Anal Bioanal Chem (2011) 399:1877–1889
DOI 10.1007/s00216-010-4586-7

ORIGINAL PAPER

Differentiation of *Panax quinquefolius* grown in the USA and China using LC/MS-based chromatographic fingerprinting and chemometric approaches

Jianghao Sun · Pei Chen

Received: 15 September 2010 / Revised: 30 November 2010 / Accepted: 7 December 2010 / Published online: 13 January 2011

Abstract American ginseng (*Panax quinquefolius*) is one of the most commonly used herbal medicines in the world. Discriminating between *P. quinquefolius* grown in different countries is difficult using traditional quantitation methods. In this study, a liquid chromatographic mass spectrometry fingerprint combined with chemometric analysis was established to discriminate between American ginseng grown in the USA and China. Fifteen American ginseng samples grown in Wisconsin and 25 samples grown in China were used. The chromatographic fingerprints, representing the chemical compositions of the samples, made it possible to distinguish samples from the two locations. In addition, it was found that some ginsenosides varied widely from *P. quinquefolius* cultivated in these two countries. *P. quinquefolius* grown in the USA is higher in ginsenoside R_c, ginsenoside R_d, quinquenoside III/pseudo-ginsenoside RC_1, malonyl ginsenoside R_{b1}, and ginsenoside R_{b2}, but lower in ginsenoside R_{b1} compared with *P. quinquefolius* grown in China. These ginsenosides may be responsible for the class separation seen using fingerprinting and chemometric approaches.

Keywords American ginseng · Fingerprint · Chemometrics · Liquid chromatography–mass spectrometry · *Panax quinquefolius*

J. Sun · P. Chen (✉)
US Department of Agriculture, Agricultural Research Service,
Beltsville Human Nutrition Research Center,
Food Composition and Method Development Laboratory,
Beltsville, MD 20705, USA
e-mail: pei.chen@ars.usda.gov

Introduction

American ginseng (*Panax quinquefolius* L.) is one of the most commonly used herbal medicines in the world and is known for its diverse pharmacological activities. The ginseng roots and their related dietary supplements have also become popular in US markets. Compared with Asian ginseng (*Panax ginseng*), American ginseng is considered to calm and nourish the human "yin" and has a "cooling effect" [1]. Preclinical studies have demonstrated that *P. quinquefolius* has beneficial health effects for the central nervous system [2–4], cardiovascular system [5–7], and immune system [8], and may contribute to cancer prevention [9–12]. These studies have led to a focus on the saponins, also known as ginsenosides, as the components providing the health benefits. Thus, the analysis and evaluation of *P. quinquefolius*, or its related products, has mainly focused on the assay of ginsenosides [13–17].

P. quinquefolius grows in the USA and Canada in the wild and has been cultivated in North America since the last century, mainly in Wisconsin. It has also been introduced to China in recent years. However, there is suspected therapeutic difference and a considerable price difference for *P. quinquefolius* cultivated in Wisconsin and China. There have been quite a few reports on discrimination among Asian ginseng (*P. ginseng*), American ginseng (*P. quinquefolius*), and notoginseng (*P. notoginseng*) based on the assay of ginsenosides [15, 18–21]. Several qualitative and quantitative analytical techniques have been used for the analysis of ginsenosides. These analytical techniques include high-performance liquid chromatography (HPLC) combined with use of an ultraviolet (UV) detector, an evaporative light scattering detector, and a mass spectrometry (MS) detector. Among these methods, liquid chromatography (LC)–MS has been the most popular for

the analysis of ginsenosides in ginseng extracts or biofluids owing to its superior sensitivity and specificity. However, discrimination between *P. quinquefolius* grown in different locations is difficult using traditional methods. There has only been one report of differentiation of American ginseng grown in China and North America by the qualitative analysis of ginsenosides with an HPLC-MS method [22]. The results showed that majonoside may be used as a marker to distinguish *P. quinquefolius* grown in China and North America and there is higher content of ginsenoside R_d in *P. quinquefolius* from North America. However, the number of samples used in that study was small and no detailed information regarding the growing year and the harvest time was provided. Schlag and McIntosh [23] reported that ginsenoside contents and compositions varied significantly among populations on the basis of the quantitation of five ginsenosides (R_{g1}, R_e, R_{b1}, R_c, and R_d) in American ginseng roots collected from ten populations grown in Maryland and suggested the use of fingerprinting techniques for regulation.

The fingerprint approach used in this study focuses on the chromatographic patterns of the samples. It is an effective approach for species differentiation and quality control of complex botanical materials [24]. Analytical methods including gas chromatography (GC), HPLC, GC-MS, and LC-MS have been successfully employed for fingerprint analysis. Among these analytical techniques, LC-MS fingerprint analysis has played an important role in the quality assessment of botanical materials [25, 26]. Spectral and chromatographic fingerprints have been considered as an acceptable approach for quality control and assessment of botanical materials. In previous studies, we have successfully employed this technique for *Ginkgo biloba* [27], dry beans [28], pycnogenol [29], grapefruit [30], broccoli [31], and skullcap [32] with UV, near-infrared, direct infusion MS, and LC-MS methods. The combination of fingerprints with chemometric analysis provided the ability to distinguish chemical differences between samples.

Chemometrics is commonly used for the study of small-molecule metabolite profiles, which aims to identify and quantify the full complement in tissues [33, 34]. It has been applied to a variety of areas, such as plant toxicology, plant science, nutrition, and systems biology [35–38]. Chemometric approaches for authentication and quality evaluation of botanical materials in teas, different species of the *Panax* genus, and Korean and Chinese red ginsengs have been reported [39–43]. Chemometrics is especially useful for identifying the components that are responsible for the chemical differences between the samples. There are quite a few commercial software packages from instrument companies, such as MarkerView™, Markerlynx™, and Sieve ™, as well as open-source programs, such as XCMS [44] and MZmine [45], that provide peak detection, noise filtration, peak alignment, and integration with public libraries for compound identification.

In this study, an LC-MS fingerprint and a chemometric approach, with multivariate statistics, was used to discriminate between *P. quinquefolius* samples grown in the USA and China. The study design is shown in Fig. 1. Variance was observed for the samples cultivated in these two locations. The results showed that there were chemical differences between *P. quinquefolius* grown in different locations and the chemometric approach was used to pinpoint the compounds (mainly ginsenosides) responsible for the differences.

Experimental

Materials

The 15 *P. quinquefolius* samples grown in the USA were purchased from the American Herbal Pharmacopeia (AHP) and were labeled from A01 to A15. The 25 *P. quinquefolius* samples grown in China were labeled from

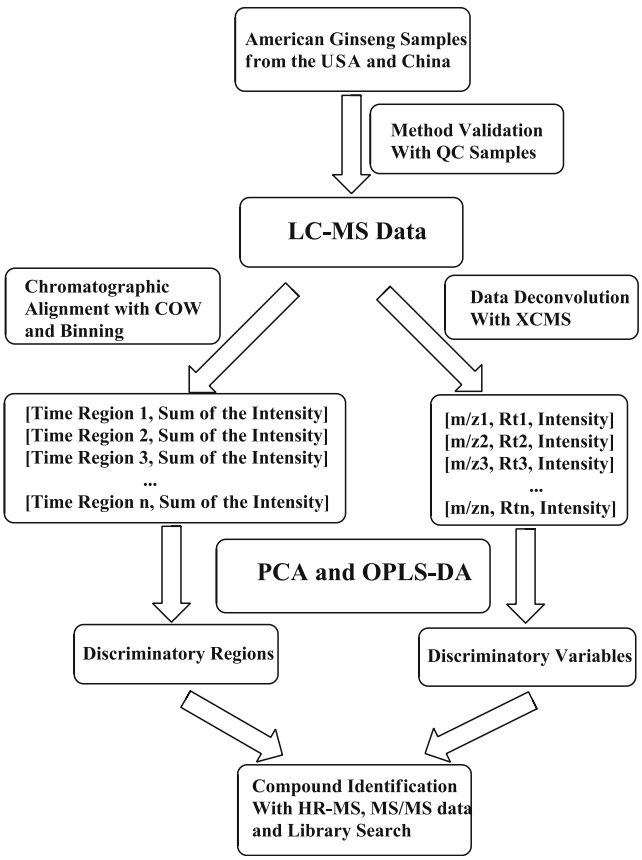


Fig. 1 Flow chart for the discrimination of American ginseng grown in the USA and China. *QC* quality control, *LC* liquid chromatography, *MS* mass spectrometry, COW correlation-optimized warping, *PCA* principal component analysis, *OPLS-DA* orthogonal projection to latent structural discriminate analysis, *HR* high resolution

B01 to B25. Samples B01 to B09 were purchased in Beijing, China; samples B10 and B11 were purchased in Guiyang, China; samplers B12 and B13 were purchased in Shangxi, China; samples B14 to B17 were purchased in Shangdong, China; samples B18 and B19 were purchased in Liaoning, China; and samples B20 to B25 were collected from local farmers in Jilin, China. All ginseng samples were root samples. The identities of these samples were confirmed using LC/MS. Vouchers of the AHP samples have been deposited in the author's laboratory. The samples from China did not come with vouchers. All the samples were stored at room temperature.

