A novel strategy to interpret the differences among white ginseng, red ginseng and American ginseng based on normalized ginsenoside profiles

Yahui Li
Hebei Medical University
Bingkun Yang
Hebei Medical University
Wei Guo
Hebei Medical University
Panpan Zhang
Hebei Medical University
Jianghua Zhang
Hebei Medical University
Jing Zhao
Hebei Medical University
Qiao Wang
Hebei Medical University
Wei Zhang
Hebei Medical University
Xiaowei Zhang
The Second Hospital of Hebei Medical University
Dezhi Kong (kongdezhi@hebmu.edu.cn)
Hebei Medical University

Research Article

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

Posted Date: May 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1623990/v1

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 based on the absolute content of ginsenosides.

Methods

The common commercial types of ginsengs in China 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. A special normalization to the datasheet was also introduced prior to multivariate statistical analysis and logistic regression analysis.

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. Our analysis strategy clearly divided the ginseng samples into three groups (i.e., WG, RG and AG groups). 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 indicated by the principal component analysis (PCA) and logistic regression analysis.

Conclusion

Our novel methodology according to ginsenoside profiles is more robust than that based on single ingredient, and could be widely used to distinguish the commercial types of ginsengs.

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 (Park J et al. 2017). "Panax" as a botanical name means "heal everything" in Greek (Wong AS et al. 2015). The history of using ginseng for its medical properties in China dates back approximately 5000 years (Mancuso C et al. 2017). It has been proven that ginseng exhibits antioxidant, anti-inflammatory,

anti-aging, anticancer, anti-apoptotic, and immune-stimulatory pharmacological activities, among others (Wong AS et al. 2015; Attele AS et al. 1999; Ong WY et al. 2015; Rokot NT et al. 2016; Wang T et al. 2016; Wang ZY et al. 2016). In recent years, ginseng has been widely sold as a dietary supplement and alternative medicine and has received great favor from users (Karmazyn M et al. 2011; Shergis JL et al. 2013).

The most common commercial types of ginseng in China are white ginseng (WG), red ginseng (RG) and American ginseng (AG), all under the *Panax* family. Each type of ginseng has slightly different uses (Chuang WC et al. 1995), 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 (Christensen LP 2009). 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 (Qi LW et al. 2011).

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 (Li W et al. 2000). 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. (Zhou QL et al. 2017). Compared to WG, RG was thought to have an increased anti-cancer property (Wong AS et al. 2015). 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 (Xiao D et al. 2015). Some ginsenoside concentrations varied by 15-fold (0.288–4.266% by *wt*) in *Panax* powders (Harkey MR et al. 2001). 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 (Zhou X et al. 2015), HPLC coupled with a UV detector or an evaporative light scattering detector (ELSD) (Lee GJ et al. 2016), and LC-MS (Zhou QL et al. 2017; Lee JW et al. 2017). Ginsenosides have such complex properties, such as diversity, similarity and complexity, that their analysis is a formidable challenge. Now 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 this paper, the ginsenosides in AG, WG and RG were analyzed using an LC-ESI-Orbitrap Fusion system and identified based the strategies shown in Fig. 1. The three types of ginsengs were well distinguished by hierarchical cluster analysis (HCA) and least squares discriminant analysis (PLS-DA) after our special normalization. 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 Fig. 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 (Shibata S et al. 1965). 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 mesh, then the fine ginseng powder was accurately weighed (1.0 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 extraction was repeated again using fresh aliquots of the solvent. When the combination of the two aliquots was achieved, 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 of 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 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 μ 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, 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

Thermo 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 Fig. 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 literature. 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 (markers) were recorded in a data matrix according to their names and peak areas. The contents of the reference markers could be calculated using the known concentrations of the standards. However, the unknown marker 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 due to weather events and laboratory errors 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 installed. 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

Classification using logistic regression is based on existing data to establish a regression formula for classification boundary lines for classification. The classification of three kinds of ginseng was further confirmed by logistic regression equation. In this paper, IBS SPSS Statistics 21 software is used to perform multivariate statistical regression and binary statistical regression to obtain classification formulas and classification scores to further increase reliability.

3. Results And Discussion

3.1. Optimization and validation of the quantitative analytical method

Since ginsenosides had higher sensitivity and clearer mass spectra in the negative ion mode, data collected in the negative ion mode were used for the component detection and characterization (Huang X et al. 2019; Yang WZ et al. 2020). 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.

