



In-depth profiling, characterization, and comparison of the ginsenosides among three different parts (the root, stem leaf, and flower bud) of *Panax quinquefolius* L. by ultra-high performance liquid chromatography/quadrupole-Orbitrap mass spectrometry

Hongda Wang^{1,2} · Chunxia Zhang^{1,2} · Tiantian Zuo^{1,2} · Weiwei Li^{1,2} · Li Jia^{1,2} · Xiaoyan Wang^{1,2} · Yuexin Qian^{1,2} · Dean Guo^{1,3} · Wenzhi Yang^{1,2}

Received: 21 July 2019 / Revised: 18 September 2019 / Accepted: 30 September 2019 / Published online: 15 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Despite *Panax quinquefolius* L. serving as a crucial source for food additives, healthcare products, and herbal medicines, unawareness of the metabolome differences among three parts (root, PQR; stem leaf, PQL; flower bud, PQF) seriously restricts its quality control. Ultra-high performance liquid chromatography/quadrupole-Orbitrap mass spectrometry (UHPLC/Q-Orbitrap-MS) was fully utilized to comprehensively identify and compare the ginsenoside compositions among PQR, PQL, and PQF. Metabolite profiling and characterization were performed by coupling reversed-phase UHPLC (a CSH C18 column) and improved untargeted data-dependent MS² acquisition in the negative mode. A novel vehicle, “Ginsenoside Sieve,” was proposed by developing fixed tolerance (± 10 mDa), discrete mass defect filtering (MDF) based on the m/z features of 499 known ginsenosides, which assisted in the screening of 71 (from 3453 ions), 89 (from 6842), and 84 (from 7369) target precursor masses for PQR, PQL, and PQF, respectively. The newly established data-dependent acquisition (DDA) approach exhibited 14% improvement in characterization of targeted components (using a PQL sample), and comparable performance in identifying the unknown, compared with conventional DDA. We could characterize 347 saponins (147 from PQR, 173 from PQL, and 195 from PQF), and 157 thereof not ever-isolated from the *Panax* genus. These potentially new saponins have 63 unknown masses. Subsequent untargeted metabolomics analysis unveiled 20 marker saponins, among which m-Rb1, Rb1, Ro, m-Rb2, and m-Rb1 isomer are the most important diagnostic for differentiating the three parts. Conclusively, the established improved DDA represents a potent ginsenoside characterization strategy, and the results obtained in this work would benefit better quality control of *P. quinquefolius*.

Keywords *Panax quinquefolius* · Ginsenoside · Q-Orbitrap · Mass defect filter · Metabolomics

Hongda Wang, Chunxia Zhang and Tiantian Zuo contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00216-019-02180-8>) contains supplementary material, which is available to authorized users.

✉ Dean Guo
daguo@simm.ac.cn

✉ Wenzhi Yang
wzyang0504@tjutcm.edu.cn

¹ Tianjin State Key Laboratory of Modern Chinese Medicine, Tianjin Key Laboratory of TCM Chemistry and Analysis, Tianjin University of Traditional Chinese Medicine, 312 Anshanxi Road, Tianjin 300193, China

² Tianjin Key Laboratory of TCM Chemistry and Analysis, Tianjin University of Traditional Chinese Medicine, 312 Anshanxi Road, Tianjin 300193, China

³ Shanghai Research Center for Modernization of Traditional Chinese Medicine, National Engineering Laboratory for TCM Standardization Technology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 501 Haik Road, Shanghai 201203, China

Introduction

Systematic elucidation of the chemical substances of traditional Chinese medicine (TCM) is a premise in exploring the therapeutic basis and elaborating the quality standards to promote its modernization and globalization [1]. Benefiting from the rapid development of chromatography/mass spectrometry, efficient, comprehensive, and accurate profiling and characterization of the metabolites from plants or biosamples have been achievable [2]. One tendency, discerned in LC-MS-based herbal component analysis, is “systematicness” that aims to separate and identify as many components as possible, and the other is “intelligentization” which is featured by significantly elevating efficiency and reproducibility due to the use of computational vehicles [3]. Powerful online chromatographic separation (by means of versatile separation mechanisms or multi-dimensional liquid chromatography [4, 5]), enhanced MSⁿ scan or diverse dissociation patterns [6], and programmed data processing [7, 8] have been reported in support of the in-depth metabolite characterization. To date, it can be accomplishable to separate and characterize more than 500 compounds from a single medicinal herb [9], which demonstrates an unimaginable potency favoring the comprehensive elucidation of herbal components and drives lead compounds discovery from nature.

Untargeted profiling strategies are preferably utilized to deconvolute the metabolome composition of medicinal herbs without the need of any pre-knowledge. Data-dependent acquisition (DDA) and data-independent acquisition (DIA) approaches have been extensively utilized to perform untargeted metabolite characterization. In the former mode (e.g., Auto MS/MS, Fast DDA, and Full MS/dd-MS², etc.), full-scan spectrum is recorded followed by intensity-ranking triggered automated MSⁿ acquisition [10]. DIA (typically including MS^E, AIF, and All Ions, etc.) can alternatively record the full-scan MS¹ and high-energy MS² data without the selection of precursor ions by quadrupole [11]. Despite no MS² information is missing by DIA, precursor-product ion matching is necessary, prior to MS² data interpretation, which can be achieved by commercial software or in-house developed algorithm [12, 13]. DDA data comparatively are easy to interpret, but its coverage of the interested components may be restricted when facing very complex chemical matrix, even if dynamic exclusion (DE) or the selection of more top-*n* precursors can be set selectively. In addition, high-resolution mass spectrometry (HRMS)-enhanced scan approaches have been reported which could improve the coverage in untargeted metabolite characterization. On one hand, precursor ion list (PIL) can be involved in DDA to improve the characterization of those co-eluting minor metabolites [14]. In-house library [15], neutral loss filtering [16], polygonal or dynamic-range mass defect filtering (MDF) [17] have been developed to generate a PIL containing a collection of target masses. On the

other hand, specific adducts or fragmentation feature can be utilized to enable the straightforward profiling of the targeted structures, such as the modified metabolites, malonylginsenosides.

Untargeted metabolomics renders a powerful analytical tool for the authentication of TCMs that involve similar chemical substances, by which the whole metabolome difference is evaluated to compare different species [18], parts [19], geographic origins [20], and the growing ages [21]. Multiple species from the genus of *Panax* L. exhibiting the tonifying effects to humans are extensively consumed as herbal medicines, healthcare products, cosmetics, etc., among which *P. ginseng* C.A. Meyer (Asian ginseng), *P. quinquefolius* L. (American ginseng), and *P. notoginseng* (Burk.) F.H. Chen (Sanchi ginseng) are the most reputable worldwide [22]. The saponins from the *Panax* genus (ginsenosides) represent a class of rich, specific, and biologically active substances closely associated with the therapeutic effects on the cardiovascular and immune systems [23]. The characteristic structure features of ginsenosides can facilitate a rapid characterization from various matrices by LC-MS. However, the complexity of herbal components (co-existing of the primary and secondary metabolites with wide spans of molecular weight, polarity, and content [1, 15, 24, 25]) and the ion species generated in ESI is far from what we imagine, which can result in a low coverage of DDA in untargeted metabolite characterization. The ESI full-scan spectrum for ginsenosides is complicated due to the presence of dimers, various adducts, and even doubly charged ions for ginsenosides [19]. Even for a well-resolved component, ginsenoside Rb1 (C₅₄H₉₂O₂₃), its MS¹ spectrum, displayed diversified ion species, such as the genuine deprotonated ion (*m/z* 1107), chlorine-adduct (*m/z* 1143), FA-adducts (*m/z* 1153), doubly charged bi-deprotonated ion (*m/z* 553) and the adducts (*m/z* 576, 599), and doubly charged dimeric FA-adduct (*m/z* 1131; see Electronic Supplementary Material (ESM) Fig. S1). In addition, medium intensity of ions with a mass difference of approximately 0.5 Da (*m/z* 1108.1012), together with the isotope peak (*m/z* 1109.1039), was observed. In this context, more powerful DDA or DIA strategies are in great urgency aiming to improve the coverage in untargeted metabolite characterization.