Standards of ginsenosides R_{b1}, R_{b2}, R_{b3}, R_c, R_d, R_e, R_{g1}, R_{g2}, R_{g3}, R_{h1}, and R_{h2} were purchased from ChromaDex (Irvine, CA, USA). The chemical structures of the ginsenosides investigated are shown in Table 1. Opitma-grade acetonitrile, methanol, and water were purchased from Thermo Fisher Scientific (Waltham, MA, USA). MS-grade formic acid was from Sigma-Aldrich (St. Louis, MO, USA). American ginseng extract standard (lot no. 1291683) was purchased from US Pharmacopoeia (Rockville, MD, USA) and was used as a quality control sample.

Sample preparation

P. quinquefolius samples were finely pulverized to homogeneity with an IKA A 11 basic S1 grinder (IKA, Wilmington, NC, USA) to the same mesh size (40 mesh). Each powder (300 mg) was dispersed and sonicated with 10 mL of 50% ethanol for 1 h. Prior to analysis, all sample solutions were filtered through a 0.20-μm polytetrafluoroethylene filter (Whatman, Florham Park, NJ, USA).

The reference ginsenosides were individually dissolved in an appropriate volume of methanol to yield 100 μg/mL stock solutions. An aliquot of each stock solution (100 μL) was mixed to give a mixed standard solution. All stock solutions and working solutions were stored at −20 °C and brought to room temperature prior to use.

Liquid chromatography

The HPLC separation was performed using an Agilent 1200 liquid chromatography system (Agilent, Santa Clara, CA, USA) equipped with a binary gradient pump (G1312B), a temperature-controlled autosampler (G1367C), and a diode-array detector (G1315C). An HPLC method was developed using a reversed-phase C_{18} column (Phenomenex Luna C_{18}, 2.1 mm×2 50 mm, 5 μm). The binary gradient elution system consisted of water (0.1% v/v formic acid) (solvent A) and acetonitrile (0.1% v/v formic acid) (solvent B) and separation was achieved using the following gradient: 0–10 min, 10–30% solvent B; 10–30 min, 30–40% solvent B; 30–40 min, 40–80% solvent B; and 40–60 min, 80–95% solvent B. The column temperature was kept constant at 40 °C. The flow rate was 0.3 mL/min and the injection volume was 1 μL. Quality control samples and blanks were injected every ten samples to monitor the potential drift of the instrument.

Mass spectrometry

MS analyses were performed with an Applied Biosystems/MDS Sciex (Toronto, ON, Canada) triple quadrupole linear ion trap mass spectrometer (Qtrap 4000), equipped with an electrospray ionization source; the interface was operated in the negative ion mode. High-purity nitrogen served as the nebulizer and drying gas with a pressure of 65 and 55 psi, respectively. The curtain gas was set at 20 psi. The turbo ion spray temperature was maintained at 500 °C and the needle voltage was optimized at −4,500 V. The declustering potential and the entrance potential were set at −150 V and −10 V, respectively. MS parameters were optimized with 10 μg/mL ginsenoside R_{g1} by using direct infusion in the mobile phase (acetonitrile/water, 1:1, v/v) with a flow rate of 0.3 mL/min. The full scan range was from 200 to 1,500 amu.

Accurate mass measurements were performed with a Thermo Fisher Exactive mass spectrometer coupled to a Thermo Accela HPLC system (Waltham, MA, USA). Sample ionization was carried out in the negative mode using the electrospray source at a capillary temperature of 250 °C, a sheath gas flow of 45 units, an auxiliary gas flow of 15 units, and an electrospray voltage of 3.0 kV. Full scan centroid data over a range from 100 to 1,500 amu were then acquired using a resolution of 30,000 at full width at half maximum.

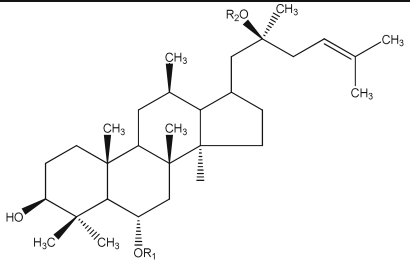
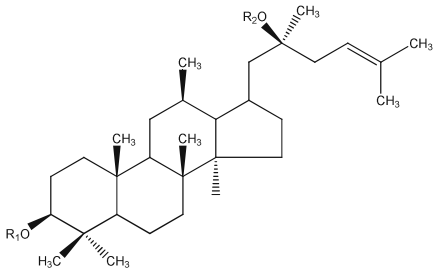
Peak alignment

For LC-MS, the total ion current chromatograms of each sample were exported to a text file using Applied Biosystems Analyst 1.5 (Applied Biosystems, Waltham, MA, USA). The text files of all samples were combined into one spreadsheet using Microsoft Excel. This Excel file was imported into MATLAB 7.0 (The MathWorks, Natick, MA, USA) and aligned using a script based on correlation-optimized warping (COW) [46]. The aligned data were then exported to Excel and the result was a 40 (observations) × 1,662 (data points) matrix which was used for further statistical analysis.

Peak finding with XCMS

The LC-MS raw files were first converted to mzData format with Wiff to mzData Translator (version 1.0, Applied Biosystems, Waltham, MA, USA,), and were subsequently processed by the XCMS package (version 1.14.1, http://metlin.scripps.edu/download/) under R (version 2.10.1, http://www.r-project.org) using optimized settings of XCMS. The "matched filter" method and a nonlinear method were

Table 1 Structures of ginsenosides investigated

Structure	Name	R1	R2
	Ginsenoside R_{g1}	-Glc	-Glc
	Ginsenoside R_{g2}	-Glc^{2}-^{1}Rha	-H
	Ginsenoside R_e	-Glc^{2}-^{1}Rha	-Glc
	Ginsenoside R_f	-Glc^{2}-^{1}Glc	-Glc
	Notoginsenoside R_2	-Glc^{2}-^{1}Xyl	-H
	Ginsenoside R_{h1}	-Glc	-H
	Ginsenoside R_{b1}	-Glc^{2}-^{1}Glc	-Glc^{6}-^{1}Glc
	Ginsenoside R_{b2}	-Glc^{2}-^{1}Glc	-Glc^{6}-^{1}Ara(p)
	Ginsenoside R_c	-Glc^{2}-^{1}Glc	-Glc^{6}-^{1}Ara(f)
	Ginsenoside R_d	-Glc^{2}-^{1}Glc	-Glc
	Ginsenoside R_{h2}	-Glc	-H
	Ginsenoside R_{g3}	-Glc^{2}-^{1}Glc	-H
	Ginsenoside R_{b3}	-Glc^{2}-^{1}Glc	-Glc^{6}-^{1}Xyl
	Quinquenoside III	-Glc^{2}-^{1}Glc (6 Ac)	-Glc
	Pseudo-ginsenoside RC_1	-Glc^{2}-^{1}Glc^{6}-Ac	-Glc
	Quinquenoside R_1	-Glc^{2}-^{1}Glc^{6}-Ac	-Glc^{6}-^{1}Glc
	Gypenoside XVII	-Glc	-Glc^{6}-^{1}Glc
	Malonyl ginsenoside R_{b1}	-Glc^{2}-^{1}Glc^{6}-malonyl	-Glc^{6}-^{1}Glc

selected for the detection of the ion features and peak alignment. The peaklist was then output to a .csv file.

Statistics

Multivariate statistical analysis, including unsupervised principal component analysis (PCA) and supervised orthogonal projection to latent structural discriminate analysis (OPLS-DA), was performed with SIMCA-P 11.5. (Umetrics, Umeå, Sweden). The *t* test was performed using Microsoft Excel 2007.