For the precision test, the peak areas of 20 ginsenosides in real samples were analyzed at 0, 12 and 24 h. The RSD values of the retention times and the peak areas of 20 ginsenosides were less than 3.2% and 9.1%, respectively (Fig. 2A). The repeatability of the assay was confirmed by extracting and analyzing five

replicates of the same samples as described above The RSD values of the areas of 20 ginsenosides were all less than 5.5%.

The recovery was used to evaluate the accuracy of the method. A known amount of six ginsenoside standards was added into a certain amount of sample. The mixture was extracted and analyzed using the method above. Performing three replicates of the test. The developed method had good accuracy, with an overall recovery of 92.3–97.1%, RSD values ranging from 5.1–10.2% (supplemental Table 2). These results indicated that the LC-MS method was precise and accurate for the quantitative determination of ginsenosides in ginseng samples.

3.2. Identity assignment and confirmation of the ginsenosides

We used the reference ginsenosides to optimize the mass chromatographic conditions and to obtain the fragmentation pathways of ginsenosides (Huang X et al. 2019). From the MS scans of reference standards, the usual precursor ions of ginsenosides were $[M\text{-}H]^-$ and $[M + CH_3COO]^-$. The negative MS/MS spectra of the product ion $[M\text{-}H]^-$ exhibited a fragmentation pattern corresponding to the successive loss of the glycosidic units until the formation of $[aglycon\text{-}H]^-$ 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 $[M\text{-}H]^-$ 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 $[M\text{-}H\text{-}Glc]^-$, m/z 459.3859 for $[M\text{-}H\text{-}2Glc]^-$, m/z 179.0565 for $[Glc\text{-}H]^-$, and m/z 161.0457 for $[Glc\text{-}H_2O]^-$, with results shown in Fig. 2B.

The ginsenosides were identified and confirmed by the strategies shown in Fig. 1, and all of 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. 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 (Wang HP et al. 2016; Xu XF et al. 2016). 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. Constituents analysis of ginseng samples

In our study, there were 81 identified ginsenosides in the three types of ginseng (supplemental Table 3). As illustrated in Fig. 3A, 2, 2 and 6 ginsenosides were only found in WG, RG and AG, respectively. As shown in Fig. 3B, ginsenoside Rs5 ($C_{44}H_{72}O_{13}$, RT 6.6 min), for 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 × analytes). Figure 4A shows the hierarchical cluster analysis (HCA) results for sample clustering based on all 59 ginsenosides. The relative distances are proportional to the correlation between samples, so a smaller relative distance means a higher similarity between samples than a pair with a larger distance (Xue, J et al. 2011). 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 according to analysis of the shared 59 ginsenosides. Moreover, with decreasing relative distances (and increasing correlations), the samples in the second branch were further sub-divided into two groups that corresponded to RG and WG. To increase the comparability among samples, the peak area of each sample was normalized by the peak area of ginsenoside Ro, since the peak areas of ginsenoside Ro among the AG, RG, WG samples were not statistically different (Fig. S2). Figure 4B shows the HCA for sample clustering after the normalization. Figure 4B more clearly shows the difference among the AG, RG and WG. Least squares discriminant analysis (PLS-DA) also proved that the normalized data more easily yielded satisfactory categorization of the samples (Fig. 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.

Although these groupings are useful for qualitative interpretation, a limitation of HCA is that the phylogenetic trees cannot be used to determine which markers cause major differences between samples (Mercier SM et al. 2013; Rathore AS et al. 2014). A principal component analysis (PCA) was performed for more rigorous interpretation of the datasets, so PCA have an advantage over HCA used alone (Sleighter RL et al. 2010).

Figure 5A and Fig. 5B 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 (Chen Y et al. 2015; Valsalan J et al. 2020). The two ranking PCs, 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 (Valsalan J et al. 2020). 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 labeled datasets (Palanisamy SK et al. 2017; Li P et al. 2018), Similar to the results of the HCA, the distances between samples in the PCA score plot were proportional to the similarities/differences between samples (Mercier SM et al. 2013). 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 (Fig. 5) was highly consistent with the previous HCA clustering (Fig. 4), which further validates the statistical results (Chen Y et al. 2015). 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 (Sleighter RL et al. 2010).

