Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

A novel method for integrating chromatographic fingerprint analytical units of Chinese materia medica: the matching frequency statistical moment method

LI Haiying[a, b†], PAN Xue[b, c†], WANG Mincun[b, c], LI Wenjiao[b, c], HE Peng[b, c], HUANG Sheng[d], HE Fuyuan[b, c, e, f*]

a. Hospital-Made Preparations Center, The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan 410007, China

b. Hunan Provincial Key Laboratory of Drugability and Preparation Modification of TCM, Changsha, Hunan 410208, China

c. School of Pharmacy, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

d. Jiuzhitang Co., Ltd., Changsha, Hunan 410205, China

e. Supramolecular Mechanism and Mathematic-Physics Characterization for Chinese Materia Medica, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

f. Property and Pharmacodynamic Key Laboratory of TCM, State Administration of Traditional Chinese Medicine, Changsha, Hunan 410208, China

ARTICLE INFO

Article history
Received 28 May 2024
Accepted 22 August 2024
Available online 25 September 2024

Keywords
Chromatographic fingerprints
Analytical units
Matching frequency statistical moment method
Chinese materia medica
Danxi Granule (丹膝颗粒, DXG)
Quality evaluation

ABSTRACT

Objective To facilitate the quality evaluation suitable for the unique characteristics of Chinese materia medica (CMM) by developing and implementing a novel approach known as the matching frequency statistical moment (MFSM) method.

Methods This study established the MFSM method. To demonstrate its effectiveness, we applied this novel approach to analyze Danxi Granules (丹膝颗粒, DXG) and its constituent herbal materials. To begin with, the ultra-performance liquid chromatography (UPLC) was applied to obtain the chromatographic fingerprints of DXG and its constituent herbal materials. Next, the MFSM was leveraged to compress and integrate them into a new fingerprint with fewer analytical units. Then, we characterized the properties and variability of both the original and integrated fingerprints by calculating total quantum statistical moment (TQSM) parameters, information entropy and information amount, along with their relative standard deviation (RSD). Finally, we compared the TQSM parameters, information entropy and information amount, and their RSD between the traditional and novel fingerprints to validate the new analytical method.

Results The chromatographic peaks of DXG and its 12 raw herbal materials were divided and integrated into peak families by the MFSM method. Before integration, the ranges of the peak number, three TQSM parameters, information entropy and information amount for each peak or peak family of UPLC fingerprints of DXG and its 12 raw herbal materials were 95.07 – 209.73, 9 390 – 183 064 μv·s, 5.928 – 21.33 min, 22.62 – 106.69 min^2, 4.230 – 6.539, and 50 530 – 974 186 μv·s, respectively. After integration, the ranges of these parameters were 10.00 – 88.00, 9 390 – 183 064 μv·s, 5.951 – 22.02 min, 22.27 – 104.73 min^2, 2.223 – 5.277, and 38 159 – 807 200 μv·s, respectively. Correspondingly, the RSD of all the aforementioned parameters before integration were 2.12% – 9.15%, 6.04% – 49.78%, 1.15% – 23.10%, 3.97% – 25.79%, 1.49% – 19.86%, and 6.64% – 51.20%, respectively. However, after integration, they changed to 0.00%, 6.04% – 49.87%, 1.73% – 23.02%, 3.84% – 26.85%, 1.17% – 16.54%, and 6.40% –

†The authors contributed equally.

*Corresponding author: HE Fuyuan, E-mail: pharmsharking@163.com.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2024.12.009

Citation: LI HY, PAN X, WANG MC, et al. A novel method for integrating chromatographic fingerprint analytical units of Chinese materia medica: the matching frequency statistical moment method. Digital Chinese Medicine, 2024, 7(3): 294-308.

48.59%, respectively. The results demonstrated that in the newly integrated fingerprint, the analytical units of constituent herbal materials, information entropy and information amount were significantly reduced ($P < 0.05$), while the TQSM parameters remained unchanged ($P > 0.05$). Additionally, the RSD of the TQSM parameters, information entropy, and information amount didn't show significant difference before and after integration ($P > 0.05$), but the RSD of the number and area of the integrated analytical units significantly decreased ($P < 0.05$).