Results and discussion

Variability of quality control samples during the analytical sequence

In a fingerprint or chemometric study, the robustness of the analytical method should be examined to guarantee the statistical difference was not from instrumental drift. A blank run was inserted into the analysis sequence every ten samples to test for cross talk. The quality control samples were measured five times during the entire analytical run and six ginsenoside peaks were selected to monitor the instrumental drift. The retention time reproducibility for the six ginsenosides was found to be consistent (less than 0.1 min) during the whole analysis. The reproducibility of the retention time was important for the subsequent peak alignment and peak picking. The variations of the peak areas of the six ginsenosides in the quality control sample were no more than 10% (Table 2), suggesting consistency of the instrumental performance during the whole analytical run.

Ginsenoside profile of the samples

The HPLC-MS chromatograms of *P. quinquefolius* grown in China and the USA are shown in Fig. 2. Nineteen ginsenosides were characterized using the reference stand-

Table 2 The retention time and peak area variations for the main peaks in ginseng samples

Peak no.	[M-H]$^-$	Retention time (mean ± SD)	Peak area (× 10^8) (mean ± SD)	RSD (%)
1	945.5	15.71±0.03	8.12±0.59	5.66
2	799.7	23.09±0.07	6.15±0.35	7.28
3	1,107.6	25.19±0.06	14.78±1.31	8.87
4	1,149.5	26.26±0.04	6.14±0.44	7.17
5	955.7	27.88±0.06	11.36±1.04	9.16
6	987.7	33.30±0.08	2.37±0.21	8.72

SD standard deviation, *RSD* relative standard deviation

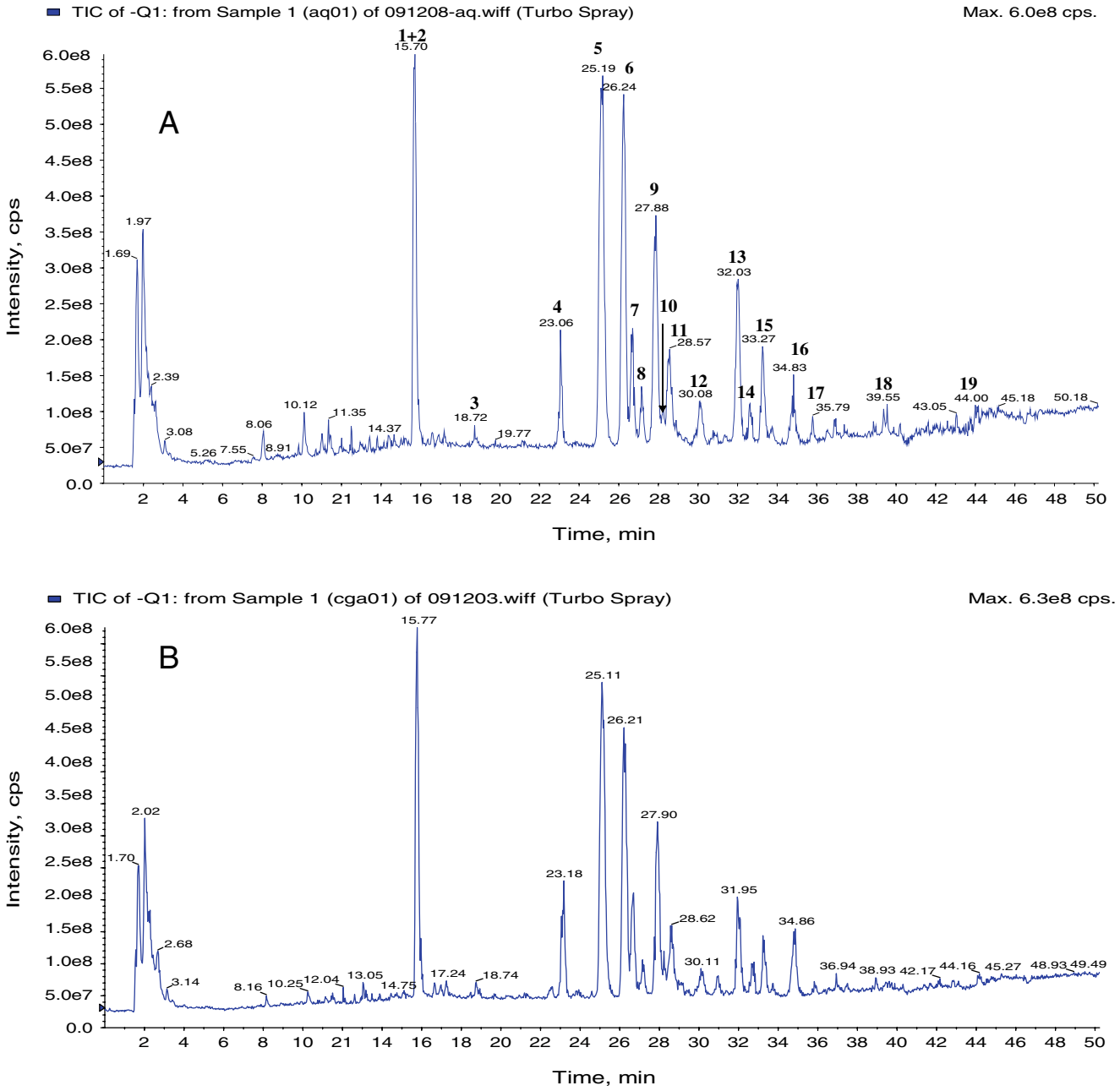


Fig. 2 Typical LC-MS chromatograms of *Panax quinquefolius* samples grown in the USA (**a**) and China (**b**). *1* ginsenoside R_{g1}, *2* ginsenoside R_e, *3* notoginsenoside R_2, *4* ginsenoside R_f, *5* ginsenoside R_{b1}, *6* ginsenoside R_{g2}, *7* ginsenoside R_c, *8* malonyl ginsenoside R_{b1}, *9 m/z* 955.5, *10* ginsenoside R_{b2}, *11* ginsenoside R_{b3}, *12* quinquenoside R_1, *13* ginsenoside R_d, *14 m/z* 793.5, *15* quinquenoside III or pseudo-ginsenoside RC_1, *16* gypenoside XVII, *17* quinquenoside III or pseudo-ginsenoside RC_1, *18* ginsenoside R_{g3}, *19* ginsenoside R_{h2}

Table 3 Ginsenosides observed in *Panax quinquefolius*

Name	Retention time (min)	Formula $[M-H]^-$	Measured mass $[M-H]^-$	Exact mass $[M-H]^-$	Mass error (ppm)
Ginsenoside R_{g1}	15.7	$C_{42}H_{71}O_{14}$	799.4855	799.4844	−0.71
Ginsenoside R_e	15.7	$C_{48}H_{81}O_{18}$	945.5426	945.5428	0.25
Notoginsenoside R_2	18.7	$C_{41}H_{69}O_{13}$	769.4730	769.4744	1.77
Ginsenoside R_f	23.1	$C_{42}H_{73}O_{14}$	799.4835	799.4844	1.79
Ginsenoside R_{b1}	25.2	$C_{54}H_{91}O_{23}$	1,107.5945	1,107.5957	1.16
Ginsenoside R_{g2}	26.2	$C_{42}H_{71}O_{13}$	783.4899	783.4900	0.15
Malonyl ginsenoside R_{b1}	26.5	$C_{57}H_{94}O_{26}$	1,193.5949	1,193.5961	−1.01
Ginsenoside R_c	26.7	$C_{53}H_{89}O_{22}$	1,077.5831	1,077.5827	−0.34
Ginsenoside R_{h1}	27.2	$C_{36}H_{61}O_9$	637.4325	637.4321	−0.62
Ginsenoside R_{b2}	28.2	$C_{53}H_{89}O_{22}$	1,077.5837	1,077.5827	0.90
Ginsenoside R_{b3}	28.6	$C_{53}H_{89}O_{22}$	1,077.5841	1,077.5827	0.93
Ginsenoside R_d	32.0	$C_{48}H_{81}O_{18}$	945.5420	945.5428	0.89
Quinquenoside III or pseudo-ginsenoside RC_1	33.3	$C_{50}H_{83}O_{19}$	987.5532	987.5534	0.21
Ginsenoside R_{g3}	39.6	$C_{42}H_{71}O_{13}$	783.4912	783.4900	−1.51
Ginsenoside R_{h2}	44.0	$C_{36}H_{61}O_8$	621.4376	621.4372	−0.65

