in the 1960s by Shibata's group [23]. We used ginsenoside as a keyword to search in the PubChem Compound database (https://www.ncbi.nlm.nih.gov/pccompound/) and found 161 records containing chemical structure information. Then, we established a new database containing 161 ginsenosides according to structure information originating from reference compounds and PubChem records using TraceFinder software (Thermo Fisher Scientific Inc. version 2.0).

2.3. Sample preparation

All ginseng samples were pulverized into powder of over 40 meshes, then the fine ginseng powder was accurately weighed (1.000 g) and extracted with 10 ml of 70% methanol (methanol: water, 70: 30, *v/v*) in an ultrasonic bath at room temperature for 30 min. The solution was centrifuged at 15,000 g for 10 min, after which the supernatant was passed through a 0.22-μm filter, and the process was subsequently followed by analysis through LC-MS/MS.

The standard stock solutions of ginsenosides were prepared independently by dissolving 2 mg of the standard into 10 ml of aqueous solution in which the proportion of methanol was approximately 70% (methanol:water, 70:30, *v/v*) to achieve a concentration of 0.2 mg/ml. The preparation of mixed standard solutions was performed through combining aliquots of each set of individual stock solutions and diluting them to the appropriate concentration. The solutions were filtered through a 0.22-μm syringe filter prior to LC-MS/MS analysis. All the solutions mentioned above were kept at 4 °C for storage and restored to room temperature before utilization.

2.4. Instruments and chromatographic conditions

The LC-MS/MS system consisted of a Thermo UltiMate® 3000 liquid chromatography system and an Orbitrap Fusion mass spectrometer equipped with a heated electrospray ionization source. Data acquisition was performed using Thermo TraceFinder Ver. 2.0 software. Chromatographic separation was achieved on a Waters T3 column (3.0 μm, 100 × 3.0 mm).

The mobile phase consisted of water/0.01% acetic acid (A) and acetonitrile (B), and the flow rate was set to 0.35 ml/min. A linear gradient was used, starting with 30% B. This proportion was held constant for 1 min and then increased linearly as follows: to 45% from 2 min to 6 min and then to 85% from 6 min to 8 min. The gradient was held constant at 85% until 12 min and then returned to the initial composition and again held constant for 4 min to re-equilibrate the column. The

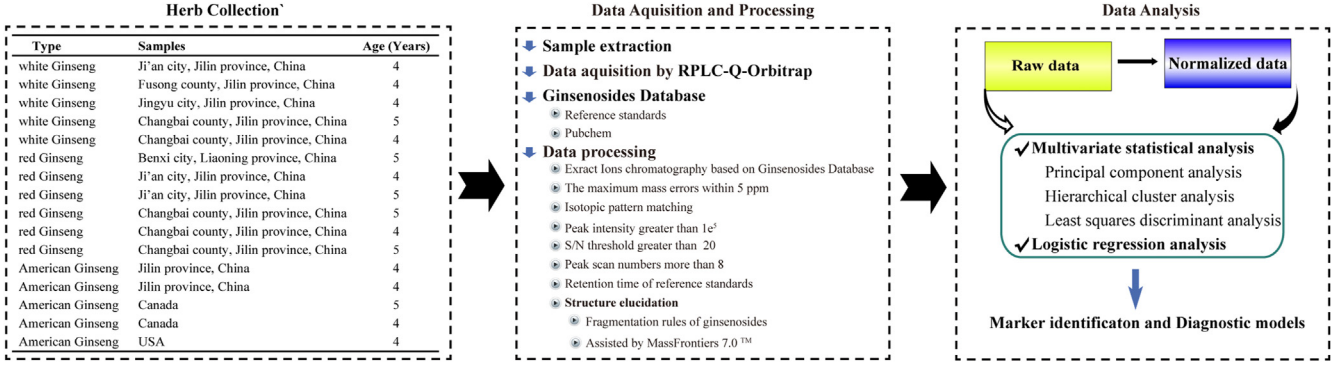

Figure 1. Schematic of the experimental workflow and the strategy to identify ginsenosides in Ginseng.

column and sample managers were maintained at 30 °C and 10 °C, respectively. An injection volume of 5 μl was used for the reference standards and samples.

For mass detection, the electrospray ionization source was operated in negative mode. The MS acquisition was performed in data-dependent acquisition (DDA) mode from 550 to 1750 Da. The mass resolution was set at 120,000 for the screening method and at 30,000 for DDA acquisition (for structural identification). The mass isolation for the DDA acquisition was set at 2 Da. The MS source parameters were set as follows: heater temperature, 335 °C; capillary temperature, 335 °C; source voltage, 3.5 kV; sheath gas flow, 42 Arb; and auxiliary gas flow, 12 Arb. The collision energy was set at 30%.

2.5. Validation of the LC-MS/MS method

The precision of the LC-MS/MS method was evaluated using three real samples at 0, 12 and 24 h for examination of the areas of 20 representative ginsenosides. To assess the recovery, the calibration curves were constructed using least-squares regression by plotting the peak areas of 6 standards (Rg1, Rb2, Rc, Rd, Re and Rf) against the concentrations. The contents of the 6 representative ginsenosides were recorded before and after spiked reference standards, and then calculated according to the following formula.

Recovery (%) = (measurement after spiking − measurement of non-spiked)/theoretical spiked amount.

2.6. Data processing

2.6.1. Ginsenosides confirmation

TraceFinder (Thermo Fisher Scientific Inc.) was used for extracting peaks based on the ginsenosides database. The strategies for confirming non-reference ginsenosides are shown in Figure 1. The filter used selected the peaks that met the following conditions for further processing: precursor mass within a 5-ppm mass tolerance window, a signal-to-noise ratio (S/N) threshold larger than 20, isotopic confirmation used, and a scan number of each peak greater than 8. The negative MS/MS spectra obtained from the deprotonated molecular $[M-H]^-$ ions were used to confirm the ginsenosides according to MS/MS (fragment ion) information from reference standards and literatures. Components of different samples appeared the same when they showed similar retention times with a tolerance of 0.15 min, accurate mass weights with a tolerance of 0.05 Da, and ion fragments. Peak integration was calculated automatically and supplemented manually.