Simultaneous monitoring of multiple chemical markers is a practical solution to authenticating TCM species as the extractives or compound formulae [26]. To exactly identify the precious TCM species (such as American ginseng) and discern the adulterant (e.g., the substitution by illegal species or parts that contain similar bioactive ingredients) is an important aspect in quality control of the TCM formulae. Despite the chemical difference between *P. quinquefolius* and the congeneric species has been widely reported by untargeted metabolomics approaches [16, 19, 27, 28], it remains unclear regarding the ginsenoside difference among the different parts of *P. quinquefolius* (root,

PQR; stem leaf, PQL; and flow bud, PQF), which urgently needs to be solved to better quality control of American ginseng. In the current work, we fully utilized the UHPLC/Q-Orbitrap-MS (ultra-high performance liquid chromatography/quadrupole-Orbitrap mass spectrometry) platform and untargeted metabolomics workflows to unveil the ginsenoside difference among three different parts of *P. quinquefolius* (PQR, PQL, and PQF; Fig. 1). Firstly, ginsenosides from PQR, PQL, and PQF were comprehensively characterized by a PIL-included Full MS/dd-MS² approach with DE (dynamic exclusion) and the IIPO (*If idle-pick others*) function enabled, after efficient reversed-phase UHPLC separation. A novel ginsenoside searching vehicle, namely “Ginsenoside Sieve,” was proposed by developing a fixed tolerance, discrete MDF, based on the *m/z* features (integer mass VS mass defect) of all the known ginsenosides. It could screen target precursors from the negative full-scan MS¹ data to yield the PIL for each sample, which was predefined in Full MS/dd-MS² to guide the negative high-energy collision-induced dissociation-MS² acquisition (HCD-MS²) for PQR, PQL, and PQF, respectively. Secondly, standardized metabolomics workflows, involving data acquisition, data processing, and pattern recognition chemometrics using the 80% of the multi-batch samples (36 batches) were established in exploring the potential marker compounds. Thirdly, a few most important markers were used to establish an artificial nerve network (ANN) model for identifying the remaining 20% of samples. The current work can deeply

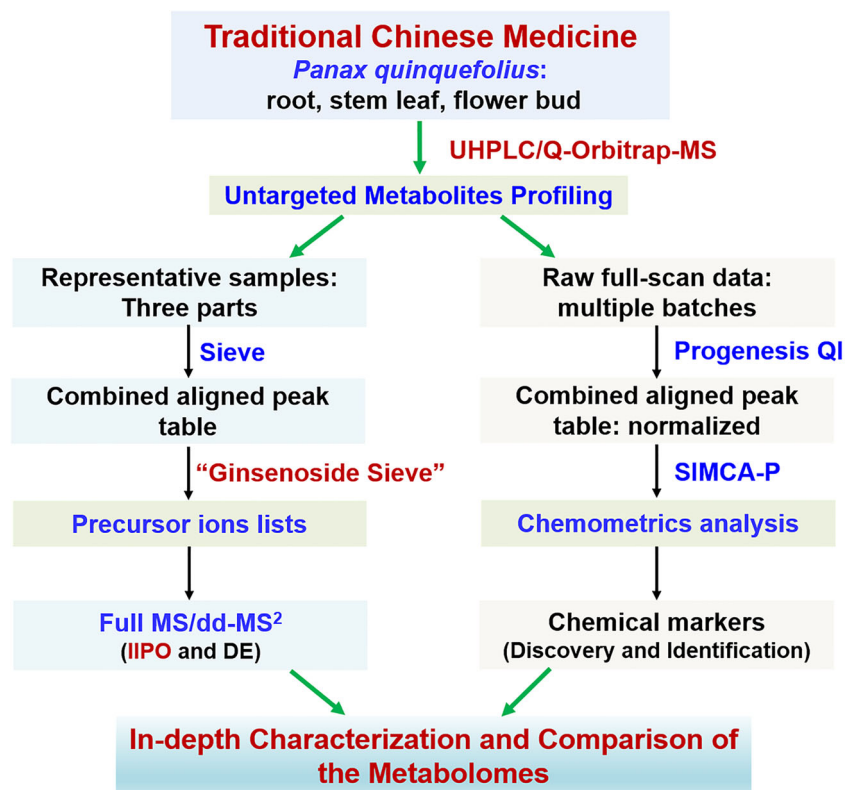
clarify the chemical compositions and unveil the marker compounds enabling the differentiation of three parts of *P. quinquefolius* (PQR, PQL, and PQF), which is crucial to the better quality control of American ginseng the related compound formulae.

Materials and methods

Chemicals and reagents

Fifty-two compounds, either isolated from the root of *P. notoginseng* or purchased from Shanghai Standard Biotech. Co., Ltd. (Shanghai, China), were used as the reference compounds in this work. The chemical structures and detailed information have been offered in Fig. 2 and ESM Table S1, respectively. Acetonitrile (Fisher, Fair lawn, NJ, USA) and formic acid (ACS, Wilmington, USA) were of HPLC grade. Ultra-pure water (18.2 MΩ cm at 25 °C) was in-house prepared using a Milli-Q water purification system (Millipore, Bedford, MA, USA). Information for the drug materials of three parts of *P. quinquefolius* (fifteen batches for each part) has been offered in ESM Table S2. Their identification was performed carefully according to Flora of China (frps.eflora.cn) and fingerprint comparison with the literature. Voucher specimens were

Fig. 1 An overall flow chart for the in-depth characterization and comparison of the metabolome differences among three different parts of *Panax quinquefolius* by the advanced UHPLC/Q-Orbitrap-MS platform



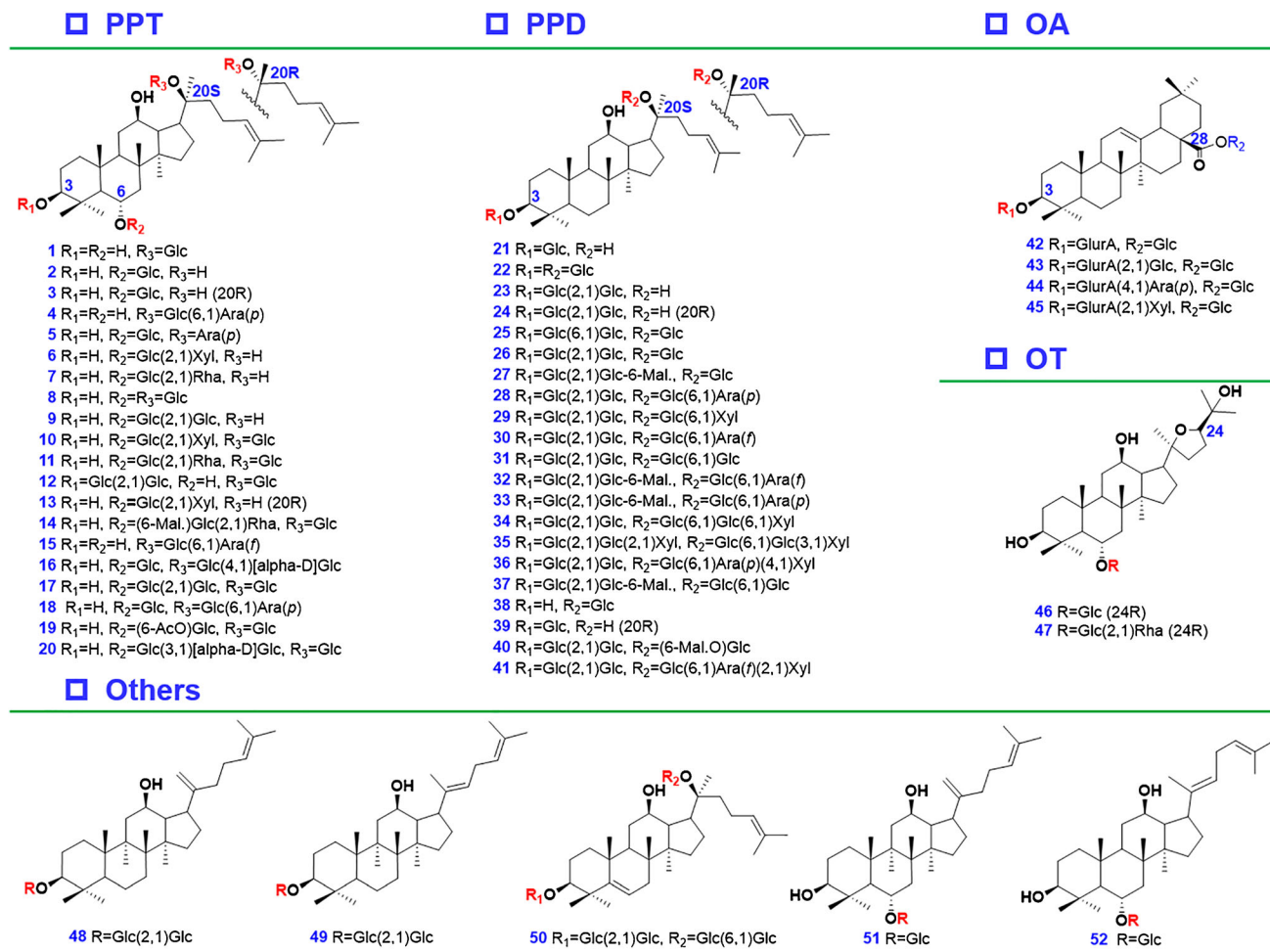


Fig. 2 Chemical structures of 52 ginsenoside reference compounds

deposited at the authors' laboratory in Tianjin University of Traditional Chinese Medicine (Tianjin, China).

Preparation of the solution of sample

An easy-to-implement ultrasound-assisted extraction method, by referring to our previous report [16], was utilized. In detail, 0.1 g accurately weighed fine powder (< 40 mesh) of each sample was soaked in a 15-mL centrifuge tube containing 6 mL 70% (v/v) methanol, and then extracted in a water bath (30 °C) with ultrasound assistance for 1 h. The extract was centrifuged at 5000 rpm for 10 min, and the supernatant was transferred into a 10-mL volumetric flask, which was further diluted to the constant volume with 70% methanol. The liquid suffered from centrifugation at 14,000 rpm for 10 min, leading to the supernatant used as the test solution (concentration 10 mg/mL of the drug material). A quality control (QC) sample was prepared by pooling the equal volume of each test solution to monitor the system stability. Three samples of PQR/PQL/PQF were prepared at higher concentration of 20 mg/mL for the first step of multicomponent characterization.