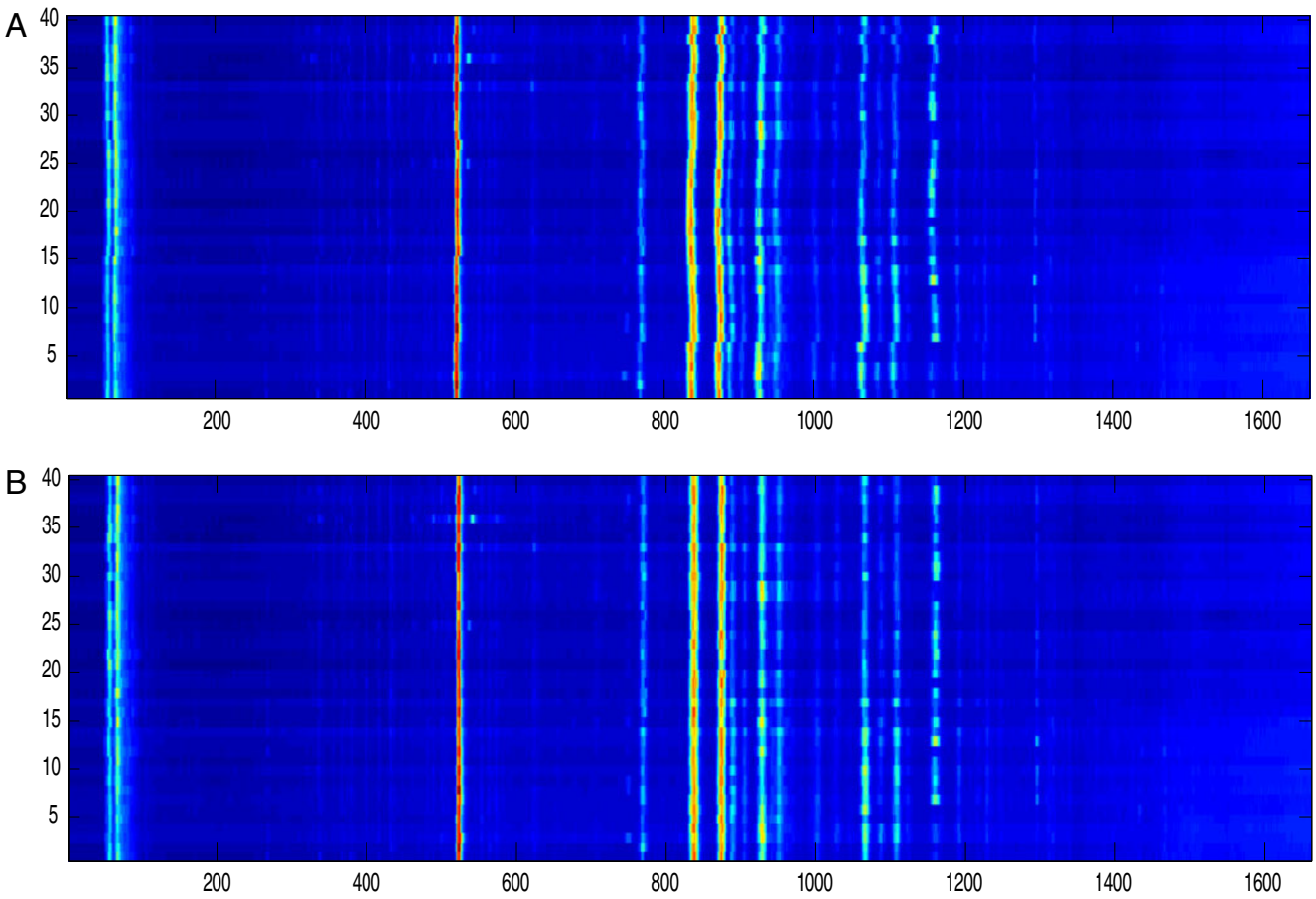


Fig. 3 The LC-MS fingerprints of the American ginseng sample before (**a**) and after (**b**) alignment

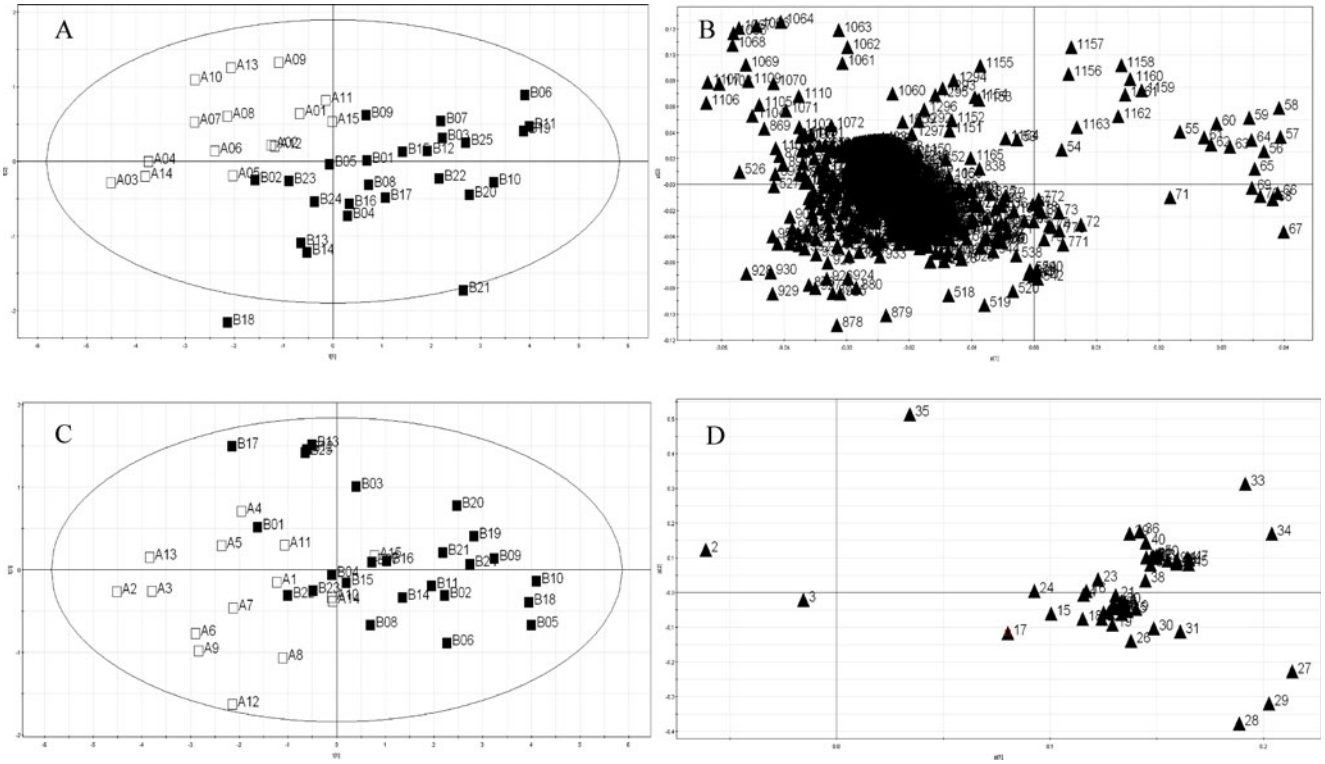


Fig. 4 PCA score plots (**a** before binning, **c** after binning) and loading plots (**b** before binning, **d** after binning) of the two kinds of American ginseng samples. *Open squares* American ginseng grown in Wisconsin, *filled squares* American ginseng grown in China

ards and literature reports. Ginsenosides R_{b1}, R_{b2}, R_{b3}, R_c, R_d, R_e, R_{g1}, R_{g2}, R_{g3}, and R_{h2} were confirmed with the reference standards and other ginsenosides were characterized according to previous studies in comparison with the retention behavior and accurate molecular weight, as listed in Table 3. It is very hard to distinguish between the ginseng samples grown in these two habitats by visual observation of the LC-MS chromatograms as the major components across samples are very similar.