2.6.2. Multivariate statistical analysis

All ginsenosides were recorded in a data matrix according to their names and peak areas. The contents of the reference ginsenoside could be calculated using the known concentrations of the standards. However, the unknown ginsenoside concentrations could not be estimated without a standard. Thus, for the consistency of multivariate analysis, all known and unknown markers were analyzed based on variations in their peak areas, which are directly proportional to their concentrations. Imputation was performed to estimate missing values using the Mice packages of the R-language program (https://www.r-project.org/) based on the random forest algorithm.

Venn diagrams were created and visualized using the R-language program with VennDiagram and ggplot2 packages. Hierarchical cluster analysis (HCA), partial least-squares discriminant analysis (PLS-DA) and principal component analysis (PCA) were also performed using the pheatmap and muma packages of the R-language program.

2.6.3. Logistic regression analysis

We used IBS SPSS Statistics 21 to conduct logistic regression analyses. The dependent variables were reported 59 metabolites, Chromatographic peak area. Conditional multiple regression calculation analyses for the 59 biomarkers were performed. And then, random selection the test results of these five biomarkers were set as dependent variable, conditional binary logistic regression analyses for the five biomarkers were performed.

3. Results and discussion

3.1. Optimization and validation of the quantitative analytical method

In the negative ion mode, ginsenosides had higher sensitivity and clearer mass spectra, so data collected in this mode were used for component detection and characterization [24, 25]. The composition of the mobile phase was investigated for improving the analyte ionization. Most of the standards contained abundant deprotonated molecular ions when the mobile phase consisted of acetonitrile and water. The water-phase additive acetic acid not only improved the LC separation but also helped to form $[M + CH_3COO]^-$ ions which were helpful to identify the precursor ions of the ginsenosides. After different concentrations of acetic acid were investigated, water containing 0.01% acetic acid was used as the optimal mobile phase.

Twenty ginsenoside peak areas were measured at 0, 12 and 24 h in real samples for the precision test. The RSD values for retention times and peak areas were less than 3.2% and 9.1%, respectively (Figure 2A). We tested the repeatability of the assay by extracting and analyzing five replicate samples, and we found that the RSD values of the areas for the 20 ginsenosides were within 5.5%.

For evaluating the method's accuracy, six ginsenoside standards were added into a sample, and then the mixture was extracted and analyzed. Our established method was with a good average recovery of 92.3%, and RSD values of 5.1–10.2% (Supplemental Table 2). Therefore, the LC-MS method was proven to be precise and accurate for quantifying ginsenosides in ginseng samples.

3.2. Identity assignment and confirmation of the ginsenosides

By using reference ginsenosides, we were able to optimized the mass chromatography conditions and obtained fragmentation pathways of ginsenosides [24]. From the MS scans of reference standards, the usual precursor ions of ginsenosides were $[M-H]^-$ and $[M + CH_3COO]^-$. Negative MS/MS spectra of the product ion $[M-H]^-$ exhibited a successive loss of the glycosidic units until the formation of $[aglycone-H]^-$ ions. Based on the neutral loss, it was easy to elucidate the sugar unit moiety according to a mass difference of 162.0547 Da indicating the presence of a glucosyl (Glc) group, of 132.0431 Da indicating the presence of a pentosyl group [arabinose (Ara) or xylose (Xyl)], of 146.0421 Da indicating the presence of a rhamnosyl (Rha) group, and of 176.0340 Da indicating the presence of a glucuronyl (Glu A) group. Figure 2B shows a representative example illustrating the fragmentation pathways of Rg3, F1 and Ro.

The first mass spectrometry data from ginsenoside Rg3 produced the analytical result of $[M - H]^-$ at m/z 783.4749 and the adduct ion $[M + CH_3COO]^-$ at m/z 829.5981, indicating that the molecular formula was $C_{42}H_{72}O_{13}$. Its characteristic MS/MS pattern contained the fragment ion m/z 459.3858, which indicated that this chemical compound belonged to the protopanaxadiol (PPD) group. The corresponding fragment ion originated from the break of the glycosidic bond, which produces peaks at m/z 621.4400 for $[M-H-Glc]^-$, m/z 459.3859 for $[M-H-2Glc]^-$, m/z 179.0565 for $[Glc-H]^-$, and m/z 161.0457 for $[Glc-H_2O]^-$, with results shown in Figure 2B.

The ginsenosides were identified and confirmed by the strategies shown in Figure 1, and all the possible ginsenosides are summarized in Supplemental Table 3. Fifteen components were unambiguously authenticated as ginsenosides Rb1,2 and 3, Rc, Rd, Re, F1,2 and 11, Ra3, Rg1,2 and 3 and 5, Rh1 and Ro by comparing the retention times, m/z values and fragment ions with those of the reference compounds.