Biplots were created by combining PCA loading plots (Fig. S4) with score plots to account for correlations between sample groups and individual markers. In Fig. 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 (Abdelhafez OH et al. 2020). 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 (Fig. S4) which corresponded with the AG sample cluster in the score plot (Fig. 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 (Li W et al. 2000) 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 (Wang X et al. 2014).

For the WG samples, most of the markers were found in Q4 of the loading plot (Fig. 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 (Fig. 6). The level of ginsenoside Rg1, which has pharmacological use through producing weak stimulation to the central nervous system, indicated that WG is more “warm” than AG (Harkey MR et al. 2001).

For the RG samples, most of the markers were found in Q3 of the loading plot (Fig. 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 (Fig. 6). Compared with American ginseng and white ginseng, the content of ginsenoside Rg3 was approximately 3-fold in red ginseng (Fig. 6). This was consistent with previous reporting that the amounts of ginsenosides Rg3 and Rg5 increased after the hot steaming process (Park EH et al. 2014). Compared with Asian white ginseng, red ginseng has stronger anticancer activities (Wong AS et al. 2015) 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 (Fig. S5). Based on our multivariate analysis results, other ratios of ginsenosides could also be suitable to distinguish the three types of ginseng.

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 (Mi JX et al. 2019). 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:

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

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:

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

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 are 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 found 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 American ginseng, red ginseng and white ginseng. Interestingly, the contents of ginsenoside Ro were relatively constant in AG, RG and WG samples. The samples were clearly divided into three groups by multivariate statistical analysis and cluster analysis (i.e., AG, RG and WG groups) after the specific data normalization. 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. This is the first time that the differences among AG, WG and RG have been illustrated clearly at the ginsenosides level.

Declarations

Acknowledgments This work was supported by National Nature Science Foundation of China (82174004 and 81872848, China), Science and Technology Project of Hebei Education Department (Grants No. QN2019045) and Scientific Research Project of Hebei Administration of Traditional Chinese Medicine (Grants No. 2016172).

Author Contribution Dezhi Kong, Xiaowei Zhang, Qiao Wang and Wei Zhang conceived and designed research. Yahui Li, Bingkun Yang, Wei Guo, Panpan Zhang, Jianghua Zhang and Jing Zhao collected and analyzed the data. Yahui Li and Dezhi Kong wrote the initial manuscript. Xiaowei Zhang, Qiao Wang and Wei Zhang revised the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by National Nature Science Foundation of China (82174004 and 81872848, China), Science and Technology Project of Hebei Education Department (Grants No. QN2019045) and Scientific Research Project of Hebei Administration of Traditional Chinese Medicine (Grants No. 2016172).

Data Availability Statements The authors declare that [the/all other] data supporting the findings of this study are available within the article [and its supplementary information files]

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to Participate The authors of this article have endorsed this article.

Conflict interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Park J, Song H, Kim SK, Lee MS, Rhee DK, Lee Y. Effects of ginseng on two main sex steroid hormone receptors: estrogen and androgen receptors. J Ginseng Res. 2017 Apr;41(2):215–221.
2. Wong AS, Che CM, Leung KW. Recent advances in ginseng as cancer therapeutics: a functional and mechanistic overview. Nat Prod Rep. 2015 Feb;32(2):256–72.
3. Mancuso C, Santangelo R. Panax ginseng and Panax quinquefolius: From pharmacology to toxicology. Food Chem Toxicol. 2017 Sep;107(Pt A):362–372.
4. Attele AS, Wu JA, Yuan CS. Ginseng pharmacology: multiple constituents and multiple actions. Biochem Pharmacol. 1999 Dec 1;58(11):1685-93.
5. Ong WY, Farooqui T, Koh HL, Farooqui AA, Ling EA. Protective effects of ginseng on neurological disorders. Front Aging Neurosci. 2015 Jul 16;7:129.
6. Rokot NT, Kairupan TS, Cheng KC, Runtuwene J, Kapantow NH, Amitani M, Morinaga A, Amitani H, Asakawa A, Inui A. A Role of Ginseng and Its Constituents in the Treatment of Central Nervous System Disorders. Evid Based Complement Alternat Med. 2016;2016:2614742.
7. Wang T, Guo R, Zhou G, Zhou X, Kou Z, Sui F, Li C, Tang L, Wang Z. Traditional uses, botany, phytochemistry, pharmacology and toxicology of Panax notoginseng (Burk.) F.H. Chen: A review. J Ethnopharmacol. 2016 Jul 21;188:234 – 58.
8. Wang ZY, Liu JG, Li H, Yang HM. Pharmacological Effects of Active Components of Chinese Herbal Medicine in the Treatment of Alzheimer's Disease: A Review. Am J Chin Med. 2016;44(8):1525–1541.
9. Karmazyn M, Moey M, Gan XT. Therapeutic potential of ginseng in the management of cardiovascular disorders. Drugs. 2011 Oct 22;71(15):1989–2008.
10. Shergis JL, Zhang AL, Zhou W, Xue CC. Panax ginseng in randomised controlled trials: a systematic review. Phytother Res. 2013 Jul;27(7):949–65.
11. Chuang WC, Wu HK, Sheu SJ, Chiou SH, Chang HC, Chen YP. A comparative study on commercial samples of ginseng radix. Planta Med. 1995 Oct;61(5):459–65.
12. Christensen LP. Ginsenosides chemistry, biosynthesis, analysis, and potential health effects. Adv Food Nutr Res. 2009;55:1–99.