UHPLC/Q-Orbitrap MS

Efficient chromatographic separation was achieved on an Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Waltham, MA, USA) configured with a CSH C18 column (2.1 × 100 mm, 1.7 μm) kept at 30 °C. A binary mobile phase, containing 0.1% formic acid in water (A) and acetonitrile (B) ran according to a gradient program: 0–2 min, 15–20% (B); 2–12 min, 20–30% (B); 12–21 min, 30–31% (B); 21–25 min, 31–35% (B); 25–32 min, 35–40% (B); 32–34 min, 40–95% (B); and 34–36 min, 95% (B). A flow rate of 0.3 mL/min was set, and the injection volume was 3 μL.

High-resolution MS data (*profile*) were recorded on a Q Exactive™ hybrid Q-Orbitrap mass spectrometer in the negative ESI mode. The source parameters were as follows: spray voltage, −3.5 kV; capillary temperature, 350 °C; gas temperature, 250 °C; sheath gas flow rate, 35 arbitrary units; aux gas flow rate, 10 arbitrary units; and sweep gas flow rate, 0 arbitrary unit. In-source CID of 0 eV was set.

In-depth ginsenoside characterization of PQR, PQL, and PQF was performed (using R-2, SL-3, and F-15) by a PIL-

including Full MS/dd-MS² approach with the IIPO function enabled. The Orbitrap analyzer scanned over a mass-to-charge ratio (m/z) range of 250–1500 at a resolution of 70,000 in full-scan MS¹ and a resolution of 13,500 in dd-MS². Automatic gain control (AGC) target for MS¹ and MS² was set at 1e6 and 1e5, respectively. Maximum injection time (IT) was separately defined at 100 ms and 50 ms for MS¹ and MS². Dynamic exclusion (DE) was also enabled with an exclusion time of 6.0 s. Isolation width was set at 6.0 m/z . Specific PIL was set for PQR, PQL, and PQF, with the detailed information offered in the following subsection. The ions showing top 5 most intense abundance were automatically selected to trigger the HCD-MS² fragmentation. The ions recorded in full-scan data showing an m/z value within the PIL by allowing a variation of 10 mDa and the intensity higher than the threshold (1.6e5) were given the highest priority to trigger HCD-MS². Mixed NCE (normalized collision energy) at 25/30/35 V was adopted to enable more balanced MS² fragmentation from high-mass to low-mass regions. An Apex trigger of 1–3 s was set to record the MS² information of precursors in higher abundance [17]. Data acquisition and processing were performed by Xcalibur 4.1 (Thermo Fisher Scientific). Elemental composition prediction of the detected components was based on the following settings: elements in use, C 0–80, H 0–130, O 0–60; mass tolerance, < 5 ppm; RDBeq (ring double-bond equivalent) ≥ 7 .

In untargeted metabolomics analysis of three parts, the UHPLC/Q-Orbitrap-MS condition was almost identical to that used in the multicomponent characterization as described above, but only the full-scan MS¹ data recorded. In the whole analysis batch, the PQR/PQL/PQF samples were injected randomly, and the QC sample was injected after the analysis of every six samples to monitor the system stability.

Establishment of the “Ginsenoside Sieve”

A rigid MDF tool with discontinuous integer mass and a fixed mass defect variation range (10 mDa) was elaborated, based on the m/z values of 499 ginsenosides that have been isolated from the *Panax* genus up to 2017. First, these 499 saponins corresponded to 169 different masses after removing the repeated ones. To the acidic ginsenosides such as carboxyl-containing OA-type and malonylginsenosides, the m/z value consistent with the deprotonated molecular mass ($[M-H]^-$) was used, while in the case of the neutral saponins (carboxyl-free PPD, PPT, and the others), their FA-adducts were considered [29]. By using the mod function of Excel, the integer mass and decimal mass were discerned. The dynamic variation range, {Decimal mass–10 mDa, Decimal mass + 10 mDa}, together with the integer mass served as the “Ginsenoside Sieve” to orthogonally screen the target precursor ions from the full-scan raw data processed by Sieve v2.2

SP2, which, ultimately, generated the PILs for PQR, PQL, and PQF, respectively.

Settings of three DDA approaches for performance comparison

Three DDA-MS² methods, involving PIL-including Full MS/dd-MS² with IIPO enabled (**M1**), PIL-including Full MS/dd-MS² (**M2**), and Full MS/dd-MS² (**M3**), were compared using a PQL sample (SL-3) to show their differentiated performance in untargeted characterization of ginsenosides. Parameter setting for **M1** had been depicted in *Subsection 2.3* and *2.4*. The settings for **M2** and **M3**, which were different from those of **M1**, are summarized as below: (i) **M2**: PIL included, but IIPO not enabled; (ii) **M3**: PIL not included and IIPO not enabled.

Untargeted metabolomics for the simultaneous differentiation of ginsenosides from PQR, PQL, and PQF

Untargeted metabolomics analysis of the ginsenosides from three parts of *P. quinquefolius* was performed on the UHPLC/Q-Orbitrap-MS platform based on the negative full-scan MS¹ data. Processing of the metabolomics data, including peak alignment and peak picking, was conducted using Progenesis QI 2.1 (Waters Corporation). Isotope and adduct fusion were applied. The adduct forms, involving $[M-H]^-$, $[M+Cl]^-$, $[M-H+FA]^-$, $[M-2H+FA]^-$, $[M-2H]^{2-}$, $[2M-H]^{2-}$, and $[M-2H+2FA]^{2-}$, in the negative mode, were selected or in-house edited. A data matrix, involving the information of t_R , m/z , and normalized abundance, was finally obtained. The “80% rule” and “30% variation” rules were utilized to filter the variables. The finally obtained variables (using 80% of the samples: 12 batches for each part) were imported into the SIMCA-P 14.1 software (Umetrics, Umea, Sweden) for pattern recognition chemometrics by PCA (principle component analysis) and OPLS-DA (orthogonal partial least squares-discriminant analysis). A VIP cutoff at 5.0 was set to filter the potential ginsenoside markers. An intelligent discriminant tool, ANN, was established based on five most important marker compounds (including m-Rb1, Rb1, Ro, m-Rb2, and m-Rb1 isomer) using SPSS 23.0 (IBM SPSS Statistics, Chicago, IL, USA), to identify the remaining unknown *P. quinquefolius* samples [19, 30].

Results and discussion

Developing a PIL-involving Full MS/dd-MS² method enabling the comprehensive ginsenoside profiling

A PIL-including Full MS/dd-MS² approach with IIPO enabled was developed, aiming to enable the in-depth profiling

and characterization of ginsenosides from three parts of *P. quinquefolius*. The most important issue involved was to establish the PILs specific for the analytes. In the current work, we firstly proposed a new ginsenoside searching vehicle, “Ginsenoside Sieve,” to screen known ginsenosides from complex herbal matrix. We utilized discrete MDF because of its high higher precision than the fixed, linear, and even polygonal algorithms [17]. In the first step, the mass defect range was carefully defined.

Two options are alternative when applying MDF to filter target components from a complex chemical matrix: ppm or mDa. The use of ppm setting in MDF can enable mass-dependent dynamic variation range of mass defect [17], while the selection of mDa renders a fixed variation range [31]. In this work, to examine the relationship between the determined mass error and the integer mass, we selected multiple components with 60 different masses from a PQF sample (F-1) and plotted a 2D scatter diagram by the determined mass error vs. integer mass (see ESM Fig. S2). Evidently, the mass error for these components in general was < 2.5 mDa. A small linearity correlation coefficient ($R^2 = 0.1403$) was calculated, which could indicate a negative correlation between the determined mass error and the molecular mass for ginsenosides. Considering the mass range of ginsenosides from PQR, PQL, and PQF (500–1500 Da), a fixed variation range, 10 mDa (equal to 10 ppm calculated for a ginsenoside with M.F. 1000), was utilized, which can enable the rigid filtering and meanwhile has the potential to discover unknown structure analogs. “Ginsenoside Sieve” was finally established based on MDF allowing a 10 mDa of mass defect variation (Fig. 3a). More importantly, we could primarily conclude the mass defect values for ginsenosides (based on the $[M-H]^-$ precursors) positively correlated to the molecular mass. As we know, increase on the molecular weight of ginsenosides largely depends on the extension of sugars (mono- to hexaglycosides). The sugars composing ginsenosides comply with an approximate constant H/O ratio ($C_6H_{10}O_5$ for Glc; $C_5H_8O_4$ for Xyl/Ara; $C_6H_{10}O_4$ for Rha; $C_6H_8O_6$ for GlurA), and accordingly, an extension of a sugar yields the variation of 50.5 mDa for Glc, 40.4 mDa for Xyl/Ara, 55.6 mDa for Rha, and 30.2 mDa for GlurA, respectively.