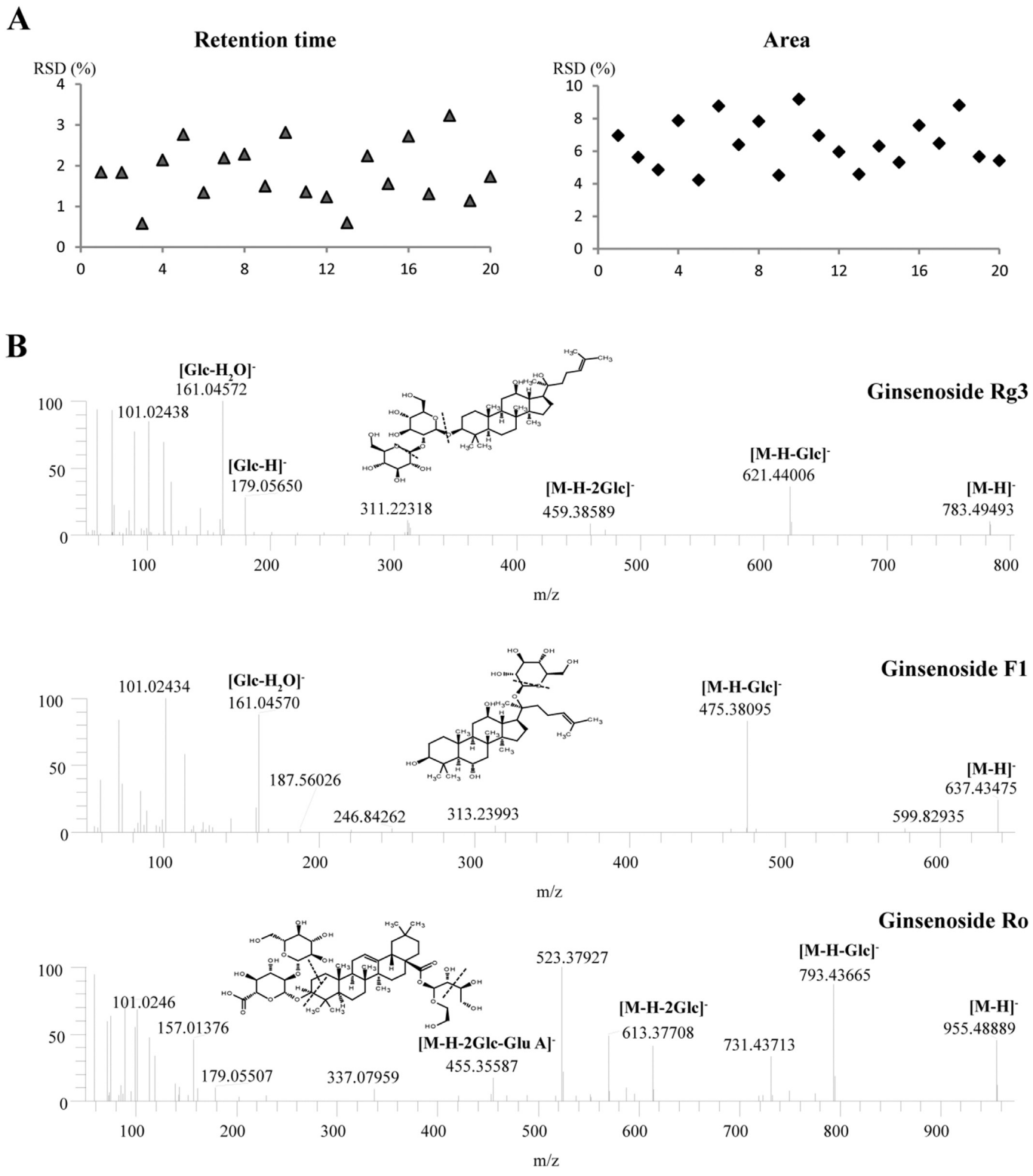

Figure 2. The precision was evaluated using the retention time and areas of 20 representative ginsenosides (A) and the ESI-Q-Orbitrap-MS/MS spectra (B) of ginsenoside Rg3, F1 and Ro.

The other components were tentatively identified by analyzing the accurate mass, isotopic ratio patterns and specific MS/MS fragment ions based on published data from known ginsenosides [26, 27]. It should be noted that isomers which had the same aglycone and sugar moiety while exhibiting the same fragmentation pathway could not be unambiguously identified.

3.3. Ginsenosides in the three types of ginsengs

In our study, there were 81 identified ginsenosides in the three types of ginsengs (Supplemental Table 3). As illustrated in Figures 3A, 2, 2 and 6 ginsenosides were only found in WG, RG and AG, respectively. As shown in Figure 3B, ginsenoside Rs5 ($C_{44}H_{72}O_{13}$, RT 6.6 min), for

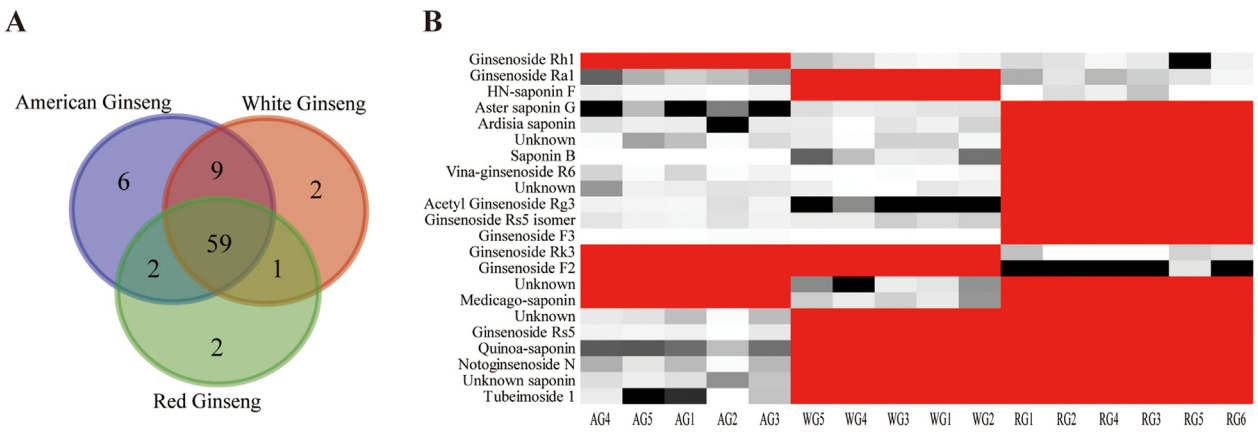

Figure 3. Venn diagram (A) depicting the relationship of ginsenosides in three ginsengs. Heatmap (B) shows 22 ginsenosides not shared by the three types of Ginseng. White to black indicates the increasing content of compounds. Red indicate the compound was not detected.

example, was only presented in AG. Ginsenoside F2 ($C_{42}H_{72}O_{13}$, RT 5.4 min) and ginsenoside Rk3 ($C_{36}H_{60}O_8$, RT 9.9 min) were only found in RG. However, more than two-thirds of the ginsenosides (59 in 81) were shared by WG, RG and AG. The results indicated that little difference was present in the ginsenoside compositions. In following analysis, we focused on the 59 shared ginsenosides to find the differences among WG, RG and AG.