Conclusion The MFSM method could reduce the analytical units of constituent herbal materials while maintain the properties and variability from their original fingerprint. Thus, it could serve as a feasible and reliable tool to reduce difficulties in analyzing multi-components within CMMs and facilitating the evaluation of their quality.

1 Introduction

Chinese materia medica (CMM) has been drawing increasing attention and receiving worldwide acceptance due to its vital role in preventing and treating diseases. Nevertheless, despite its extensive worldwide application, one of the major concerns hindering its modernization and globalization is how to ensure its stability and consistency in quality. Given that constituent herbal materials in CMM are influenced by a variety of factors, including herbal origins, geographical regions, harvest time, processing methods, and preparation techniques, distinct disparities in quality exist among different CMMs [1-3]. Currently, there are few appropriate analytical methods focusing on integrating the analytical units of the chromatographic fingerprint to reduce the difficulty in assessing the quality of CMM.

At present, to establish new approaches for assessing the quality of CMM with multiple components, researchers have proposed various methods based on chemical components, medicinal effects or both. Among them, chromatographic fingerprint analysis plays an important role in the quality control of CMM with complex components [4-6]. It serves as a comprehensive method normally combined with similarity analysis, chemometrics or quantitative analysis to ensure the authentication of species as well as the consistency and stability of CMM properties, and facilitate the evaluation of its quality [7-11]. Characterized by the intricate pharmacodynamics of the components of CMM and their preparation processes, a succession of methods for quality control and evaluation has already been developed. These include, but are not limited to, the correlation coefficient method, cluster analysis, angle cosine method, fuzzy tip T-distribution method, Euclidean distance method, hyper-information feature digitization, and total statistical moment (similarity) method. However, these methods often streamline the analysis by focusing on common components, correlating them with medicinal effects, and subsequently identifying the active components, leading to the loss of some active components. Moreover, the pharmacological essence of CMM transcends mere aggregation of individual chemical constituents; it likely encompasses a dynamic ensemble of components that engage in self-organization, self-assembly, and self-replication processes. Therefore, how to integrate the component families into a new fingerprint according to their active characteristics and establish a new analytical unit to characterize the imprinting template of the component families without losing the characteristics of the parameters from previous fingerprint analysis has become a bottleneck in the theoretical, industrial, and applied research on CMM.

This study initially developed and validated a reasonable and reliable analytical matching frequency statistical moment (MFSM) method applied to chromatographic fingerprints. Using this method, analytical units were reduced without losing any chromatographic peaks through peak division and integration, with the properties and variability of the fingerprints retained. Danxi Granules (丹膝颗粒, DXG) and its raw herbal materials were used as an example to verify the effectiveness of the method. DXG, a well-known traditional Chinese medicine (TCM), has been widely applied with excellent outcomes in preventing and treating patients with stroke hemiplegia of the Yin deficiency and wind movement type and Qi deficiency and blood stasis type [12].

2 Materials and methods

2.1 MFSM method

In this study, the chromatographic fingerprints of DXG and its herbal materials were obtained with the use of ultra-performance liquid chromatography (UPLC), which were compressed and integrated into a new fingerprint through a novel method, the MFSM. To validate the effectiveness of this method, the properties and variability from different batches were compared between the original and novel fingerprints. Figure 1 shows the flowchart of establishing and validating the MFSM method. First, the original chromatographic parameters, such as the number of peaks, retention time, peak area, and matching frequency, were obtained through peak calibration and matching, and the total quantum statistical moment

(TQSM) parameters, information entropy, information content and their relative standard deviation (RSD) values were calculated according to the retention time and peak area. Subsequently, a novel integrated fingerprint with new analytical units was obtained by partially dividing and integrating the chromatographic peaks into peak families using the MFSM method. The aforementioned parameters and their RSD values of the novel fingerprint were also obtained. Finally, the MFSM method was evaluated and validated by comparing all these parameters and their RSD values before and after integration. The *t* test was employed for statistical analysis.