The elaborated MDF tool subsequently was applied to screen target masses from the precursor lists for the generation of individual PIL for PQR, PQL, and PQF. The negative full-scan MS^1 data of representative PQR (R-2), PQL (SL-3), and PQF (F-15) samples were processed by the Sieve software yielding three precursor ion tables. The scatter plots of all precursor ions detected from PQR, PQL, and PQF are exhibited in Figs. 3b–d. The metabolomes of three different parts showed impressive complexity but with remarkable discriminatory features. PQL (Fig. 3c) and PQF (Fig. 3d) in general contain more high-mass saponins (M.F. > 1300) and low-mass components (M.F. < 600) than PQR (Fig. 3b). Further filtering

of these three chemical matrices led to the generation of three PILs involving 71 (PQR-PIL), 89 (PQL-PIL), and 84 (PQF-PIL) masses, respectively, which were included in the Full MS/dd- MS^2 approaches. Notably, this improved DDA strategy, by incorporating MDF-resultant PILs and IIPO function, can facilitate the sensitive characterization of the target components and simultaneous identification of those unknown ones via one injection analysis.

Superiority of the established Full MS/PIL/dd- MS^2 approach with IIPO enabled (**M1**) was demonstrated by comparing with another two DDA methods, Full MS/PIL/dd- MS^2 with IIPO disabled (**M2**) and Full MS/dd- MS^2 in which PIL was not included and IIPO disabled (**M3**). Due to the same top-n setting, five MS^2 spectra were recorded after one full scan in all three methods. Differences among **M1**–**M3** are embodied in the criteria to automatically trigger the MS^2 acquisition. In **M1**, targeted masses found in the full-scan spectrum were given the highest priority (despite the intensity was not ranked among top 5), and if none was hit, the untargeted masses could also trigger the MS^2 fragmentation. In the case of **M2**, only the targeted masses hit from MS^1 scan were fragmented, and if idle, the instrument continued to perform full scans. For **M3**, the instrument regularly recorded one full scan followed by five MS^2 fragmentations of those top 5 most intense ions, whether they had targeted or untargeted masses. Differentiated settings of these three DDA methods aforementioned led to the remarkably different performances which can be summarized in two aspects. First, due to the differentiated MS^2 acquisition rules, the time distributed to MS^1 and MS^2 between **M2** and the others was largely different (MS^2 acquisition occurs only when target masses are found in MS^1 for **M2**), which thus resulted in the acquisition of much less MS^2 spectra (2169-**M2** VS 11943-**M1** and 11,776-**M3**; effective separation time from 2 to 36 min) but much more full scans (6034-**M2** VS 3231-**M1** and 3297-**M3**) recorded by **M2**. In another word, the capacity in recording MS^2 spectra between **M1** and **M3** (conventional DDA setting) was almost the same, and the only difference was the highest priority given to target masses in **M1**. Second, the capacities of profiling target masses and untargeted masses among **M1**–**M3** were significantly different. Numbers of the components with the targeted masses enabled by **M1** (290) were larger than those of **M2** (263) and **M3** (254), while the components having untargeted masses characterized by **M1** (1978) and **M3** (2001) were much higher than **M2** (76) (Fig. 4a). Comparatively, **M1** was the most potent in characterizing the compounds with a target mass, which also showed comparable performance for the characterization of those untargeted components. The superior capacity of **M1** over the other two, in sensitive characterization of the targeted compounds could be evidenced from the major components (Fig. 4b). An illustration diagram using a target mass at m/z 887.50 could in detail exhibit from which components these differences were generated (Fig. 4c). These

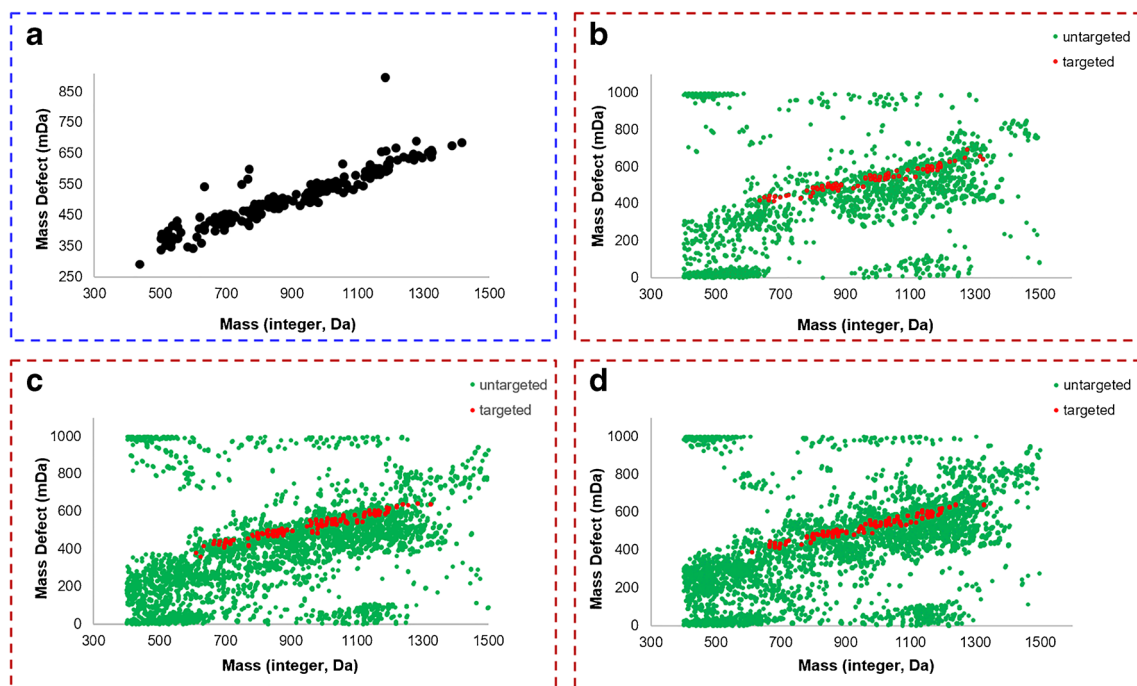


Fig. 3 Development of a discrete, fixed variation range MDF vehicle by plotting the integer mass (horizontal axis, Da) versus mass defect (vertical axis; mDa) of 169 masses representing 499 known ginsenosides isolated

from the *Panax* genus (a), and its application to PQR (b), PQL (c), and PQF (d) for screening the target precursors (in red dots)

evidences testified that, the elaborated Full MS/PIL/dd-MS² approach with IIPO enabled is a very powerful, improved DDA strategy that can facilitate the sensitive characterization of both the targeted and untargeted components from a complex herbal matrix.

In-depth profiling and characterization of the ginsenosides from PQR, PQL, and PQF by analyzing the negative HCD-MS² data

Comprehensive characterization of the multicomponents simultaneously from PQR, PQL, and PQF was performed based on the high-accuracy HCD-MS² data obtained on the Q Exactive Q-Orbitrap mass spectrometer by the established improved DDA approach. To enhance the reliability of MS-oriented identification, multiple solutions, including comparison with 52 reference compounds (including t_R , MS¹, and MS²), elemental composition analysis (element in use, mass tolerance, and RDBeq), fragmentation pathway interpretation, and searching an in-house ginsenoside library (499 ginsenosides recorded), had been employed. The fragmentation behaviors of different subclasses of ginsenoside reference compounds (Fig. 2) were initially studied, and the diagnostic product ions (DPI) useful for identification of the sapogenins and sugars were summarized. As a result, we could identify or tentatively characterize 347 saponins, including 147 from PQR, 173 from PQL, and 195 from PQF, respectively (see ESM Table S3). Among them, 157 ginsenosides may have

not been isolated from the *Panax* genus. Taking it into account that the negative ESI fragmentation behaviors of ginsenosides have been extensively reported in our previous reports [9, 16, 19], we do not emphasize the characterization process in the current work. Characterization of five subclasses of ginsenosides (e.g., PPT, PPD, OA, OT, and malonylated) was briefly described using typical examples. A summary of the structure features of these 347 components characterized from three parts of *P. quinquefolius* is shown in ESM Fig. S3. Under the current condition, neutral ginsenosides (PPT/PPD/OT-) could give FA-adduct as the base peak, which was different from the genuine deprotonated precursor form generated for carboxyl-containing OA- and malonyl ginsenosides [29].

Ginsenosides belonging to the PPT and PPD types represent two most important subclasses of tetracyclic saponins for most of the *Panax* species [22]. A total of 81 PPT-type (accounting for 23.34% of the total amount) and 88 PPD-type (25.36%) ginsenosides were characterized from three parts of *P. quinquefolius*, which were featured by the regular neutral eliminations of the sugars generating the sapogenin ions at m/z 475.38 and 459.38, respectively [9, 19, 29]. These fragmentation features could be evidenced in the MS² data of a PPT-type reference compound, vinaginsenoside R4 (Fig. 2). Successive cleavages of three Glc residues (3×162.05 Da) were observed by dissociating the precursor ion m/z 961.5411. A product ion at m/z 475.3804 was consistent with the deprotonated PPT sapogenin [9]. High abundances of low-