3.4. Multivariate statistical analysis of the shared ginsenosides

The following multivariate statistical analysis was based on the 16 × 59 data matrix (samples × ginsenosides). Figure 4A shows the hierarchical cluster analysis (HCA) results for sample clustering based on all 59 ginsenosides. Because relative distances are proportional to sample correlation, a smaller relative distance indicates greater sample similarity

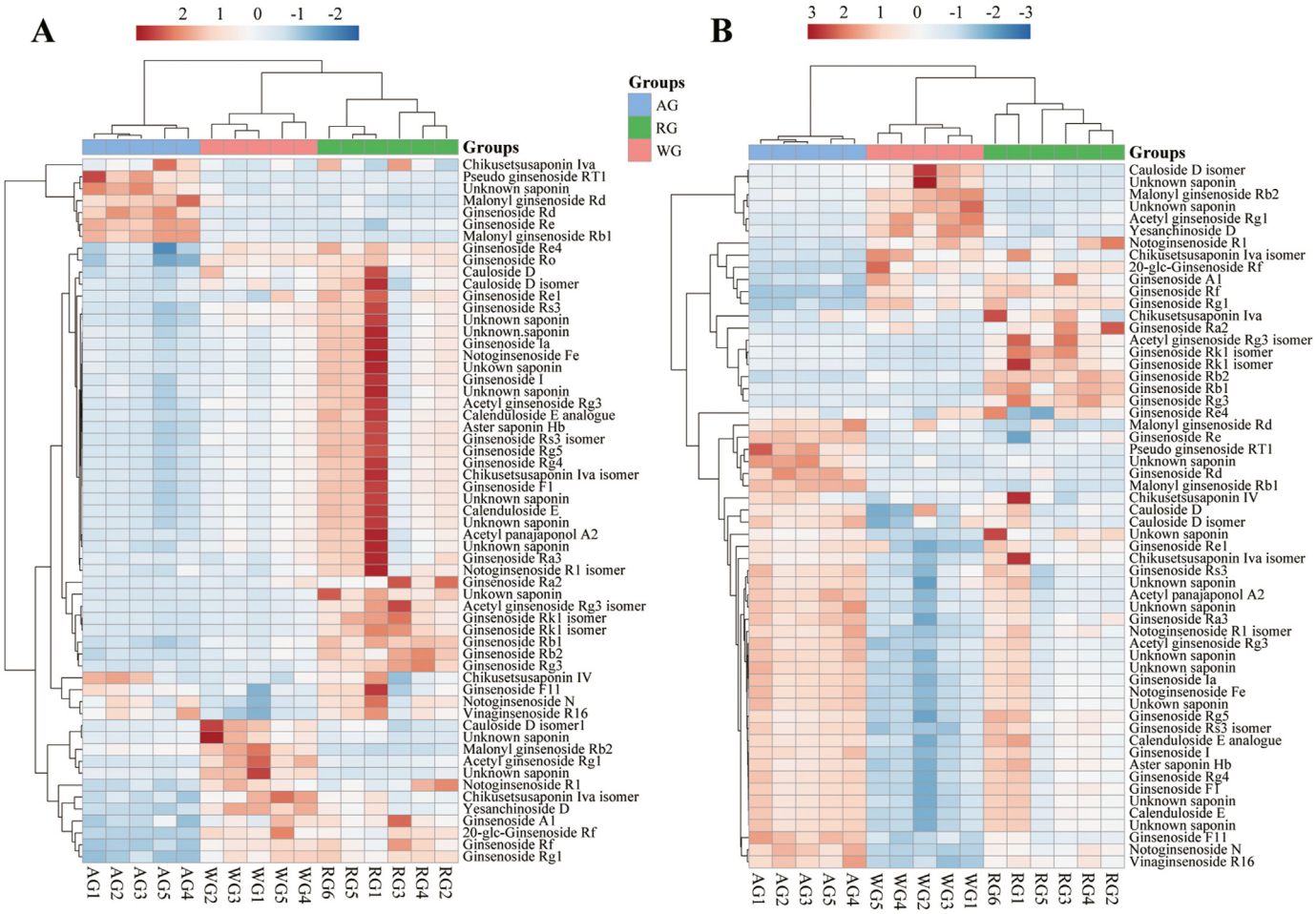

Figure 4. Heatmap depicting using (A) or (B). Each row, a compound; each column, a sample. The dendrograms on the left and at the top of the heatmap represent: correlations of compounds for individual cases; correlations of samples. Blue to red indicates the increasing.

than a pair with a larger distance [28]. Two major branches separate the 16 samples into two groups. The first branch (Group 1/3) included samples from AG, and the second branch (Group 2/3) included samples collected from RG and WG. This division of samples indicated that the AG samples were significantly different from the RG and WG samples. Moreover, with decreasing relative distances, the samples in the second branch were further sub-divided into two groups that corresponded to RG and WG.

Since we focus on the difference among the AG, RG, WG samples, we selected the ginsengs from different origins to fully consider the effects of various factors such as ginseng cultivation type, cultivation year and growth climate on the ginsenoside content. Our results also found the content of ginsenosides varied greatly among the samples. For example, the content of Rg1 in their highest concentrated samples was 3.2-fold higher and of Rb1 was nearly 10-fold higher than the lowest concentrated samples (data not shown). Therefore, it is a challenge to directly distinguish among AG, RG, WG by the absolute content of each ginsenoside.