Comparison of *P. quinquefolius* samples using chromatographic fingerprint datasets

An approach using chromatographic fingerprints in combination with PCA was used to examine the chemical differences between *P. quinquefolius* from the two sources. For a chromatographic fingerprint study, peak alignment is a crucial step for the subsequent chemometric analysis. The COW algorithm has shown great potential for alignment correction in chromatographic fingerprints. The method uses piecewise linear stretching or compression of segments, allowing relatively small changes in a segment length [47, 48] and alignment of the chromatograms as shown in Fig. 3. The aligned data were exported into Excel to form a 40× 1,662 data matrix (samples × data points). Mean centering and Pareto scaling were performed for data preprocessing prior to PCA. Separation between the two sets of *P. quinquefolius* samples was observed in the PCA score plot (Fig. 4a). It is very hard to interpret the loading plot (Fig. 4b) owing to the large number of data points involved. Each variable (data point) represents the loading of the total ion intensity at that retention time. To reduce the complexity of the dataset and increase the data interpretability, the original dataset was binned at 1-min intervals, which reduced the original matrix to a 40× 50 matrix. PCA was subsequently carried out on this simplified dataset and the separation was not significantly altered (Fig. 4c). A two-component PCA model cumulatively accounted for 78.1% of variation with 67.9% predictability Q^2. In the PCA score plot, some *P. quinquefolius* samples grown in China, e.g., B01 and B22, were chemically similar to the samples grown in the USA.

Finding the potential discriminatory ginsenosides using chromatographic fingerprint datasets

From the PCA loading plot (Fig. 4d), the variables 2, 3, 27, 28, 29, 33, 34, and 35 contribute to more than 50% variance for the population separation. Therefore, the major ions with retention times in these regions (Fig. 5) were extracted from the raw data file and the intensity of these ions was compared between the two sample sets with a *t* test.

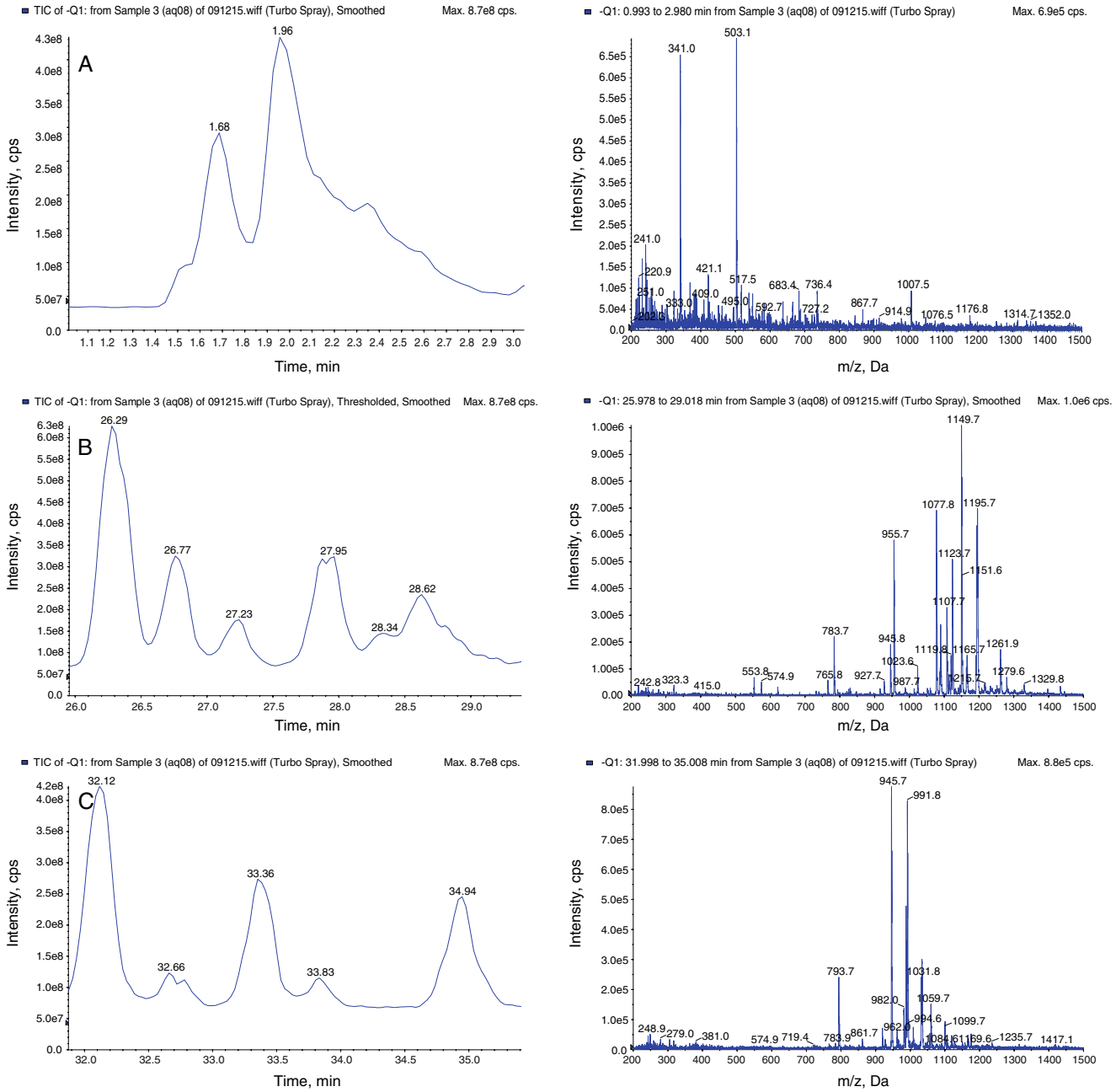


Fig. 5 Three discriminatory regions from the PCA loading plot and the corresponding full-scan mass spectra: **a** 1–3 min; **b** 26–30 min; **c** 32–35 min

The results suggested that the ions with *m/z* 945.8, 1,107.8, 1,193.5, 987.8, 1,077.8, 1,123.8, 981.7, 1,263.7, and 1,113.7 are significantly different ($P<0.05$) in the two sets of samples (Fig. 6). The ions at *m/z* 945.8, 1,107.8, 1,077.8, and 1,123.8 were confirmed as ginsenosides R_d, R_{b1}, and R_c and the formic acid adduct of ginsenoside R_c, using reference standards with retention times. The other ions were identified according to accurate mass measurements, MS/MS fragmentation characteristics, retention behaviors, and literature reports. Confidence in the tentative compound identification was enhanced with the sub-parts-per-million mass accuracy of the Thermo LC-Exactive MS system, together with the ring-plus-double-bond equivalent and isotope distribution. Using the ion at *m/z* 987.8, for example, the accurate mass measurement suggested the elemental composition is $C_{50}H_{83}O_{19}$ ([M-H]$^-$ 987.5532, 0.2 mDa, 0.2 ppm). As shown in Fig. 7, the MS/MS data suggested the loss of three glycosidic groups and one acyl group. The product ion at *m/z* 459.3851 corresponded to the deprotonated ion of the aglycone protopanaxadiol. From a ChemSpider (http://www.chemspider.com) library search, quinquenoside III or pseudo-ginsenoside RC_1 was found to be consistent with the mass

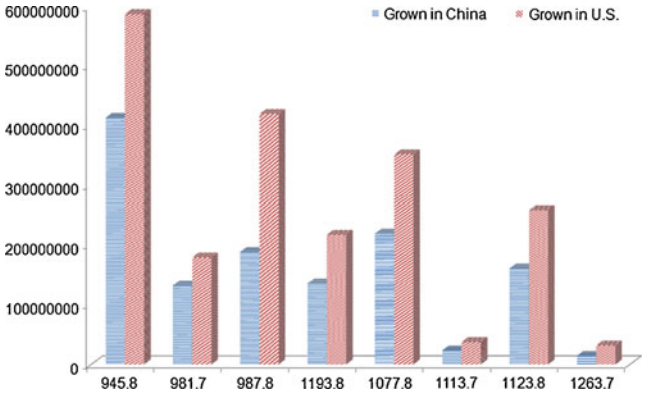


Fig. 6 The discriminatory ginsenosides in American ginseng grown in China and the USA

fragmentation behavior. So the ion at *m/z* 987.8 (retention time 33.3 min) was identified as quinquenoside III/pseudo-ginsenoside RC_1. Similarly, the ion at *m/z* 1,093.8 (retention time 26.5 min) was identified as malonyl ginsenoside R_{b1}. However, ions at *m/z* 981.7, 1,263.7, and 1,113.7 were not identified since no compounds with those molecular weights were reported in *P. quinquefolius*. A previous study reported that the majonoside isomer may be used as a marker to distinguish *P. quinquefolius* grown in China and the USA [22]. The authors showed two chromatograms: the one for *P. quinquefolius* grown in China had a large majonoside peak (peak 14 in [22]), whereas the one for *P. quinquefolius* grown in the USA did not have that peak at all. It is very hard to believe that the same species grown in two locations could have such a discrepancy in chemical composition. The results from our 40 chromatograms indicate no such discrepancy exists between *P. quinquefolius* grown in the two locations. The chromatograms for the two sets of samples are almost indistinguishable to the naked eye, and only with the assistance of the chemometric analysis could we discriminate one group from the other.