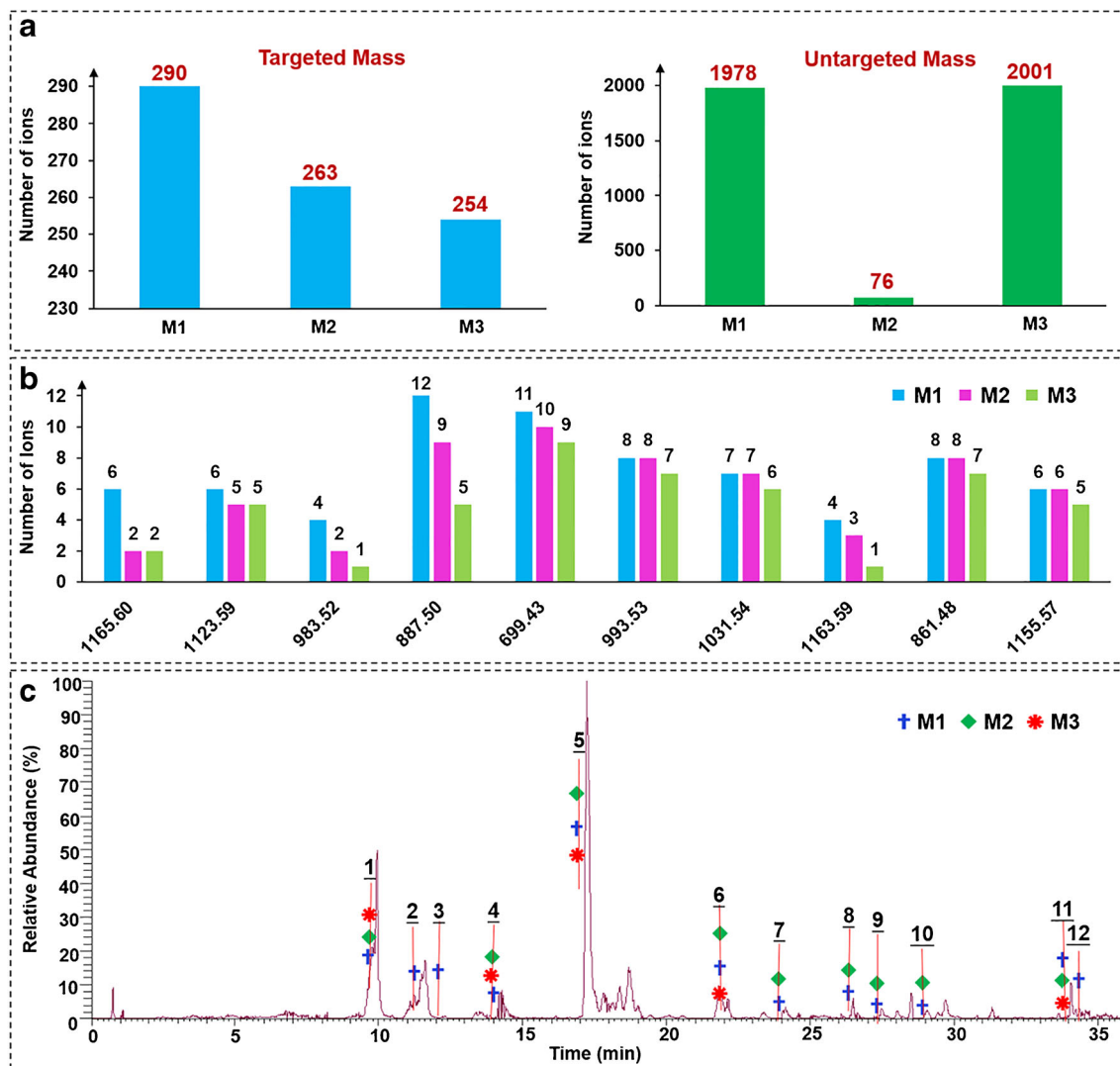


Fig. 4 Comparing the performance of three DDA approaches in profiling ginsenosides using a PQL sample. **M1**: Full MS/PIL-IPO/dd-MS²; **M2**: Full MS/PIL/dd-MS²; **M3**: Full MS/dd-MS²

mass fragments at m/z 221.0670, 161.0450, and 101.0234 were the product ions associated with the oligosaccharide of GlcGlc (Fig. 5a). We could characterize the presence of Xyl, which was used to express all the pentose characterized by NL of 132.04 Da based on the transition m/z 1061 > 929, Glc (m/z 929 > 767), Rha (m/z 767 > 621), and Glc (m/z 621 > 459), as well as the PPD saponin (m/z 459.3865), from the MS² spectrum of an unknown compound **274#**, which gave FA-adduct precursor ion at m/z 1107.6002 (t_R 28.00 min, C₅₃H₉₀O₂₁; ESM Table S3). Since no hit was found in the in-house ginsenoside library, we could primarily characterize compound **274#** as an unknown ginsenoside (PPD-2Glc-Rha-Xyl).

Ginsenosides involving a pentacyclic oleanolic acid saponin are also the characteristic saponins for the *Panax* genus. Nineteen OA-type ginsenosides (5.48% of the total) were characterized, the HCD-MS/MS fragmentation of which

showed the characteristic OA saponin ion at m/z 455.35 as well as the neutral losses of GlurA (176.03 Da), Glc (162.05 Da), and Xyl (132.04 Da). Ginsenoside Ro, a rich saponin in both *P. ginseng* and *P. quinquefolius*, showed the deprotonated precursors at m/z 955.4934, which could be dissociated by sequentially eliminating the attached sugar residues yielding rich product ions at m/z 793.4399 ([M-H-Glc]⁻) and 455.3548 ([OA-H]⁻). Different from the PPT/PPD-saponins, the transition of m/z 793 > 731 should be assigned as the neutral elimination of CO₂ + H₂O (Fig. 5b). In the case of an unknown compound **149#** (t_R 15.77 min, C₅₄H₈₆O₂₄) which gave the deprotonated precursors at m/z 1117.5480 (RDB_{eq} 12.5), almost identical product ions to those of Ro were observed in its MS² spectrum, in addition to one more Glc residue. According to the fragment at m/z 337.0802, we assumed that the additional Glc was attached to 28-Glc. Our in-house ginsenoside library gave no hit for the

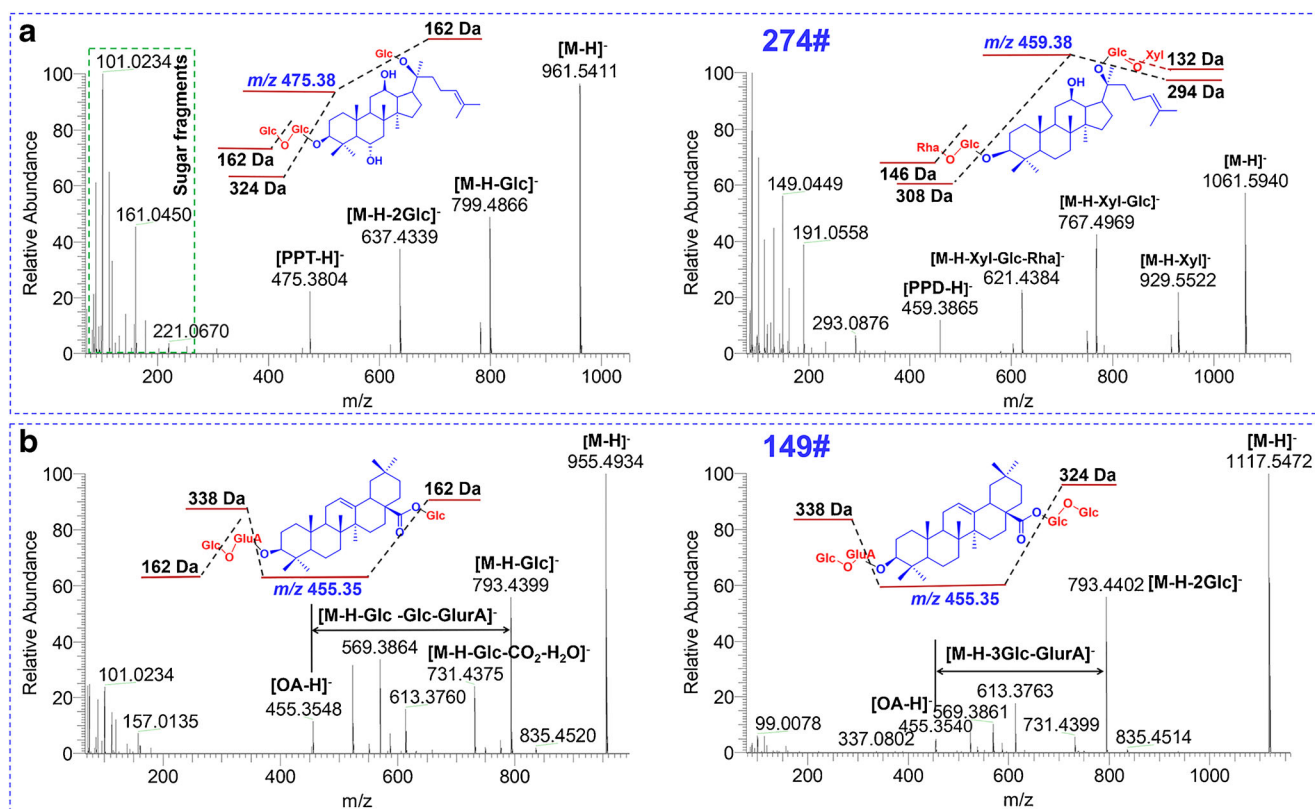


Fig. 5 Illustration for the structural elucidation of different subclasses of ginsenosides from PQR, PQL, and PQF, based on the negative HCD-MS² data. **a** PPD-/PPT-type. **b** OA-type

established structure, and accordingly, the unknown compound **149#** was tentatively characterized as OA-GlurA-3Glc.