To improve the efficiency of multivariate statistical analysis, we tried to normalize the ginsenosides matrix. Because the peak areas of ginsenoside Ro among the AG, RG, WG samples were without statistically different (Figure S2), the peak area of each sample was normalized by the peak area of ginsenoside Ro. Figure 4B shows the HCA for sample clustering after the normalization. Figure 4B more clearly shows the difference among the AG, RG and WG. The least squares discriminant analysis (PLS-DA) also proved that the normalized data more easily yielded satisfactory categorization of the samples (Figure S3). PLS-DA provided a 100% success rate in the prediction ability in terms of variety. The results indicated that the 59 ginsenosides in the ginseng samples could be used as indicators for determination of the ginseng variety, and data normalization was necessary for the classification.

Although these groupings are helpful for qualitative interpretation, a limitation of HCA is that the phylogenetic trees cannot be used to determine which markers cause major differences between samples [29, 30]. For a more thorough interpretation of the datasets, a principal component analysis (PCA) was utilized, which has an advantage over HCA alone [31].

Figure 5A and B are the PCA score plots generated based on the peak areas of all 59 ginsenosides without or with normalization. It is clearly to see that the first principal component (PC1) can explain the maximum variance in the data, and the second principal component (PC2) represents the maximum amount of variance in the other direction [32, 33]. PC1 and PC2 described 41% and 30% of the total variability in the original observations, respectively, and together they accounted for 71% of the total variance. In Supplemental Table 4, the loading of variables showed that ginsenosides Ra3, F11 and Re primarily formed PC1. PC2 was related to malonyl ginsenoside Rb2, acetyl ginsenoside Rg1, ginsenoside Rg4, ginsenoside Rs3, etc. PC3 was not prominent, as it only explained 8% of the total variance, and its inclusion would provide little additional information. As has been shown in other PCA studies with large labelled datasets [34, 35], Similar to the results of the HCA, the distances between samples in the PCA score plot were proportional to the similarities/differences between samples [29]. Since PC1 (41%) explained the most of the total variance, the same distance value along the PC1 axis indicated the greatest difference between samples. Therefore, AG which were vertically separated were more distinct to the horizontally separated WG and RG clusters samples. The PCA clustering (Figure 5) was highly consistent with the previous HCA clustering (Figure 4), which further validates the statistical results [32]. Therefore, the high consistency between HCA, which was generated based on 100% of the original variance, and PCA clustering which accounted for 71% of the total variance, indicated that the PC1 and PC2 were sufficient to provide a trustworthy linear relationship model, and was further validate the advantage of PCA which study the significant markers [31].

Biplots were created by combining PCA loading plots (Figure S4) with score plots to account for correlations between sample groups and individual markers. In Figure S4, every detected single point represents one ginsenoside, loading values plotted on the PC1 and PC2 axis. The differences and/or similarities among the markers were shown in the score scatter plot [36]. Therefore, a key contribution loading value of 1.0 was chosen to distinguish significant and non-significant markers for further analysis.

For the AG samples, most of the markers were found in Q2 of the loading plot (Figure S4) which corresponded with the AG sample cluster in the score plot (Figure 5A). Thirty-one markers, including their names and m/z values, are listed in Supplemental Table 5. In our study, it was found that American ginseng contained little ginsenoside Rf and higher levels of ginsenosides F11, Re and Rd. These results were consistent with those of previous reports [14] and proved that our data analysis processing was robust and reliable. These distinctive ginsenosides are related to the therapeutic implication of AG for neurodegenerative diseases associated with neuroinflammation [37].

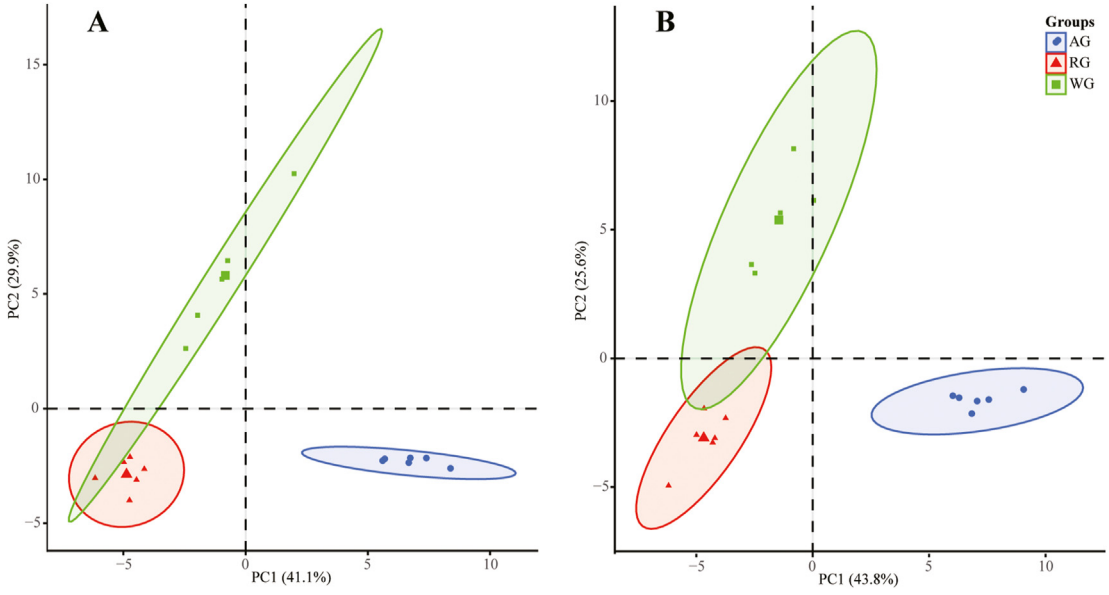

Figure 5. The principal component analysis (PCA) showed that 16 samples can be divided into three groups without (A) or with (B) normalization. AG, American Ginseng; RG, red Ginseng; WG, white Ginseng.