13. Qi LW, Wang CZ, Yuan CS. Ginsenosides from American ginseng: chemical and pharmacological diversity. Phytochemistry. 2011 Jun;72(8):689–99.
14. Li W, Gu C, Zhang H, Awang DV, Fitzloff JF, Fong HH, van Breemen RB. 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. 2000 Nov 1;72(21):5417–22.
15. Zhou QL, Zhu DN, Yang YF, Xu W, Yang XW. 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. 2017 Apr 15;137:1–12.
16. Xiao D, Yue H, Xiu Y, Sun X, Wang Y, Liu S. Accumulation characteristics and correlation analysis of five ginsenosides with different cultivation ages from different regions. J Ginseng Res. 2015 Oct;39(4):338–44.
17. Harkey MR, Henderson GL, Gershwin ME, Stern JS, Hackman RM. Variability in commercial ginseng products: an analysis of 25 preparations. Am J Clin Nutr. 2001 Jun;73(6):1101–6.
18. Zhou X, Razmovski-Naumovski V, Chan K. 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. 2015 Jun 6;10:13.
19. Lee GJ, Shin BK, Yu YH, Ahn J, Kwon SW, Park JH. Systematic development of a group quantification method using evaporative light scattering detector for relative quantification of ginsenosides in ginseng products. J Pharm Biomed Anal. 2016 Sep 5;128:158–165.
20. Lee JW, Choi BR, Kim YC, Choi DJ, Lee YS, Kim GS, Baek NI, Kim SY, Lee DY. Comprehensive Profiling and Quantification of Ginsenosides in the Root, Stem, Leaf, and Berry of Panax ginseng by UPLC-QTOF/MS. Molecules. 2017 Dec 4;22(12):2147.
21. Shibata S, Tanaka O, Soma K, Ando T, Iida Y, & Nakamura H. Studies on Saponins and Sapogenins of Ginseng. The Structure of Panaxatriol. Tetrahedron Letters. 1965 42, 207–213.
22. Huang X, Liu Y, Zhang Y, Li SP, Yue H, Chen CB, Liu SY. Multicomponent assessment and ginsenoside conversions of Panax quinquefolium L. roots before and after steaming by HPLC-MSn. J Ginseng Res. 2019 Jan;43(1):27–37.
23. Yang WZ, Shi XJ, Yao CL, Huang Y, Hou JJ, Han SM, Feng ZJ, Wei WL, Wu WY, Guo DA. 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. 2020 Jan 5;177:112813.
24. Wang HP, Zhang YB, Yang XW, Zhao DQ, Wang YP. 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. 2016 Oct;40(4):382–394.
25. Xu XF, Cheng XL, Lin QH, Li SS, Jia Z, Han T, Lin RC, Wang D, Wei F, Li XR. Identification of mountain-cultivated ginseng and cultivated ginseng using UPLC/oa-TOF MSE with a multivariate statistical sample-profiling strategy. J Ginseng Res. 2016 Oct;40(4):344–350.