OT-type ginsenosides represent a subclass of ginseng saponins, among which 24(*R*)-pseudoginsenoside F11 is a marker to differentiate *P. quinquefolius* from *P. ginseng* [32]. A total of 25 ginsenosides belonging to this type (7.20% of the total amount) got characterized, which exhibited a characteristic sapogenin ion at m/z 491.38 [9]. High intensity of FA-adduct precursors were detected at m/z 845.4931 (RDBeq 7.5) for the reference compound, 24(*R*)-p-F11. Characteristic product ions, including m/z 799.4883 ($[M-H]^-$), 653.4285 ($[M-H-Rha]^-$), 491.3764 ($[OT-H]^-$), and 205.0724 (a fragment of the disaccharide chain GlcRha), were dissociated (see ESM Fig. S4). Compound **161#** (t_R 17.25 min, $C_{45}H_{74}O_{17}$) gave high intensity of precursors at m/z 885.4888 (RDBeq 9.5), the HCD-MS² of which exhibited rich product ions at m/z 841.4934 ($[M-H-CO_2]^-$) and 799.4855 ($[M-H-Mal]^-$). We could infer the presence of a single malonyl substituent. The other fragments, consisting of m/z 653.4249, 491.3782, and 205.0722 (in contrast to m/z 221.0670 generated from the reference compound vinaginsenoside R4 depicted above), could suggest the attachment of a disaccharide chain GlcRha. Thus, compound **161#** was characterized as malonyl-OT-GlcRha, a novel ginsenoside structure.

Malonylginsenosides are widely distributed in various *Panax* species, particularly fresh *P. ginseng* and the flower

buds [16, 19, 22]. Compared with the neutral ginsenosides, they are acylated with one or two polar malonyl substituent(s) (mostly on the sugar moiety) showing an additional mass of 86.0004 ($C_3H_2O_3$) or 172.0008 Da ($2 \times C_3H_2O_3$). We were able to characterize 85 malonyl-substituted ginsenosides (24.50% of the total). The negative HCD-MS² fragmentation features of malonylginsenosides involve the unique neutral elimination of CO_2 and the whole malonyl group ($C_3H_2O_3$), as well as common NL of sugars and the generation of sapogenin anions typically observed at m/z 459.38 and 475.38 [16]. These MS/MS fragmentation features were readily embodied in the HCD-MS² spectrum of the reference compound, malonylginsenoside Rd (see ESM Fig. S4). In the case of an unknown compound **278#** (t_R 28.18 min, $C_{59}H_{96}O_{27}$), rich precursor ions were observed at m/z 1235.6108 (RDBeq 11.5). Its HCD-MS² spectrum displayed the product ions of m/z 1149.6086, 1107.6012, and 1089.5879, which were assigned as $[M-H-Mal]^-$, $[M-H-Mal-Ace]^-$, and $[M-H-Mal-Ace-H_2O]^-$, respectively. Because of the observation of abundant fragments consistent with Rb1 (e.g., m/z 1107.6012, 945.5463, 783.4888, 621.4400, 459.3868, and 221.0669), collectively, we could characterize compound **278#** as simultaneous malonyl and acetyl ginsenoside Rb1 or its isomer, an unreported structure for the *Panax* genus.

In addition, 49 ginsenosides (14.12% of the total) were identified or tentatively characterized which involved a

sapogenin other than PPD, PPT, OA, or OT. Their structure information is fully provided in ESM Table S3. Interestingly, these 347 ginsenosides characterized from three parts of *P. quinquefolius* have 63 unknown masses, which could demonstrate the high potency of the improved DDA method in the discovery of novel ginsenoside structures, in addition to the sensitive profiling of the known ones. However, we have to acknowledge that a large proportion of these components were characterized tentatively, which necessitates further researches for validating their structures. On one hand, additional dimensions of structure information can be obtained, such as t_R and the measurement of ion mobility-derived collision cross section (CCS) [33]. Large-scale prediction of CCS has been reported, by which predicted CCS values are useful for isomer differentiation [34]. On the other hand, LC-MS guiding phytochemical isolation can be performed to selectively isolate some compounds in pure forms with their structures fully established by NMR.

Holistic comparison of the ginsenosides among PQR, PQL, and PQF by untargeted metabolomics workflows

Untargeted metabolomics, based on the negative full-scan MS¹ data of 36 batches of *P. quinquefolius* samples (ESM Table S2), was utilized to probe into the potential markers

enabling the differentiation among three different parts of *P. quinquefolius*. The first-step processing of the multibatch MS data by Progenesis QI generated 5548 metabolic features, which were further filtered by “80% rule” and “30% variation” leading to the retaining of 4229 ions. These ions were used as the variates for pattern recognition chemometrics by PCA (unsupervised) and OPLS-DA (supervised) [16, 19]. Score plot of PCA indicated good data quality, as the QC samples were tightly clustered (see ESM Fig. S5). R2X(cum) and Q2(cum) were 0.871 and 0.693, respectively, which could indicate the acceptable fitness and predictability for the PCA model. Remarkably differentiated metabolomes among three parts were witnessed, and 42 batches of *P. quinquefolius* samples were grouped into three remarkable clusters (PQR/PQL/PQF). Supervised OPLS-DA was further applied for the discovery of the potential markers. The fitted OPLS-DA classifier exhibited good fitness (Q^2 0.946) and predictability (R^2X 0.871; R^2Y 0.975). Score plot of OPLS-DA showed a holistic difference among PQR, PQL, and PQF, particularly between the root and the other two (Fig. 6a). The chance permutation test indicated the OPLS-DA model was not over-fitted (see ESM Fig. S6). VIP plot (variable importance in projection) can unveil the importance of each variable to the observed clustering, and thus enables the discovery of potential markers. A total of 347 ions showed VIP > 1.0, and