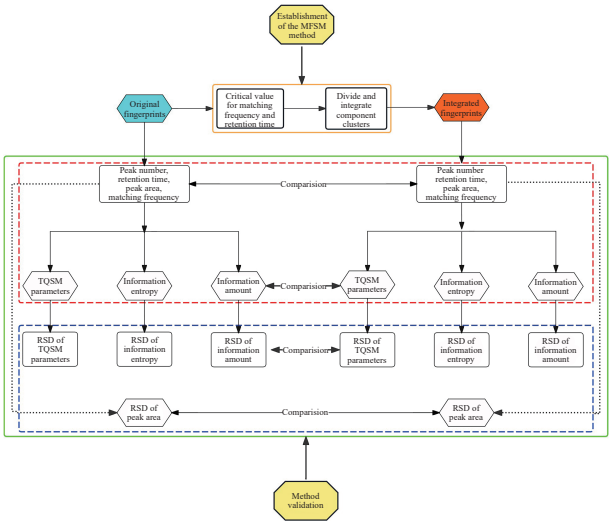

Figure 1 The flowchart of establishing and validating the MFSM method

2.1.1 Calculation of the TQSM parameters, information entropy, and information amount

(i) Calculation of the TQSM parameters. Based on the famous statistical moment theory, we employed the TQSM method to qualitatively and quantitatively analyze the properties of the chromatographic fingerprint of CMM [13-15], which was first integrated into the response curve of the peaks versus time to obtain the zeroth moment of total quanta (AUC_T); the mean value of the response curve of the peaks projected on the time coordinate was calculated as the first moment of total quanta (MRT_T) and their variance to the time span was calculated as the second moment of total quanta (VRT_T), as detailed below. The calculation of the TQSM parameters is explicitly explained in the cited literature [16].

(a) The zeroth moment of total quanta (AUC_T) represents the area under the curve (AUC) of the chromatographic fingerprints, which encompasses all chromatographic peak areas integrated versus time from zero to infinity. It can be calculated by Equation (1).

$$AUC_T = \sum_{i=1}^{n} A_i \tag{1}$$

Where there are n peaks in a chromatogram, and A_i represents the peak area of the ith peak.

(b) The first moment of total quanta (MRT_T) is defined as the mean value of the retention time of the chromatographic peaks. It can be calculated by Equation (2).

$$MRT_T = \sum_{i=1}^{n} A_i \lambda_i \Big/ \sum_{i=1}^{n} A_i \tag{2}$$

Where there are n peaks in a chromatogram, and A_i represents the peak area of the ith peak, λ_i as the retention time of the ith peak.

(c) The second moment of total quanta (VRT_T) is defined as the variance of the retention time of the chromatographic peaks and can be described as the discrete degree of residence time, as given in Equation (3).

$$VRT_T = \sum_{i=1}^{n} A_i \left(\sigma_i^2 + \lambda_i^2\right) \Big/ \sum_{i=1}^{n} A_i - \lambda_i^2 \tag{3}$$

Where there are n peaks in a chromatogram, and A_i represents the peak area of the ith peak, λ_i as the retention time of the ith peak, and σ_i as the standard deviation of the retention time of the ith peak.

On one hand, AUC_T can be used to quantitatively characterize the constituent components of CMM and the quantum alternation of the holistic components. On the other hand, MRT_T and VRT_T can be used to qualitatively examine the chromatographic properties and describe the variation of their peaks. Therefore, these TQSM parameters (AUC_T, MRT_T, and VRT_T) were employed to characterize the amount and properties of the chromatogram. Namely, they were determined by the chemical compositions and concentration of the CMM. Previous studies have affirmed that the TQSM approach exhibits superior performance in terms of anti-interference, additive effects, coupling interactions, and statistical analysis when compared with alternative methods [13-16].

(ii) Calculation of information entropy and information amount. The distinct peaks observed in the chromatographic fingerprints of CMM correspond to individual chemical components, each regarded as a discrete unit of information. The height of each peak in the fingerprint reflects the likelihood of its chemical component's presence, which is influenced by the chromatographic conditions and interactions among the components. Consequently, the presence or absence of chromatographic peaks in the fingerprints of CMM can be characterized using information entropy and information amount [14-17]. Information entropy serves as a probabilistic measure of the presence of values, thereby characterizing the degree of information units within each component. Meanwhile, the information amount represents the quantity of information carried by the fingerprints.