26. Xue, J. H., Lee, C., Wakeham, S. G., & Armstrong, R. A. Using principal components analysis (PCA) with cluster analysis to study the organic geochemistry of sinking particles in the ocean.Organic Geochemistry. 2011; 42(4), 356–367.
27. Mercier SM, Diepenbroek B, Dalm MC, Wijffels RH, Streefland M. Multivariate data analysis as a PAT tool for early bioprocess development data. J Biotechnol. 2013 Sep 10;167(3):262–70.
28. Rathore AS, Mittal S, Pathak M, Arora A. Guidance for performing multivariate data analysis of bioprocessing data: pitfalls and recommendations. Biotechnol Prog. 2014 Jul-Aug;30(4):967–73.
29. Sleighter RL, Liu Z, Xue J, Hatcher PG. Multivariate statistical approaches for the characterization of dissolved organic matter analyzed by ultrahigh resolution mass spectrometry. Environ Sci Technol. 2010 Oct 1;44(19):7576-82.
30. Chen Y, McPhedran KN, Perez-Estrada L, Gamal El-Din M. 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. 2015 Apr 1;511:230–7.
31. Valsalan J, Sadan T, Venketachalapathy T. Multivariate principal component analysis to evaluate growth performances in Malabari goats of India. Trop Anim Health Prod. 2020 Sep;52(5):2451–2460.
32. Palanisamy SK, Trisciuoglio D, Zwergel C, Del Bufalo D, Mai A. Metabolite profiling of ascidian Styela plicata using LC-MS with multivariate statistical analysis and their antitumor activity. J Enzyme Inhib Med Chem. 2017 Dec;32(1):614–623.
33. Li P, Dai W, Lu M, Xie D, Tan J, Yang C, Zhu Y, Lv H, Peng Q, Zhang Y, Guo L, Ni D, Lin Z. Metabolomic analysis reveals the composition differences in 13 Chinese tea cultivars of different manufacturing suitabilities. J Sci Food Agric. 2018 Feb;98(3):1153–1161.
34. Abdelhafez OH, Othman EM, Fahim JR, Desoukey SY, Pimentel-Elardo SM, Nodwell JR, Schirmeister T, Tawfike A, Abdelmohsen UR. Metabolomics analysis and biological investigation of three Malvaceae plants. Phytochem Anal. 2020 Mar;31(2):204–214.
35. Wang X, Wang C, Wang J, Zhao S, Zhang K, Wang J, Zhang W, Wu C, Yang J. 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. 2014 Apr;79:642–56.
36. Park EH, Kim YJ, Yamabe N, Park SH, Kim HK, Jang HJ, Kim JH, Cheon GJ, Ham J, Kang KS. Stereospecific anticancer effects of ginsenoside Rg3 epimers isolated from heat-processed American ginseng on human gastric cancer cell. J Ginseng Res. 2014 Jan;38(1):22–7.
37. Mi JX, Zhang YN, Lai Z, Li W, Zhou L, Zhong F. Principal Component Analysis based on Nuclear norm Minimization. Neural Netw. 2019 Oct;118:1–16.

Figures

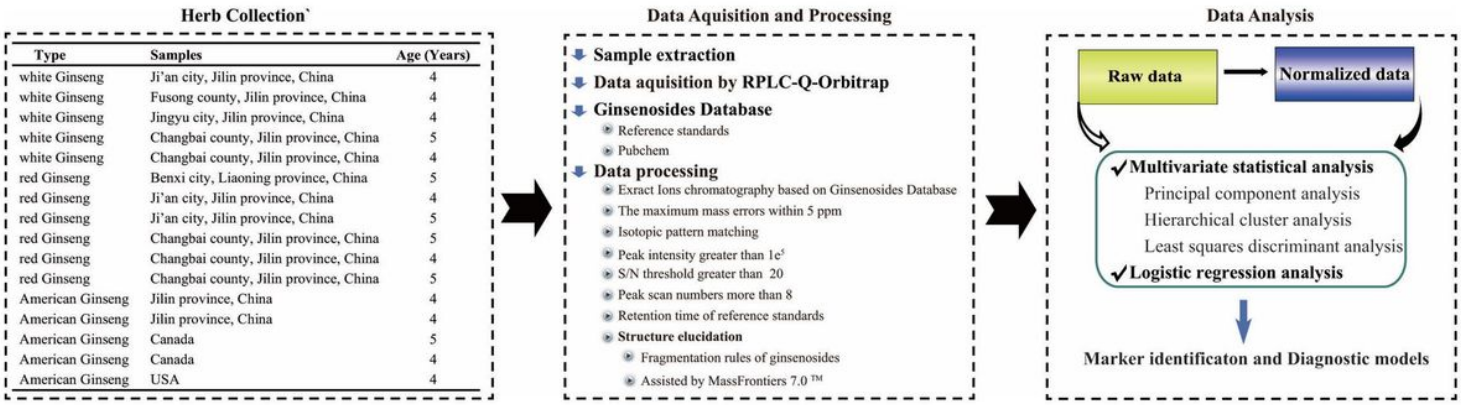

Figure 1

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

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.