Comparison of *P. quinquefolius* samples using a chemometric approach

A fingerprint/PCA study was performed on the binned spectra data. Chemometric approaches were recently introduced to the quality control and assessment of botanical materials. The chemometric approach is different from the PCA approach in data preprocessing steps. For the LC-MS-based chemometric study, instead of using a 2D chromatographic fingerprint (total

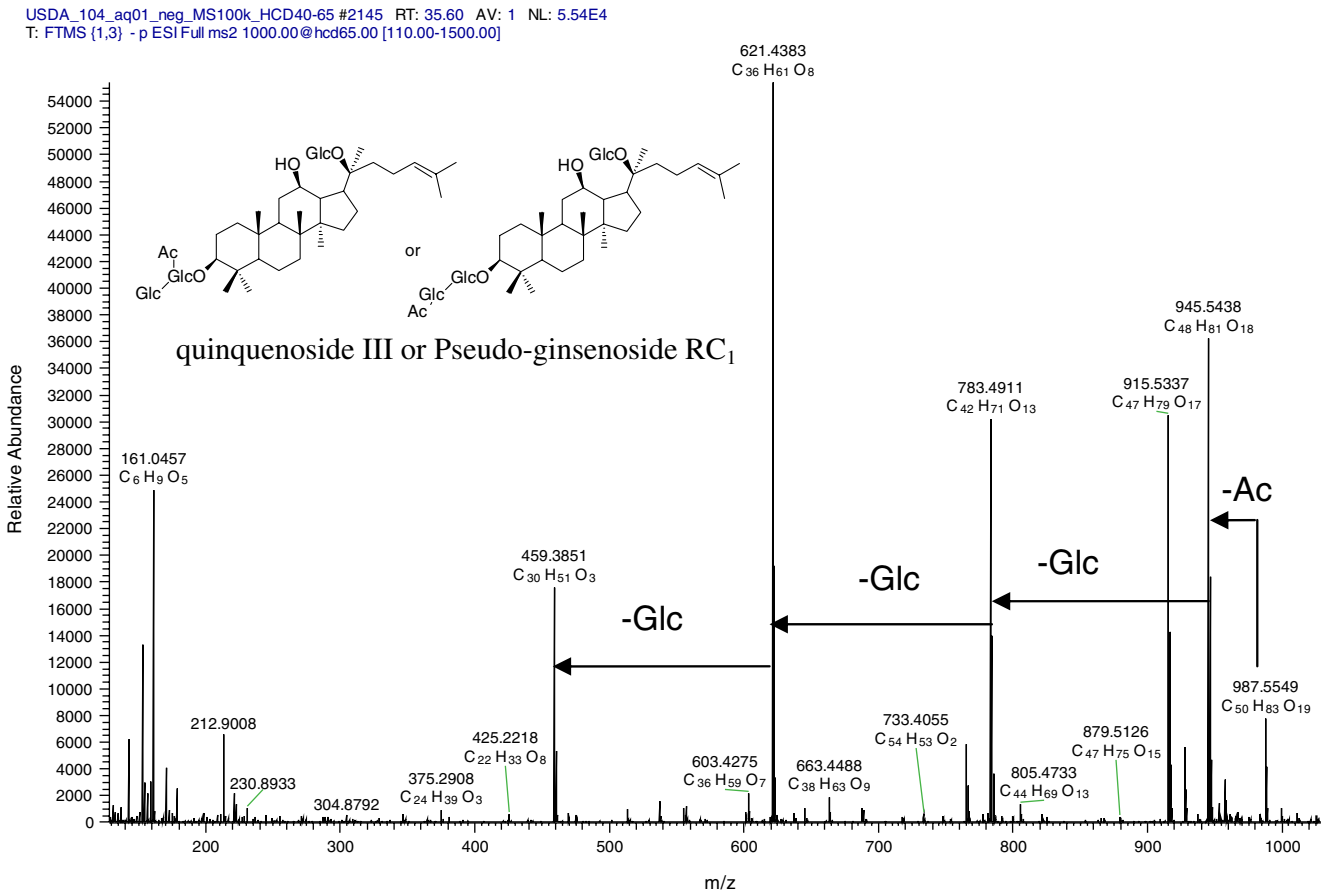


Fig. 7 The MS/MS fragmentation of $[M-H]^-$ 987.5549

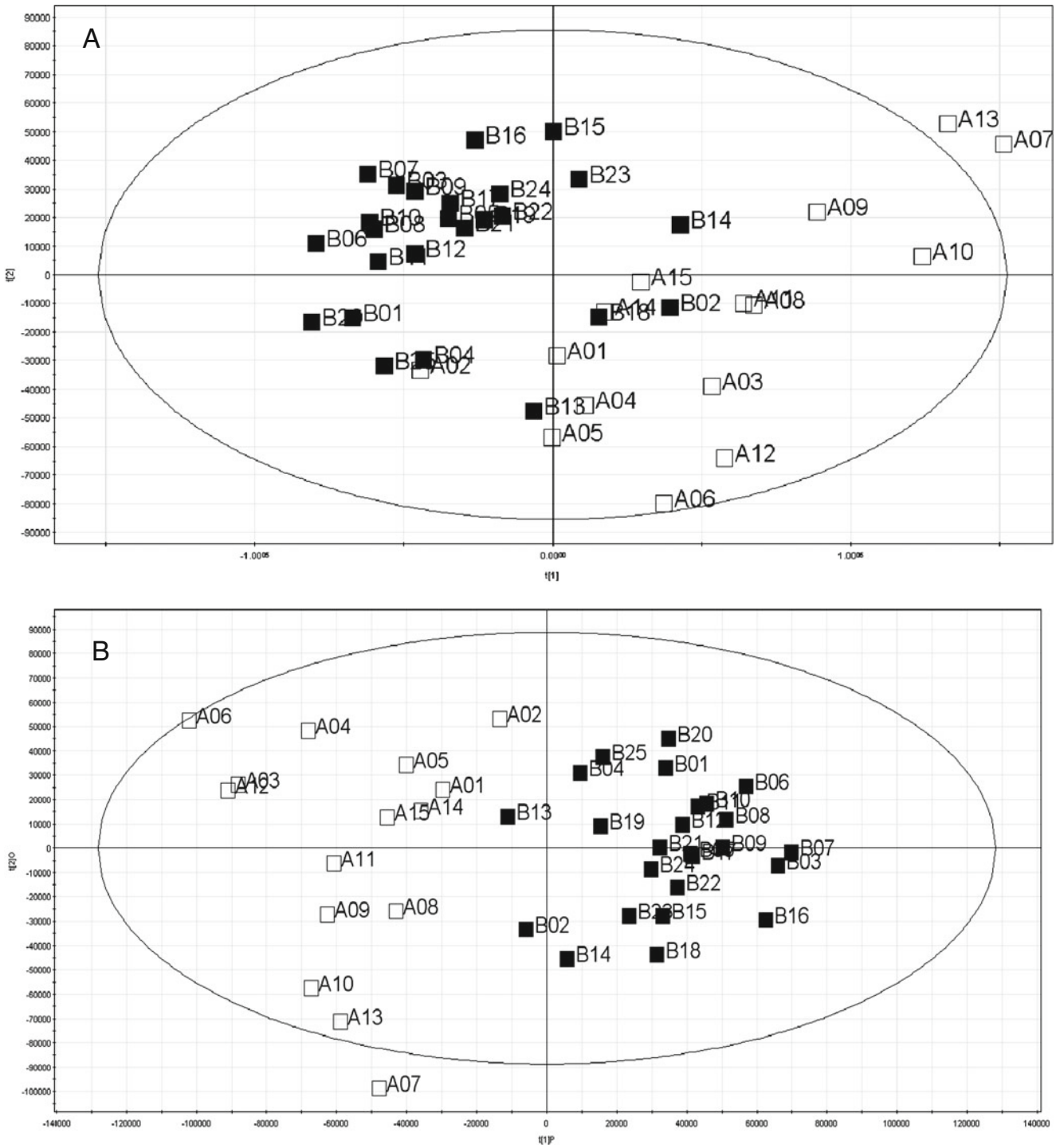


Fig. 8 The PCA score plot (**a**) and the OPLS-DA score plot (**b**)of the American ginseng grown in the USA and China. *Open squares* American ginseng grown in Wisconsin, *filled squares* American ginseng grown in China