For the WG samples, most of the markers were found in Q4 of the loading plot (Figure S4), and 13 of these markers (including ginsenosides Rg1, Rb2, acetyl ginsenoside Rg1, etc.) are listed in Supplemental Table 5. Higher levels of ginsenosides Rg1 and Rb2 were found in WG samples (Figure 6). The level of ginsenoside Rg1, which has pharmacological use through producing weak stimulation to the central nervous system, indicated that WG is "warmer" than AG [19].

For the RG samples, most of the markers were found in Q3 of the loading plot (Figure S4), and 17 of these markers (including ginsenosides Rg3, Rg5, Rs3 and malonyl ginsenosides Rb1, Rb2, etc.) are listed in Supplemental Table 5. It was observed that the content of ginsenoside Rg3 was the highest in RG among the three types of ginseng (Figure 6). Compared with American ginseng and white ginseng, the content of ginsenoside Rg3 was approximately 3-fold in red ginseng (Figure 6). This was consistent with previous reporting that the amounts of ginsenosides Rg3 and Rg5 increased after the hot steaming process [38]. Compared with Asian white ginseng, red ginseng has stronger anticancer activities [2] due to the changes in these ginsenosides.

Moreover, we could use the ratios of some ginsenosides to easily illustrate the differences among AG, RG and WG. For example, we determined the contents of ginsenosides Re, Rg1, Rg3 and Ro, then calculated the ratios of Re/Ro, Rg1/Ro and Rg3/Ro. The maximum values of Re/Ro (0.600), Rg1/Ro (0.033) and Rg3/Ro (0.046) were obtained in AG, WG and RG, respectively (Figure S5). Based on our multivariate analysis results, other ratios of ginsenosides could also be suitable to distinguish the three types of ginseng.

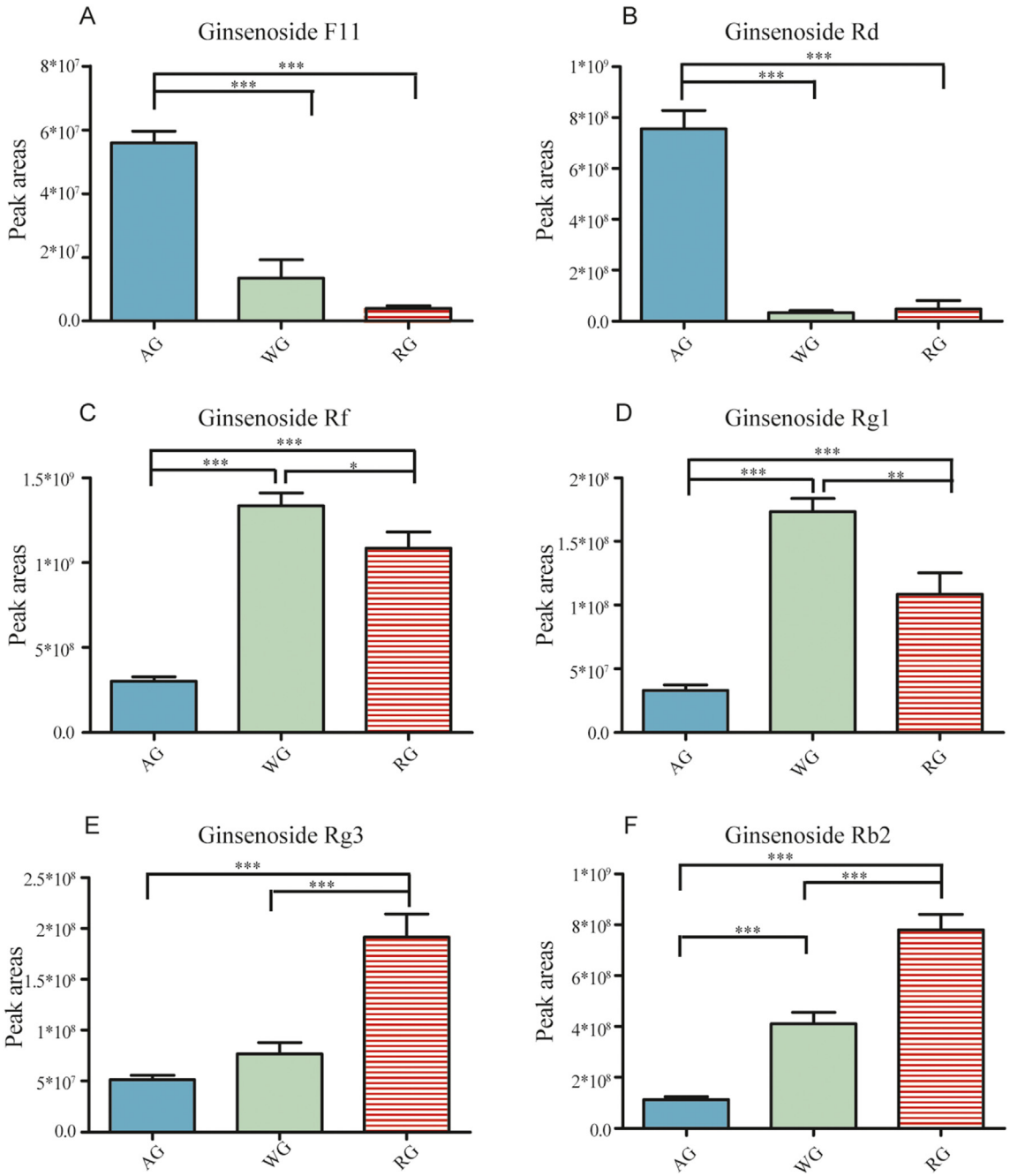

Figure 6. Based on the PCA loading plot, the representative ginsenoside makers for American Ginseng, red Ginseng and white Ginseng. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

3.5. Logistic regression analysis of the shared ginsenosides

Extracting principal components (PCs) by directly projecting the data using transformation matrices results in incorrectly mapped samples to their true locations in the low-dimensional feature subspace if some elements of the samples are perturbed [39]. Due to this weakness of PCA, logistic regression was introduced in this experiment for model classification to further increase the accuracy of the results.