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.

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.

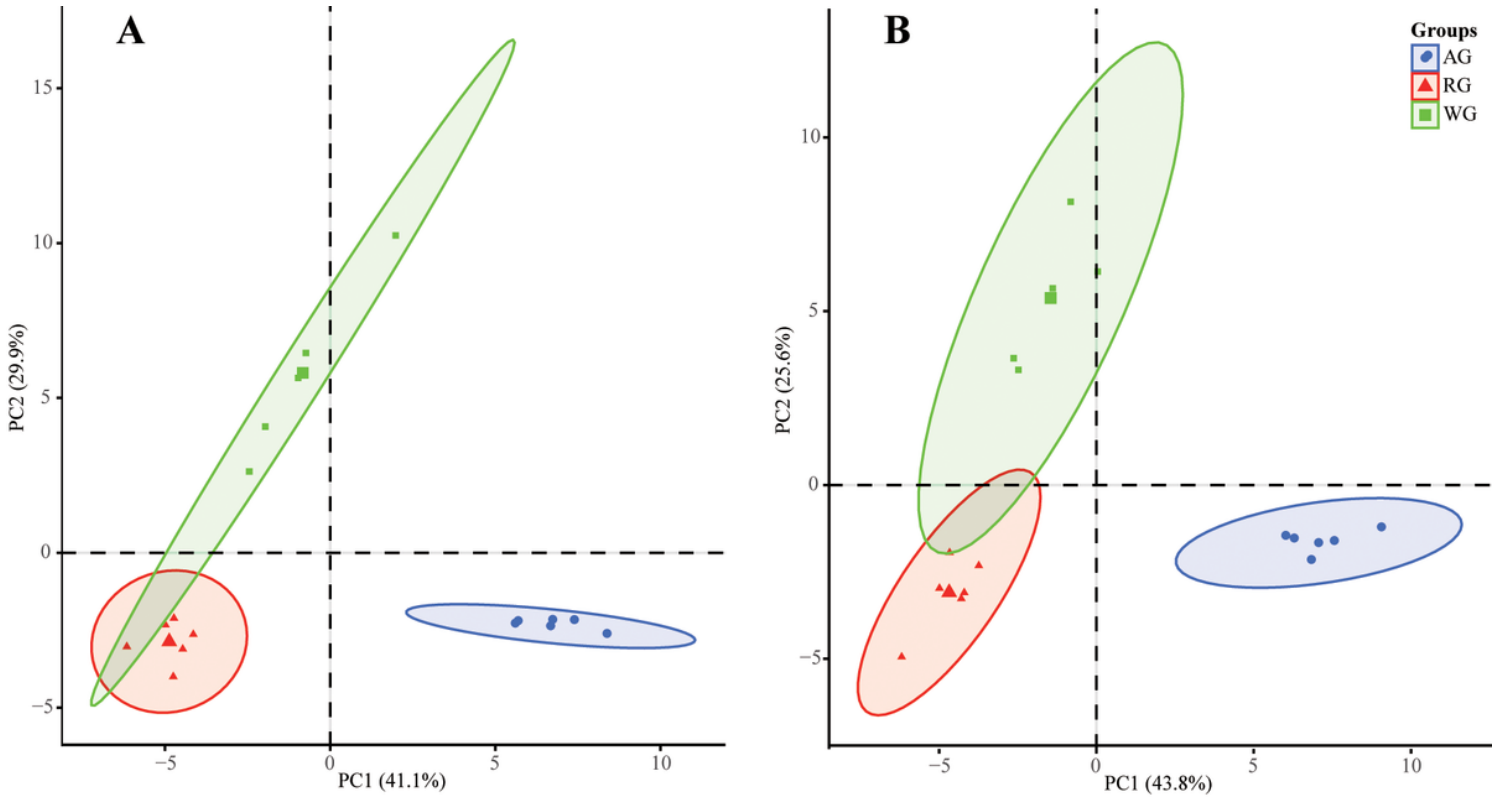

Figure 5

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.