ion intensity vs time), a 3D dataset (MS spectrum vs time) was used. The data preprocessing involves the peak detection, alignment, and denoising, which maximizes the useful information and minimizes the instrumental noise in the dataset. Then the 3D dataset was deconvoluted into 2D format using chemometric software. The final data were stored as *m*/*z*, retention time, and the ion intensity for that *m*/*z*. The variables for the subsequent multivariate statistical analysis can be traced back to the ions in the dataset by interpretation of the loading plot. Thus, potential biomarkers (ions that are responsible for the classification) can be determined.

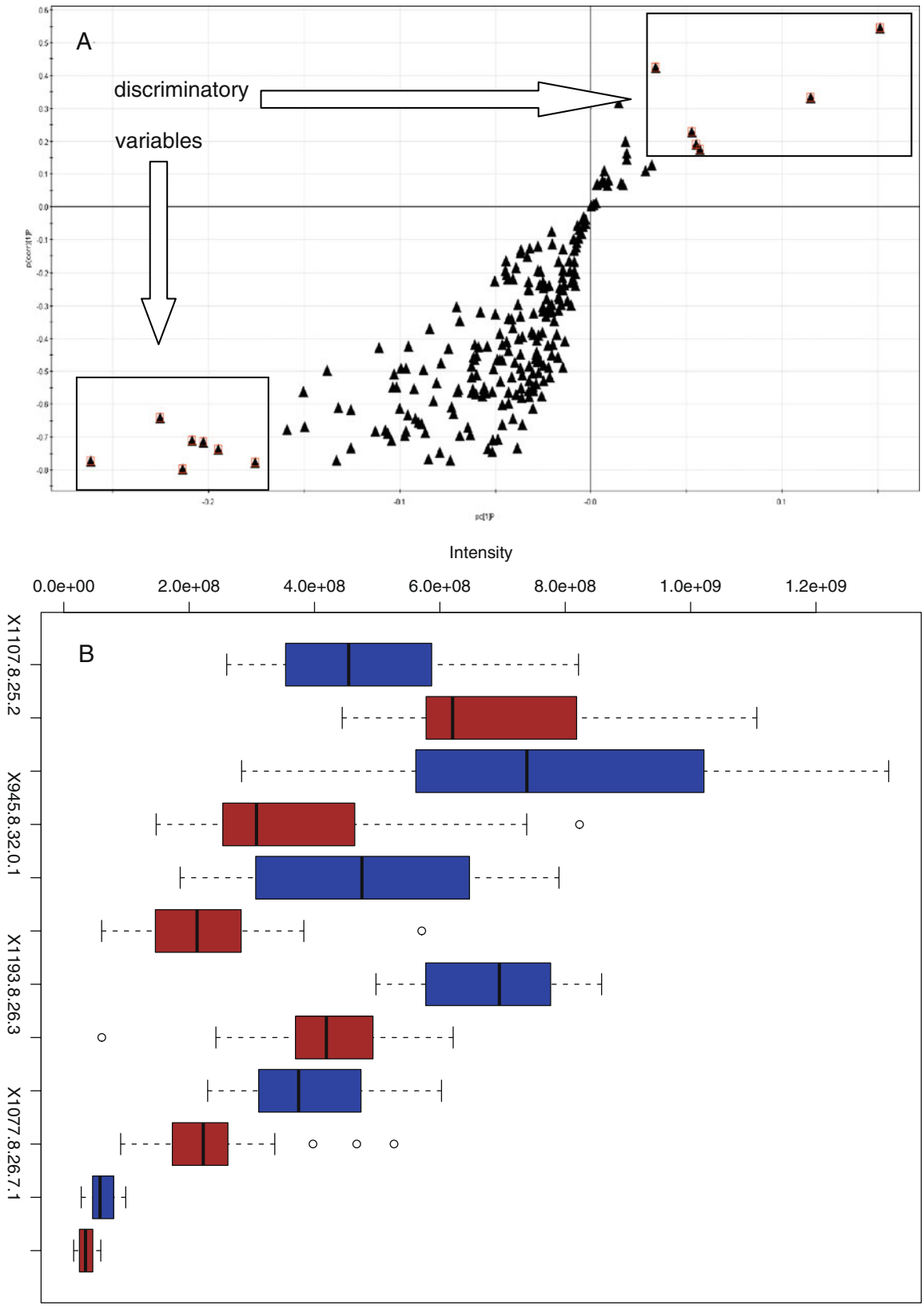


Fig. 9 The S-plot (**a**) and the box plot (**b**) showing the major ginsenoside variation in American ginseng grown in China (*blue*) and the USA (*brown*). *1* ginsenoside R_{b1}, *2* ginsenoside R_d, *3* quinquenoside III or pseudo-ginsenoside RC_1, *4* malonyl ginsenoside R_{b1}, *5* ginsenoside R_c, *6* unidentified

Table 4 The discriminatory variables identified from an S-plot

Variables (*m/z*, RT)	VIP value	Possible identification
945.8, 32.0	3.86	Ginsenoside R_d[a]
987.3, 33.3	3.12	Quinquenoside III or pseudo-ginsenoside RC_1
1,077.8, 26.7	2.34	Ginsenoside R_c[a]
1,107.8, 25.2	2.32	Ginsenoside R_{b1}[a]
1,193.7, 26.3	2.03	Malonyl ginsenoside R_{b1}
981.7, 32.0	1.54	Unidentified

RT retention time, *VIP* variable importance projection

[a] Identified with reference standards

XCMS [44], an open-source software package based on R, was used in this study for peak detection and alignment. After the XCMS processing, 234 peaks were found and subjected to multivariate analysis. Pareto scaling and mean centering was used for the data pretreatment before PCA. As shown in Fig. 8a, better separation in the PCA score plot was observed using this approach compared with using the simple chromatographic fingerprint/PCA approach. However, samples B02, B13, B14, and B18 fall into the grown in the USA group and sample A02 falls into the grown in China group. The first two principal components cumulatively accounted for 57.5% of the variation.

A supervised OPLS-DA statistical model was carried out to focus on the discrimination of the two groups. OPLS-DA can separate predictive from nonpredictive (orthogonal) variation, which can be used for discrimination differences between sample sets. The OPLS-DA approach has been demonstrated in some recent chemometric studies for the discrimination of botanical materials [41, 49, 50]. The OPLS-DA model resulted in one predictive and one orthogonal (1+1) component with a cross-validated predictive ability $Q^2(Y)$ of 76.7% and a total explained variance $R^2(X)$ of 76.9%. In addition, a permutation test (n=100) was carried out to evaluate the overfitting of the model. This test showed an R^2 intercept of 0.517 and a Q^2 intercept of 0.743, which validated the model. As shown in Fig. 8b, the OPLS-DA model successfully separated the two sample sets. No sample was misclassified during the permutation test.

Finding the potential discriminatory ginsenosides using chromatographic fingerprint datasets

An S-plot (Fig. 9a) was employed to identify the constituents that were responsible for the differentiation of the two sample sets. The S-plot was first introduced by Wiklund et al. [51]. It is proposed as a tool for visualization and interpretation of multivariate classification models, e.g., OPLS-DA model of two classes. The S-plot plots both the covariance and the correlation between the constituents and the modeled class designation based both on the contributions to the model and on their reliability. From the S-plot, the discriminatory variables were picked out by the variable importance projection (VIP) value. The variables with larger VIP values (more than 1) were more relevant for sample classification. The VIP values of the marker ions are listed in Table 4, showing six variables mainly contributed to the sample classification. The variables representing adduct, isotope, and fragment ions were filtered out manually. These discriminatory variables were identified and described in the previous section. The box plot (Fig. 9b) showed the differences of these discriminatory ginsenosides between the two groups of samples. A univariate *t* test was performed to confirm the discriminatory variables. All these six ginsenosides were significantly different between the groups (P<0.05). These compounds fulfilled the criterion of marker compounds because of the obvious variance in two kinds of American ginseng samples.