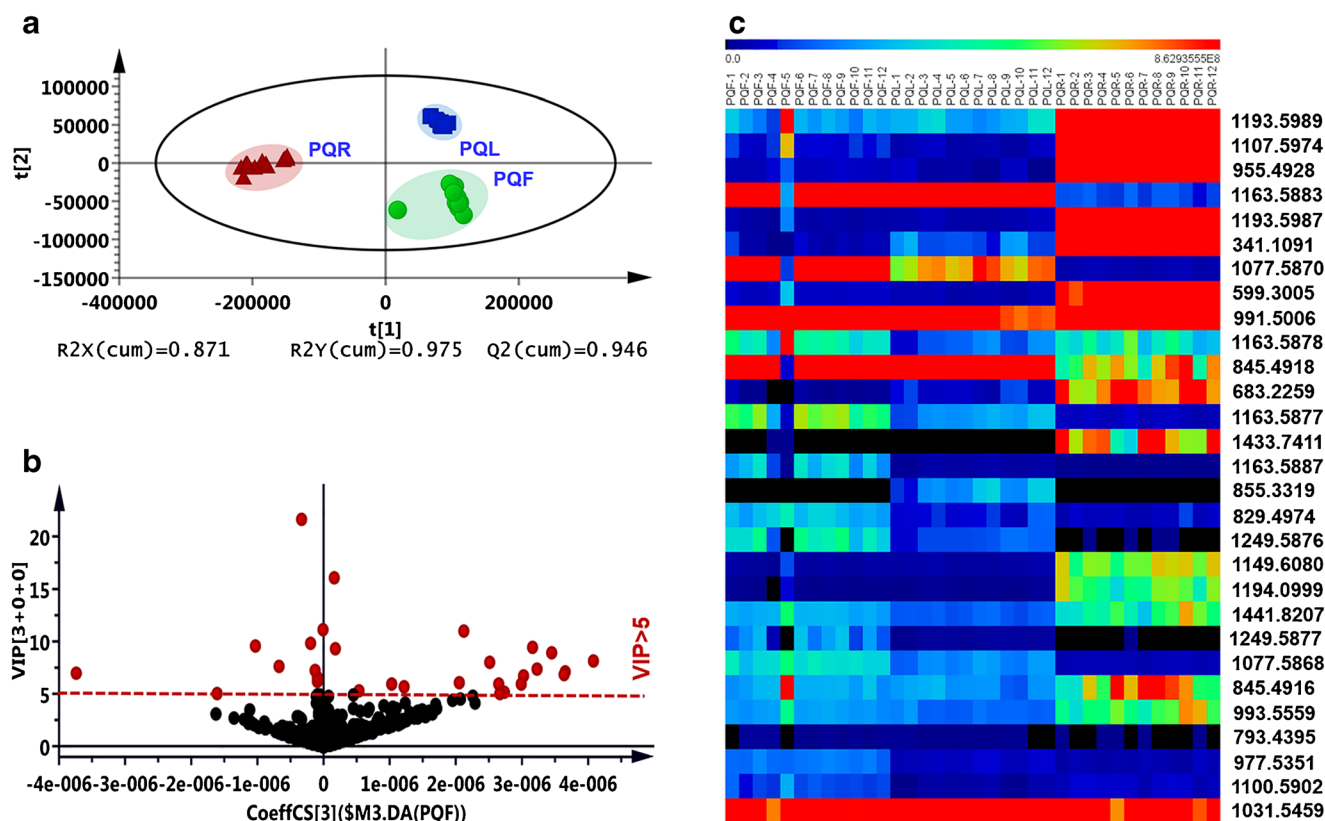
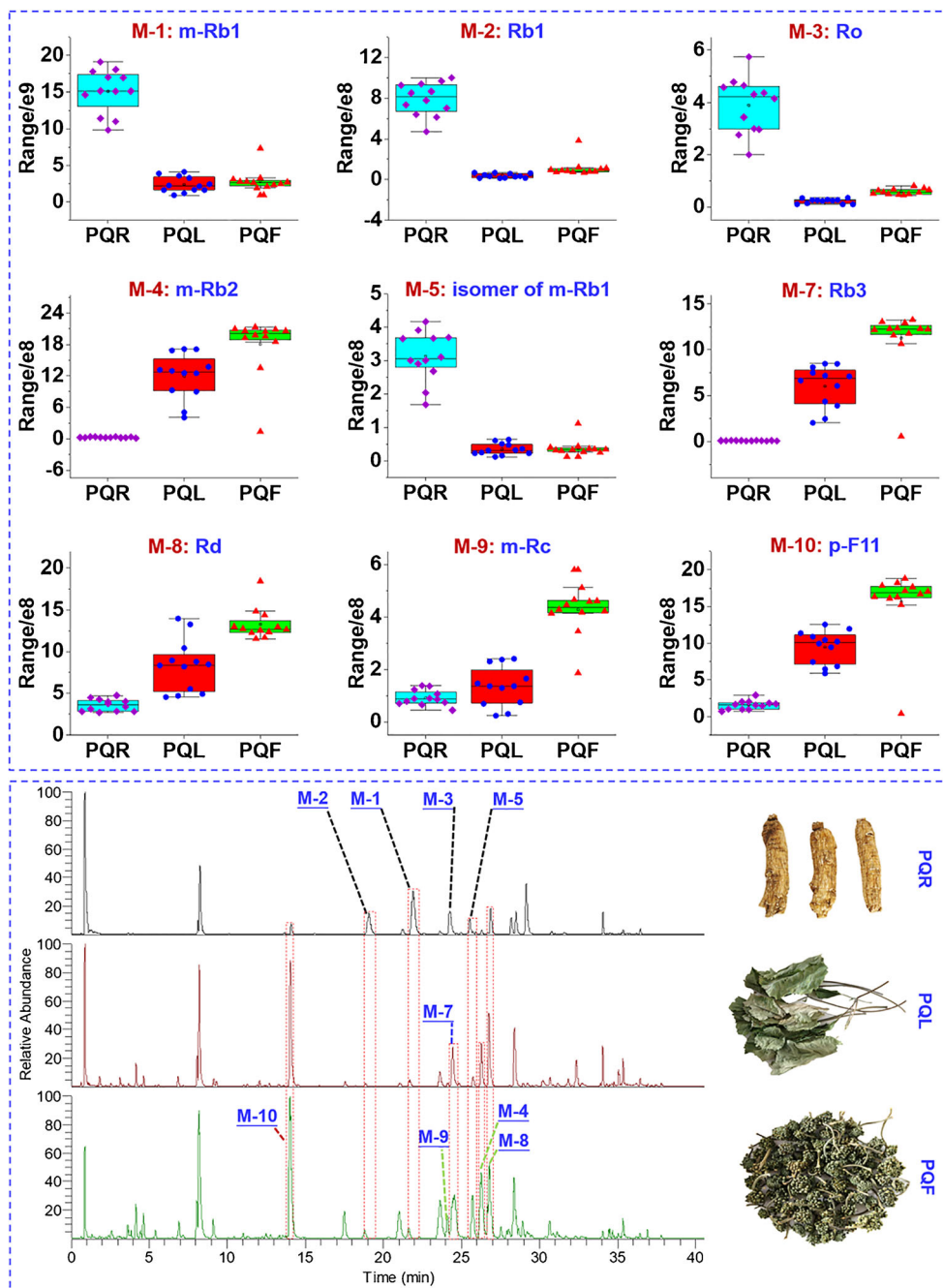


Fig. 6 Multivariate statistical analysis of different sample parts of *P. quinquefolius* to unveil differentiated ions. **a** Score plot of OPLS-DA. **b** VIP plot with the cutoff set at 5.0. **c** Heat map visualizing 29 differentiated ions

29 thereof with VIP > 5.0 (Fig. 6b). A heat map showing the content variations of these 29 differential ions is given in Fig. 6c. By searching the identification table (see ESM Table S3) and additional targeted MS/MS experiments using a Vion IMS-QTOF hybrid high-resolution mass spectrometer (Waters Corporation), these 29 ions were assigned to 24 potential marker compounds (see ESM Table S4). Twenty compounds were characterized by various tools, and 13 thereof were confirmed by comparison with the reference compounds. Box charts displaying the content difference of top 9 identified markers (m-Rb1, Rb1, Ro, m-Rb2, m-Rb1 isomer,

Rb3, Rd, m-Rc, and p-F11) are shown in Fig. 7a, which had been annotated in the base peak chromatograms (BPC) of the representative samples (Fig. 7b). In addition, we gave a primary comparison of the $^1\text{H-NMR}$ spectra for the samples representing three parts of *P. quinquefolius* (PQR: R-2; PQL: SL-3; PQF: F-15; 500 MHz, in pyridine- d_5). Despite the unbiased response for all the metabolomes can be achievable by NMR, actually it is rather difficult to exactly identify ginsenoside compounds from the $^1\text{H-NMR}$ spectra. The presence of numerous ginsenoside analogs, such as a number of PPD-type and PPT-type saponins, led to severe overlapping of

Fig. 7 A box chart (a) showing the content difference (absolute intensity values) of top 9 identified markers among PQR, PQL, and PQF, and base peak chromatograms of the representative samples (b) displaying the corresponding peaks



the signals (ESM Fig. S8), which provided little discriminatory information that could be correlated to single ginsenoside markers as LC-MS did. For instance, the chemical shifts around 3.47 ppm ($\delta_{\text{H-3}}$ [35]) could reveal more abundant PPT ginsenosides in PQR and PQF than PQL, which were consistent with the content difference of the major PPT-type ginsenoside markers Rg1 and Re (Rg2 is relatively rare), deduced by LC-MS analysis (ESM Table S4).

Based on the identified markers, key features for identifying and differentiating among three different parts of *P. quinquefolius* are summarized as follows:

- (i). the root (American ginseng, PQR) contains much more abundant m-Rb1 (**M-1**), Rb1 (**M-2**), Ro (**M-3**), and m-Rb1 isomer (**M-5**), compared with the other two parts;
- (ii). the stem leaf (PQL) and flower bud (PQF) show similar saponin composition, with richer m-Rb2 (**M-4**), Rb3 (**M-7**), and p-F11 (**M-10**), than the root;
- (iii). Rb3, p-F11, m-Rb3 (**M-12**) and its isomer (**M-13**), and m-Rd (**M-24**), may be the markers for differentiating between PQL and PQF, based on additional OPLS-DA (see ESM Fig. S7).

An ANN model was established based on the 36 batches of *P. quinquefolius* samples (80% of the sample) using top 5 identified markers, involving m-Rb1, Rb1, Ro, m-Rb2, and m-Rb1 isomer (see ESM Table S5). The trained model enabled 100% of percent correct for unknown PQR, PQF, and PQL samples, with confidence of 1.00 (see ESM Table S6).

Conclusions

For the sake of deep deconvolution of plant metabolome and probing into potential markers useful for authentication of TCM, in the current work, an improved DDA approach and untargeted metabolomics workflows were established by use of a powerful UHPLC/Q-Orbitrap-MS platform. PIL was incorporated into Full MS/dd-MS² with the IIPPO function (*If idle-pick others*) enabled presented a simultaneously targeted and untargeted metabolite characterization strategy, which exhibited remarkable improvement on the coverage of the interested components and comparable performance in acquiring the fragmentation information of unknown masses, in contrast to the conventional DDA. A novel “Ginsenoside Sieve” was established based on a fixed variation range, discrete MDF algorithm by analyzing the m/z features of all known ginsenoside structures. By these efforts, we could separate and identify 347 components from three parts of *P. quinquefolius* (PQR, PQL, and PQF). Particularly, 157 thereof have not been isolated from the *Panax* genus, and 63 new masses are reported. Multivariate statistical analysis revealed 20 potential marker components, of which m-Rb1,

Rb1, Ro, m-Rb2, and m-Rb1 isomer were five most important. An ANN discriminatory model could achieve precise identification of unknown *P. quinquefolius* samples using these five markers. The improved DDA method established in this work is proven as a potent ginsenoside characterization strategy that can enhance the sensitivity in identifying targeted compounds, and enable uncompromising performance in exploring potentially new structures. The results obtained provide new insights into the ginsenoside composition complexity of different parts of *P. quinquefolius*, which will greatly benefit the quality control of American ginseng.