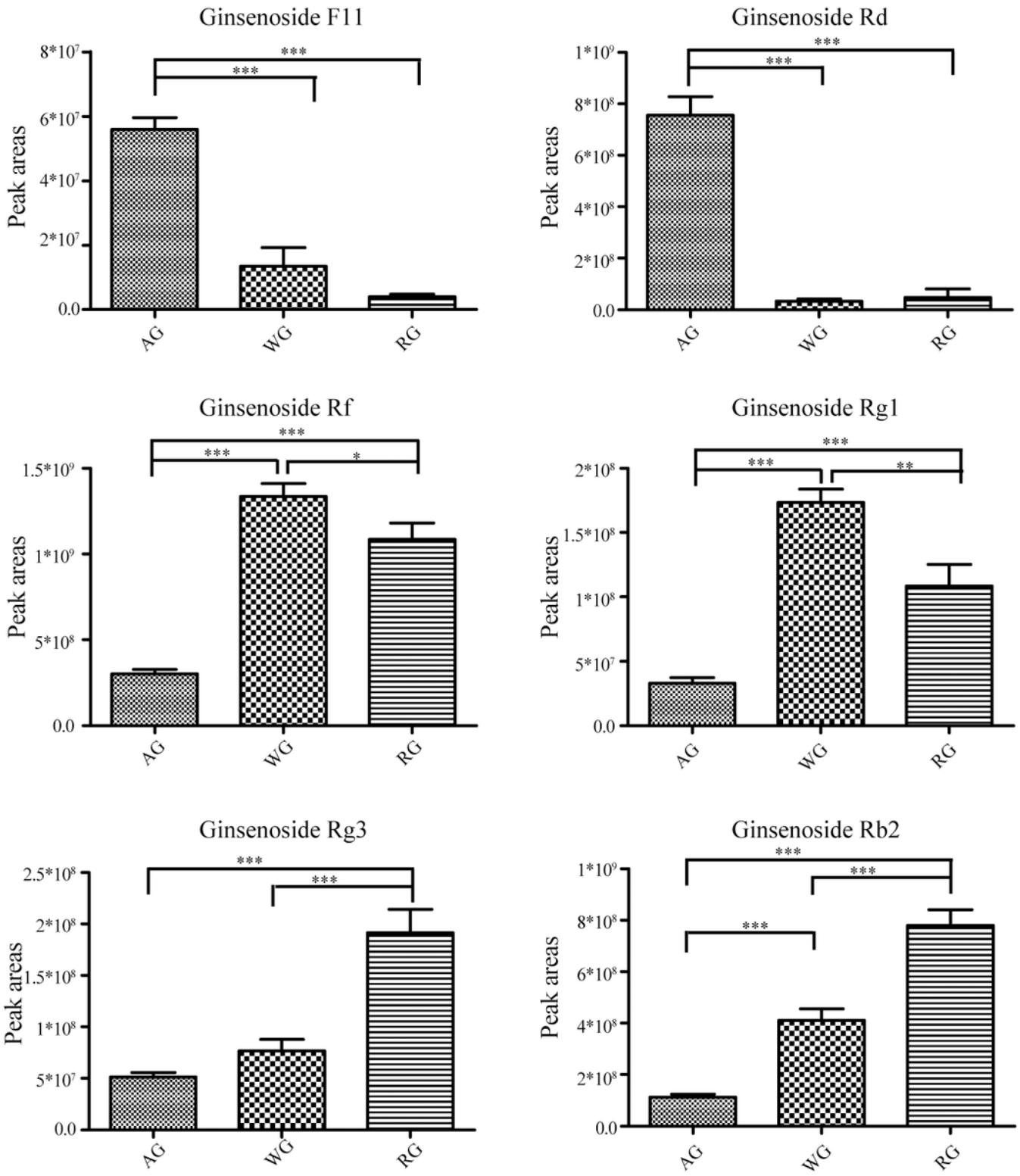

Figure 6

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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S1.tif
- Fig.S2.tif
- fig.S3.tif
- fig.S4.tif
- Fig.S5.tif
- supplementTable.docx