The presence and absence of the chromatographic peaks in the fingerprints of CMM can be expressed by the information amount (ΔS_{iq}) and calculated with Equation (4).

$$\Delta S_{iq} = -\frac{A_T}{W_0 \ln_2} \sum_{i=1}^{m} W_i \ln W_i \quad (4)$$

Where there are m peaks in a chromatogram, W_i represents the percentage of the peak area of the ith peak, $\ln w_i$ is the Napierian logarithm of W_i, $\ln_2$ is a constant, W_0 represents the total component content, and A_T represents the total peak area. The information entropy (ΔS_{ie}) was calculated with Equation (5).

$$\Delta S_{ie} = -\frac{1}{\ln_2} \sum_{i=1}^{m} W_i \ln W_i \quad (5)$$

Where there are m peaks in a chromatogram, and W_i is the percentage of the peak area of the ith peak. Information entropy and information amount jointly expressed the integral information contained in the chromatographic fingerprints, including the number of peaks and peak area.

2.1.2 Standards for dividing and integrating component families As a biomolecular complex supramolecular, CMM conveys its medicinal properties and efficacy through its imprinting templates. The chromatographic peaks within the fingerprint spectrum were statistically analyzed through chromatographic techniques, and the types and contents of CMM components were determined using TQSM parameters, alongside the information entropy, and information amount derived from the fingerprint spectrum. The number of imprinting templates component families of CMM was determined by the convex-concave distribution of the matching frequency of chromatographic peaks with respect to retention time, which was statistically divided by the matching frequency of chromatographic peaks in each time period. The imprinting template attribution of CMM component families was obtained by dividing the component family attribution among the imprinting template component families in the CMM fingerprint using band matching frequencies. The imprinting template component family within the CMM fingerprint represents a novel fingerprinting approach that leverages the imprinting template concept for component families. This method employs band matching frequencies to determine new retention times or the intensity of parameters, and incorporates the area or capacity parameters of each component cluster as additional peak area or capacity parameters [13-16].

In the analysis of chromatographic fingerprints, a part of a molecule randomly interacted with the stationary phase of the chromatograph, leading to imprinting-phenomena such as adsorption-migration-desorption-readsorption-remigration. The holistic imprinting characteristics were reflected in the form of retention time [18]. Consequently, the average retention time was a comprehensive index to reflect the holistic imprinting properties of all the components. Variations in the average retention time of component families reflected alterations in their chemical composition, holistic imprinting characteristics, and therapeutic efficacy. Once the imprinting evaluation index based on the average retention time was established, it enabled the clarification of the chemical composition and efficacy properties of component families. The imprinting behavior of component families within a chromatogram showed that the frequency of chromatographic peaks appeared in a pattern that followed a convex-concave distribution as a function of retention time. These peaks' statistical matching frequencies were computed using Student's t-distribution, and subsequently, the number of component families could be obtained based on the critical value of these statistical matching frequencies.

(i) Determination of the critical value of matching frequency. The Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicines (2012A) software was utilized to establish fingerprints from S batches of the same herb. This involved peak calibration and matching to determine key parameters such as the number of peaks, retention time, peak area, and matching frequency. The matching frequency refers to the frequency with which a particular peak appears at a specific retention time among the S batches of samples. First, a statistical analysis was carried out on the matching frequencies of the chromatographic fingerprints. Second, the average matching frequency and its standard deviation were calculated. The confidence level was then established, and the critical value for the matching frequency, specific to the sample size, was determined using Equation (6). Subsequently, components with a matching frequency exceeding the critical threshold were identified as marker or common components. This process enabled the classification of component families and the determination of their total count.

$$M_{PC} = \overline{M_P} + t_{(\alpha,v)} \times \mathrm{SD} \quad (6)$$

Where M_{PC} is defined as the critical value of matching frequencies of chromatographic peaks required to be classified as a component family. In addition, $\overline{M_P}$ and SD are the mean value and standard deviation of the matching frequencies, respectively, and $t_{(\alpha,v)}$ is the t critical value under the confidence coefficient $(1 - \alpha)$ and the degree of freedom (v). The number of chromatographic peaks that exhibit a matching frequency exceeding the critical value defines the number of distinct component families.

Equation (6) showed that when the critical value of a matching frequency was greater than or equal to the total batch number, the latter value was adopted as the critical value for the matching frequency. Generally, the confidence level $(1 - \alpha)$ was set at 95% or 99%, which implied that α was set at 2.5% or 0.005%, respectively, on one side.

The number of component families was affected by the total batch number and the overall matching similarity. As a result, when both the total batch number and the holistic matching similarity were held constant, the number of the component family was correspondingly fixed.