Funding information This work was financially supported by National Natural Science Foundation of China (Grant No. 81872996), State Key Research and Development Project-Special Research of TCM Modernization (Grant No. 2017YFC1702104, 2018YFC1707904, and 2018YFC1707905), and Tianjin Municipal Education Commission Research Project (Grant No. 2017ZD07).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Yang WZ, Zhang YB, Wu WY, Huang LQ, Guo DA, Liu CX. Approaches to establish Q-markers for the quality standards of traditional Chinese medicines. *Acta Pharm Sin B*. 2017;7:439–46.
2. Zhang WD, Jiang HH, Yang JX, Song GS, Wen D, Liu WQ, et al. A high-throughput metabolomics approach for the comprehensive differentiation of four Pulsatilla Adans herbs combined with a nontargeted bidirectional screen for rapid identification of triterpenoid saponins. *Anal Bioanal Chem*. 2019;411:2071–88.
3. Yao CL, Yang WZ, Si W, Shen Y, Zhang NX, Chen HL, et al. An enhanced targeted identification strategy for the selective identification of flavonoid *O*-glycosides from *Carthamus tinctorius* by integrating offline two-dimensional liquid chromatography/linear ion-trap-Orbitrap mass spectrometry, high-resolution diagnostic product ions/neutral loss filtering and liquid chromatography-solid phase extraction-nuclear magnetic resonance. *J Chromatogr A*. 2017;1491:87–97.
4. Pirok BWJ, Stoll DR, Schoenmakers PJ. Recent developments in two-dimensional liquid chromatography: fundamental improvements for practical application. *Anal Chem*. 2019;91:240–63.
5. Stephan S, Jakob C, Hippler J, Schmitz OJ. A novel four-dimensional analytical approach for analysis of complex samples. *Anal Bioanal Chem*. 2016;408:3751–9.
6. Przybylski C, Bonnet V. Discrimination of cyclic and linear oligosaccharides by tandem mass spectrometry using collision-induced dissociation (CID), pulsed-Q-dissociation (PQD) and the higher-energy *C*-trap dissociation modes. *Rapid Commun Mass Spectrom*. 2013;27:75–87.
7. Liu S, Liang YZ, Liu HT. Chemometrics applied to quality control and metabolomics for traditional Chinese medicines. *J Chromatogr B*. 2016;1015–1016:82–91.
8. Ren DB, Ran L, Yang C, Xu ML, Yi LZ. Integrated strategy for identifying minor components in complex samples combining mass defect diagnostic ions and neutral loss information based on ultra-performance liquid chromatography-high resolution mass

- spectrometry platform: *Folium Artemisiae Argyi* as a case study. *J Chromatogr A*. 2018;1550:35–44.
9. Qiu S, Yang WZ, Shi XJ, Yao CL, Yang M, Liu M, et al. A green protocol for efficient discovery of novel natural compounds characterization of new ginsenosides from the stems and leaves of *Panax ginseng* as a case study. *Anal Chim Acta*. 2015;893:65–76.
 10. Tan T, Lai CJS, Zeng SL, Li P, Liu EH. Comprehensive profiling and characterization of quassinoids from the seeds of *Brucea javanica* via segment and exposure strategy coupled with modified mass defect filter. *Anal Bioanal Chem*. 2016;408:527–33.
 11. Li H, Qin Q, Shi XJ, He J, Xu GW. Modified metabolites mapping by liquid chromatography-high resolution mass spectrometry using full scan/all ion fragmentation/neutral loss acquisition. *J Chromatogr A*. 2019;1583:80–7.
 12. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;12:523.
 13. Li H, Cai YP, Guo Y, Chen FF, Zhu ZJ. MetDIA: targeted metabolite extraction of multiplexed MS/MS spectra generated by data-independent acquisition. *Anal Chem*. 2016;88:8757–64.
 14. Zhang JY, Wang ZJ, Zhang Q, Wang F, Ma Q, Lin ZZ, et al. Rapid screening and identification of target constituents using full scan-parent ions list-dynamic exclusion acquisition coupled to diagnostic product ions analysis on a hybrid LTQ-Orbitrap mass spectrometer. *Talanta*. 2014;124:111–22.
 15. Lu JX, Zhang CX, Hu Y, Zhang MH, Wang YH, Qian YX, et al. Application of multiple chemical and biological approaches for quality assessment of *Carthamus tinctorius* L. (safflower) by determining both the primary and secondary metabolites. *Phytomedicine*. 2019;58:152826.
 16. Shi XJ, Yang WZ, Qiu S, Yao CL, Shen Y, Pan HQ, et al. An in-source multiple collision-neutral loss filtering based nontargeted metabolomics approach for the comprehensive analysis of malonyl-ginsenosides from *Panax ginseng*, *P. quinquefolius*, and *P. notoginseng*. *Anal Chim Acta*. 2017;952:59–70.
 17. Fu LL, Ding H, Han LF, Jia L, Yang WZ, Zhang CX, et al. Simultaneously targeted and untargeted multicomponent characterization of Erzhi Pill by offline two-dimensional liquid chromatography/quadrupole-Orbitrap mass spectrometry. *J Chromatogr A*. 2019;1584:87–96.
 18. Park HW, In G, Kim JH, Cho BG, Han GH, Chang IM. Metabolomic approach for discrimination of processed ginseng genus (*Panax ginseng* and *Panax quinquefolius*) using HPLC-QTOF MS. *J Ginseng Res*. 2014;38:59–65.
 19. Qiu S, Yang WZ, Yao CL, Qiu ZD, Shi XJ, Zhang JX, et al. Nontargeted metabolomic analysis and “commercial-homophyletic” comparison-induced biomarkers verification for the systematic chemical differentiation of five different parts of *Panax ginseng*. *J Chromatogr A*. 2016;1453:78–87.
 20. Li Y, Jin Y, Yang SP, Zhang WW, Zhang JZ, Zhao W, et al. Strategy for comparative untargeted metabolomics reveals honey markers of different floral and geographic origins using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry. *J Chromatogr A*. 2017;1499:78–89.
 21. Shin JS, Park HW, In G, Seo HK, Won TH, Jang KH, et al. Metabolomic approach for discrimination of four- and six-year-old red ginseng (*Panax ginseng*) using UPLC-QToF-MS. *Chem Pharm Bull*. 2016;64:1298–303.
 22. Yang WZ, Hu Y, Wu WY, Ye M, Guo DA. Saponins in the genus *Panax* L. (Araliaceae): a systematic review of their chemical diversity. *Phytochemistry*. 2014;106:7–24.
 23. Kang S, Min H. Ginseng, the ‘immunity boost’: the effects of *Panax ginseng* on immune system. *J Ginseng Res*. 2012;36:354–68.
 24. Yan Y, Song QQ, Chen XJ, Li J, Li P, Wang YT, et al. Simultaneous determination of components with wide polarity and content ranges in *Cistanche tubulosa* using serially coupled reverse phase-hydrophilic interaction chromatography-tandem mass spectrometry. *J Chromatogr A*. 2017;1501:39–50.
 25. Qi LW, Wang CZ, Yuan CS. Ginsenosides from American ginseng: chemical and pharmacological diversity. *Phytochemistry*. 2011;72:689–99.
 26. Yang WZ, Qiao X, Li K, Fan JR, Bo T, Guo DA, et al. Identification and differentiation of *Panax ginseng*, *Panax quinquefolium*, and *Panax notoginseng* by monitoring multiple diagnostic chemical markers. *Acta Pharm Sin B*. 2016;6:568–75.
 27. Wang YP, Choi HK, Brinckmann JA, Jiang X, Huang LF. Chemical analysis of *Panax quinquefolius* (North American ginseng): a review. *J Chromatogr A*. 2015;1426:1–15.
 28. Kim SH, Shin YS, Choi HK. Nano ESI-MS-based lipidomics to discriminate between cultivars, cultivation ages, and parts of *Panax ginseng*. *Anal Bioanal Chem*. 2016;408:2109–21.
 29. Yang WZ, Shi XJ, Yao CL, Huang Y, Hou JJ, Han SM, et al. A novel neutral loss/product ion scan-incorporated integral approach for the untargeted characterization of comparison of the carboxyl-free ginsenosides from *Panax ginseng*, *Panax quinquefolius*, and *Panax notoginseng*. *J Pharm Biomed Anal*. 2020;177:112813.
 30. Agatonovic-Kustrin S, Loescher CM. Qualitative and quantitative high performance thin layer chromatography analysis of *Calendula officinalis* using high resolution plate imaging and artificial neural network data modelling. *Anal Chim Acta*. 2013;798:103–8.
 31. Pan HQ, Yao CL, Yang WZ, Yao S, Huang Y, Zhang YB, et al. An enhanced strategy integrating offline two-dimensional separation and step-wise precursor ion list-based raster-mass defect filter: characterization of indole alkaloids in five botanical origins of *Uncaria Ramulus Cum Unicis* as an exemplary application. *J Chromatogr A*. 2018;1563:124–34.
 32. Li L, Luo GA, Liang QL, Hu P, Wang YM. Rapid qualitative and quantitative analyses of Asian ginseng in adulterated American ginseng preparations by UPLC/Q-TOF-MS. *J Pharm Biomed Anal*. 2010;52:66–72.
 33. May JC, Morris CB, McLean JA. Ion mobility collision cross section compendium. *Anal Chem*. 2017;89:1032–44.
 34. Zhou ZW, Shen XT, Tu J, Zhu ZJ. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Anal Chem*. 2016;88:11084–91.
 35. Teng RW, Li HZ, Chen JT, Wang DZ, He YN, Yang CR. Complete assignment of ^1H and ^{13}C NMR data for nine protopanaxatriol glycosides. *Magn Reson Chem*. 2002;40:483–8.
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.