Using the 59 metabolites shared by the three types of ginseng and the corresponding peak areas as independent variables, the logistic regression was carried out and the results were shown in Supplemental Table 6. Taking WG as the control, after performing multiple regression calculation, the classification equation of WG and AG is obtained:

$$\text{Species} = \text{-0.895} + \sum(\beta \times \text{Area})$$

where β is the value corresponding to each marker; Area is the corresponding value of each marker chromatographic peak area.

Taking WG as the control, the classification equation of WG and RG is obtained:

$$\text{Species} = -1.719 + \sum(\beta \times \text{Area})$$

where β is the value corresponding to each marker; Area is the chromatographic peak area corresponding to each marker.

Five ginsenosides were randomly selected as independent variables, and perform binary logistic regression. After regression analysis, the AG-WG, AG-RG, WG-RG can be obviously distinguished. In the AG-WG classification, the positive judgment probability for species = AG/WG is 100%, and its classification effect is significant. In the AG-RG classification, the positive judgment probability for species = AG is 100%, and that for species = RG is 85.7% with high total positive judgment probability (94.4%). In the WG - RG classification, the positive judgment probability for specie = RG is 100%, and that for the specie = WG is 80%, and the total positive judgment rate is 91.7%. The significance of its classification is lower than the first two cases Supplemental Table 7. It is to say, the difference between AG and WG/RG is higher than the discrimination between WG and RG. The results of the regression equation were consistent with the aforementioned results, so our analysis was credible.

However, it should be noted that only the ginsenosides with higher responses in the negative ESI mode were measured in this study. There are many ginsenosides with low content that should be further studied. In addition, future work is also needed for the identification of the "unknown ginsenosides" mentioned in this paper.

4. Conclusion

In this paper, 81 ginsenosides were identified (including 76 tentatively assigned ginsenosides) in ginseng samples using an optimized LC-Q-Orbitrap MS/MS method coupled with a ginsenoside-identifying strategy. A majority of the ginsenosides (59 of 81) were all shared by AG, RG and WG samples. Interestingly, the contents of ginsenoside Ro were relatively constant in AG, RG and WG samples. When the shared ginsenosides datasheet was normalized by ginsenoside Ro, our analysis strategy clearly divided the ginseng samples into three groups (i.e., WG, RG and AG groups). The results also indicated that RG and WG samples had more unity in the content of ginsenosides, while the relative content of 59 ginsenosides in RG and WG samples was significantly different from that in AG samples. To find the markers among AG, RG and WG, several state-of-the-art statistical analysis methods including HCA, PCA, PLS-DA and logistic regression analysis were performed based on the ginsenoside profiles. Our novel methodology based on ginsenoside profiles is more robust than existing methods, and data normalization is required to improve the efficiency of multivariate statistical analysis.

Declarations

Author contribution statement

Yahui Li: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Bingkun Yang, Wei Guo, Panpan Zhang, Jianghua Zhang, Jing Zhao: Performed the experiments; Analyzed and interpreted the data.

Qiao Wang, Wei Zhang, Xiaowei Zhang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Dezhi Kong: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Classification of three types of ginseng samples by Dezhi Kong was supported by National Natural Science Foundation of China [82174004 & 81872848], Technology Project of Hebei Education Department [QN2019045], Scientific Research Project of Hebei Administration of Traditional Chinese Medicine [2016172].

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e12044.

References

[1] J. Park, et al., Effects of ginseng on two main sex steroid hormone receptors: estrogen and androgen receptors, J. Ginseng Res. 41 (2) (2017) 215–221.
[2] A.S. Wong, C.M. Che, K.W. Leung, Recent advances in ginseng as cancer therapeutics: a functional and mechanistic overview, Nat. Prod. Rep. 32 (2) (2015) 256–272.
[3] C. Mancuso, R. Santangelo, Panax ginseng and Panax quinquefolius: from pharmacology to toxicology, Food Chem. Toxicol. 107 (Pt A) (2017) 362–372.
[4] A.S. Attele, J.A. Wu, C.S. Yuan, Ginseng pharmacology: multiple constituents and multiple actions, Biochem. Pharmacol. 58 (11) (1999) 1685–1693.
[5] W.Y. Ong, et al., Protective effects of ginseng on neurological disorders, Front. Aging Neurosci. 7 (2015) 129.
[6] N.T. Rokot, et al., A role of ginseng and its constituents in the treatment of central nervous system disorders, Evid. Based Complement. Alternat. Med. 2016 (2016), 2614742.
[7] T. Wang, et al., Traditional uses, botany, phytochemistry, pharmacology and toxicology of Panax notoginseng (Burk.) F.H. Chen: a review, J. Ethnopharmacol. 188 (2016) 234–258.
[8] Z.Y. Wang, et al., Pharmacological effects of active components of Chinese herbal medicine in the treatment of Alzheimer's disease: a review, Am. J. Chin. Med. 44 (8) (2016) 1525–1541.
[9] M. Karmazyn, M. Moey, X.T. Gan, Therapeutic potential of ginseng in the management of cardiovascular disorders, Drugs 71 (15) (2011) 1989–2008.
[10] J.L. Shergis, et al., Panax ginseng in randomised controlled trials: a systematic review, Phytother. Res. 27 (7) (2013) 949–965.
[11] W.C. Chuang, et al., A comparative study on commercial samples of ginseng radix, Planta Med. 61 (5) (1995) 459–465.
[12] L.P. Christensen, Ginsenosides chemistry, biosynthesis, analysis, and potential health effects, Adv. Food Nutr. Res. 55 (2009) 1–99.
[13] L.W. Qi, C.Z. Wang, C.S. Yuan, Ginsenosides from American ginseng: chemical and pharmacological diversity, Phytochemistry 72 (8) (2011) 689–699.
[14] W. Li, et al., Use of high-performance liquid chromatography-tandem mass spectrometry to distinguish Panax ginseng C. A. Meyer (Asian ginseng) and Panax quinquefolius L. (North American ginseng), Anal. Chem. 72 (21) (2000) 5417–5422.