By using the chromatographic fingerprint and chemometric approaches, we found that chemical differences existed between *P. quinquefolius* samples cultivated in the USA and China. The ginsenosides responsible for the chemical differences were identified. *P. quinquefolius* grown in the USA contained more ginsenoside R_c, ginsenosideR_d, quinquenoside III//pseudo-ginsenoside RC_1, malonyl ginsenoside R_{b1}, and ginsenoside R_{b2}, but less ginsenoside R_{b1} than its counterpart grown in China.

Conclusion

In this study, both the traditional fingerprint approach and the traditional chemometric approach were successfully used for the differentiation of *P. quinquefolius* samples grown in the USA and China. Six ginsenosides were found to be mainly responsible for the chemical differences of *P. quinquefolius* samples grown in the two locations. The results suggested that ginsenosides R_{b1}, R_{b2}, R_c, and R_d, quinquenoside III/pseudo-ginsenoside RC_1, and malonyl ginsenoside R_{b1} may be used as marker compounds to distinguish *P. quinquefolius* grown in China and North America. Also, as demonstrated in this study, the chemometric approach is promising for quality control and assessment of botanical materials.

Acknowledgements This research was supported by the Agricultural Research Service of the US Department of Agriculture and an

Interagency Agreement with the Office of Dietary Supplements of the National Institutes of Health.

References

1. Dharmananda S, Feldman JF (1979) Plant Physiol 63:1049–1054
2. Yuan CS, Wang X, Wu JA, Attele AS, Xie JT, Gu M (2001) Phytomedicine 8:178–183
3. Liu D, Li B, Liu Y, Attele AS, Kyle JW, Yuan CS (2001) Eur J Pharmacol 413:47–54
4. Sloley BD, Pang PK, Huang BH, Ba F, Li FL, Benishin CG, Greenshaw AJ, Shan JJ (1999) J Psychiatry Neurosci 24:442–452
5. Yuan CS, Attele AS, Wu JA, Lowell TK, Gu Z, Lin Y (1999) Am J Chin Med 27:331–338
6. Mehendale SR, Wang CZ, Shao ZH, Li CQ, Xie JT, Aung HH, Yuan CS (2006) Eur J Pharmacol 553:209–214
7. Elmer GW, Lafferty WE, Tyree PT, Lind BK (2007) Ann Pharmacother 41:1617–1624
8. Zhou DL, Kitts DD (2002) Can J Physiol Pharmacol 80:1030–1033
9. Xie JT, Wang CZ, Zhang B, Mehendale SR, Li XL, Sun S, Han AH, Du W, He TC, Yuan CS (2009) Biol Pharm Bull 32:1552–1558
10. Jin Y, Hofseth AB, Cui X, Windust AJ, Poudyal D, Chumanevich AA, Matesic LE, Singh NP, Nagarkatti M, Nagarkatti PS et al (2010) Cancer Prev Res (Phila Pa) 3:339–347
11. Lee TK, O'Brien KF, Wang W, Sheng C, Wang T, Johnke RM, Allison RR (2009) Open Nucl Med J 1:1–8
12. Peralta EA, Murphy LL, Minnis J, Louis S, Dunnington GL (2009) J Surg Res 157:261–267
13. Wang X, Sakuma T, Asafu-Adjaye E, Shiu GK (1999) Anal Chem 71:1579–1584
14. Chan TW, But PP, Cheng SW, Kwok IM, Lau FW, Xu HX (2000) Anal Chem 72:1281–1287
15. Li W, Gu C, Zhang H, Awang DV, Fitzloff JF, Fong HH, van Breemen RB (2000) Anal Chem 72:5417–5422
16. Ji QC, Harkey MR, Henderson GL, Gershwin ME, Stern JS, Hackman RM (2001) Phytochem Anal 12:320–326
17. Ligor T, Ludwiczuk A, Wolski T, Buszewski B (2005) Anal Bioanal Chem 383:1098–1105
18. Li L, Luo GA, Liang QL, Hu P, Wang YM (2010) J Pharm Biomed Anal 52:66–72
19. Wang JR, Leung CY, Ho HM, Chai S, Yau LF, Zhao ZZ, Jiang ZH (2010) Chem Biodivers 7:975–983
20. Wang CZ, Aung HH, Ni M, Wu JA, Tong R, Wicks S, He TC, Yuan CS (2007) Planta Med 73:669–674
21. Sloley BD, Lin YC, Ridgway D, Semple HA, Tam YK, Coutts RT, Lobenberg R, Tam-Zaman N (2006) J AOAC Int 89:16–21
22. Ma X, Xiao HB, Liang XM (2006) Chromatographia 64:31–36
23. Schlag EM, McIntosh MS (2006) Phytochemistry 67:1510–1519
24. Zhong XK, Li DC, Jiang JG (2009) Curr Med Chem 16:3064–3075
25. Liang YZ, Xie P, Chan K (2004) J Chromatogr B Anal Technol Biomed Life Sci 812:53–70
26. Li S, Han Q, Qiao C, Song J, Lung Cheng C, Xu H (2008) Chin Med 28:3–7
27. Chen P, Ozcan M, Harnly J (2007) Anal Bioanal Chem 389:251–261
28. Harnly JM, Pastor-Corrales MA, Luthria DL (2009) J Agric Food Chem 57:8705–8710
29. Chen P, Song F, Lin LZ (2009) J AOAC Int 92:624–632
30. Chen P, Harnly JM, Lester GE (2010) J Agric Food Chem 58:4545–4553
31. Luthria DL, Lin LZ, Robbins RJ, Finley JW, Banuelos GS, Harnly JM (2008) J Agric Food Chem 56:9819–9827
32. Lin LZ, Harnly JM, Upton R (2009) Phytochem Anal 20:298–306
33. Fiehn O (2002) Plant Mol Biol 48:155–171
34. Daviss B (2005) Scientist 19:25–28
35. Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R (2002) Plant Cell 14:1437–1440
36. German JB, Roberts MA, Fay L, Watkins SM (2002) J Nutr 132:2486–2487
37. Hollywood K, Brison DR, Goodacre R (2006) Proteomics 6:4716–4723
38. Weckwerth W (2003) Annu Rev Plant Biol 54:669–689
39. Tarachiwin L, Ute K, Kobayashi A, Fukusaki E (2007) J Agric Food Chem 55:9330–9336
40. Pongsuwan W, Bamba T, Harada K, Yonetani T, Kobayashi A, Fukusaki E (2008) J Agric Food Chem 56:10705–10708
41. Ku KM, Choi JN, Kim J, Kim JK, Yoo LG, Lee SJ, Hong YS, Lee CH (2010) J Agric Food Chem 58:418–426
42. Son HS, Hwang GS, Kim KM, Kim EY, van den Berg F, Park WM, Lee CH, Hong YS (2009) Anal Chem 81:1137–1145
43. Lee EJ, Shaykhutdinov R, Weljie AM, Vogel HJ, Facchini PJ, Park SU, Kim YK, Yang TJ (2009) J Agric Food Chem 57:7513–7522
44. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) Anal Chem 78:779–787
45. Katajamaa M, Miettinen J, Oresic M (2006) Bioinformatics 22:634–636
46. Tomasi G, van den Berg F, Andersson C (2004) J Chemom 18:231–241
47. van Nederkassel AM, Daszykowski M, Eilers PH, Heyden YV (2006) J Chromatogr A 1118:199–210
48. Christin C, Smilde AK, Hoefsloot HC, Suits F, Bischoff R, Horvatovich PL (2008) Anal Chem 80:7012–7021
49. Choi JN, Kim J, Lee MY, Park DK, Hong YS, Lee CH (2010) J Agric Food Chem 58:4258–4267
50. Li SL, Song JZ, Qiao CF, Zhou Y, Qian K, Lee KH, Xu HX (2010) J Pharm Biomed Anal 51:812–823
51. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J (2008) Anal Chem 80:115–122