[15] Q.L. Zhou, et al., Simultaneous quantification of twenty-one ginsenosides and their three aglycones in rat plasma by a developed UFLC-MS/MS assay: Application to a pharmacokinetic study of red ginseng, J. Pharm. Biomed. Anal. 137 (2017) 1–12.
[16] A. Ludwiczuk, B. Kołodziej, T. Wolski, The content and the composition of ginsenosides in different parts of American ginseng (Panax quinquefolium L.), Acta Agrobot. 59 (1) (2006) 507–514.
[17] W. Shi, et al., Investigation of ginsenosides in different parts and ages of Panax ginseng, Food Chem. 102 (3) (2007) 664–668.
[18] D. Xiao, et al., Accumulation characteristics and correlation analysis of five ginsenosides with different cultivation ages from different regions, J. Ginseng Res. 39 (4) (2015) 338–344.
[19] M.R. Harkey, et al., Variability in commercial ginseng products: an analysis of 25 preparations, Am. J. Clin. Nutr. 73 (6) (2001) 1101–1106.
[20] X. Zhou, V. Razmovski-Naumovski, K. Chan, A multivariate analysis on the comparison of raw notoginseng (Sanqi) and its granule products by thin-layer chromatography and ultra-performance liquid chromatography, Chin. Med. 10 (2015) 13.
[21] G.J. Lee, et al., Systematic development of a group quantification method using evaporative light scattering detector for relative quantification of ginsenosides in ginseng products, J. Pharm. Biomed. Anal. 128 (2016) 158–165.
[22] J.W. Lee, et al., Comprehensive profiling and quantification of ginsenosides in the root, stem, leaf, and berry of Panax ginseng by UPLC-QTOF/MS, Molecules 22 (12) (2017).
[23] S. Shibata, et al., Studies on saponins and sapogenins of ginseng. The structure of panaxatriol, Tetrahedron Lett. 42 (1965) 207–213.
[24] X. Huang, et al., Multicomponent assessment and ginsenoside conversions of Panax quinquefolium L. roots before and after steaming by HPLC-MS(n), J. Ginseng Res. 43 (1) (2019) 27–37.
[25] W.Z. Yang, et al., A novel neutral loss/product ion scan-incorporated integral approach for the untargeted characterization and comparison of the carboxyl-free ginsenosides from Panax ginseng, Panax quinquefolius, and Panax notoginseng, J. Pharm. Biomed. Anal. 177 (2020), 112813.
[26] H. Wang, et al., Rapid characterization of ginsenosides in the roots and rhizomes of Panax ginseng by UPLC-DAD-QTOF-MS/MS and simultaneous determination of 19 ginsenosides by HPLC-ESI-MS, J. Ginseng Res. 40 (4) (2016) 382–394.
[27] X. Xu, et al., Identification of mountain-cultivated ginseng and cultivated ginseng using UPLC/oa-TOF MSE with a multivariate statistical sample-profiling strategy, J. Ginseng Res. 40 (4) (2016) 344–350.
[28] J. Xue, et al., Using principal components analysis (PCA) with cluster analysis to study the organic geochemistry of sinking particles in the ocean, Org. Geochem. 42 (4) (2011) 356–367.
[29] S.M. Mercier, et al., Multivariate data analysis as a PAT tool for early bioprocess development data, J. Biotechnol. 167 (3) (2013) 262–270.
[30] A.S. Rathore, et al., Guidance for performing multivariate data analysis of bioprocessing data: pitfalls and recommendations, Biotechnol. Prog. 30 (4) (2014) 967–973.
[31] R.L. Sleighter, et al., Multivariate statistical approaches for the characterization of dissolved organic matter analyzed by ultrahigh resolution mass spectrometry, Environ. Sci. Technol. 44 (19) (2010) 7576–7582.
[32] Y. Chen, et al., An omic approach for the identification of oil sands process-affected water compounds using multivariate statistical analysis of ultrahigh resolution mass spectrometry datasets, Sci. Total Environ. 511 (2015) 230–237.
[33] J. Valsalan, T. Sadan, T. Venketachalapathy, Multivariate principal component analysis to evaluate growth performances in Malabari goats of India, Trop. Anim. Health Prod. 52 (5) (2020) 2451–2460.
[34] S.K. Palanisamy, et al., Metabolite profiling of ascidian Styela plicata using LC-MS with multivariate statistical analysis and their antitumor activity, J. Enzym. Inhib. Med. Chem. 32 (1) (2017) 614–623.
[35] P. Li, et al., Metabolomic analysis reveals the composition differences in 13 Chinese tea cultivars of different manufacturing suitabilities, J. Sci. Food Agric. 98 (3) (2018) 1153–1161.
[36] O.H. Abdelhafez, et al., Metabolomics analysis and biological investigation of three Malvaceae plants, Phytochem. Anal. 31 (2) (2020) 204–214.
[37] X. Wang, et al., Pseudoginsenoside-F11 (PF11) exerts anti-neuroinflammatory effects on LPS-activated microglial cells by inhibiting TLR4-mediated TAK1/IKK/NF-κB, MAPKs and Akt signaling pathways, Neuropharmacology 79 (2014) 642–656.
[38] E.H. Park, et al., Stereospecific anticancer effects of ginsenoside Rg3 epimers isolated from heat-processed American ginseng on human gastric cancer cell, J. Ginseng Res. 38 (1) (2014) 22–27.
[39] J.X. Mi, et al., Principal component analysis based on nuclear norm minimization, Neural Network. 118 (2019) 1–16.