(ii) Determination of the critical value of retention time. The component families in the CMM fingerprints were identified according to the critical values of retention time and matching frequency. The peaks with retention time less than the critical value were assigned to the former component family. In contrast, those greater than the critical value were classified as the latter component family. When the retention time of the peak was equal to the critical value, it was restricted to the last significant digits of $t_{R,j}$: the odd number was classified to the former component family, while the even number was distributed to the latter component family. The data were processed as follows.

Suppose there exist n component families in the fingerprints, sequentially ordered by retention time as N_{pc1}, N_{pc2}, N_{pc3}, ..., N_{pci}, ..., N_{pcn}. The critical retention times for each component family were calculated according to the statistical moment of the matching frequencies. For example, the critical retention time between N_{pc1} and N_{pc2} was calculated using Equation (7).

$$t_{Rpc1=}\sum_{j=N_{pc1}}^{N_{pc2}} M_{p,j}t_{R,j} \Big/ \sum_{j=N_{pc1}}^{N_{pc2}} M_{P,j} \tag{7}$$

Where t_{Rpc1} refers to the critical value for dividing N_{pc1} to N_{pc2}, N_{pc1} is the first component family, N_{pc2} is the second component family, $M_{P,j}$ means matching frequencies of all these components from N_{pc1} to N_{pc2}, and $t_{R,j}$ stands for the retention time of all these components from N_{pc1} to N_{pc2}.

Similarly, the critical value of retention time for dividing N_{pci} and $N_{pc(i+1)}$ can be calculated by Equation (8).

$$t_{Rpci} = \sum_{j=N_{pci}}^{N_{pc(i+1)}} N_{P,j}t_{R,j} \Big/ \sum_{j=N_{pci}}^{N_{pc(i+1)}} N_{P,j} \tag{8}$$

Where t_{Rpci} indicates the critical value of retention time from N_{pci} and $N_{pc(i+1)}$, N_{pci} is the ith component family, and $N_{pc(i+1)}$ is the $(i+1)$th component family, $N_{P,j}$ stands for the matching frequencies of these components between N_{pci} and $N_{pc(i+1)}$, $t_{R,j}$ is the retention time of these components between N_{pci} and $N_{pc(i+1)}$.

When comparing the retention time of every single peak to the critical value of retention time, the peaks with retention time less than t_{Rpci} were assigned to the N_{pci}, whereas those greater than t_{Rpci} were classified as $N_{pc(i+1)}$. When the peak retention time was equal to t_{Rpci}, it was categorized based on the terminal digits of $t_{R,j}$: the odd numbers were classified into the N_{pci}, while even numbers into the $N_{pc(i+1)}$.

2.1.3 Determination of new chromatographic parameters of component families A series of component families in chromatographic fingerprints were divided and integrated based on the critical value of matching frequencies and retention time of chromatographic peaks. Equations (9) and (10) facilitate the integration and computation of novel chromatographic parameters, including the updated retention time and area for peak families. For example, the integrated retention time for N_{pci} can be calculated by Equation (9).

$$t_{Rpi} = \sum_{j=t_{Rpc(i-1)}}^{t_{Rpci}} N_{P,j}t_{R,j} \Big/ \sum_{j=t_{Rpc(i-1)}}^{t_{Rpci}} N_{P,j} \tag{9}$$

Where t_{Rpi} is the new retention time of N_{pci}, and t_{Rpci} is that of the ith component family. Correspondingly, $t_{Rpc(i-1)}$ is that of the $(i-1)$th component family. $N_{P,j}$ stands for the matching frequencies of these components between N_{pci} and $N_{pc(i+1)}$, $t_{R,j}$ is the retention time of these components between N_{pci} and $N_{pc(i+1)}$.

The new peak family area of N_{pci} (A_{Rpi}) can also be integrated using Equation (10).

$$A_{Rpi} = \sum_{j=t_{Rpc(i-1)}}^{t_{Rpci}} A_{R,j} \tag{10}$$

Where A_{Rpi} is the newly integrated area of peak family N_{pci}, and $A_{R,j}$ is the single peak area from $t_{Rpc(i-1)}$ to t_{Rpci}.

The TQSM parameters, information entropy, and information number of the integrated fingerprints were calculated by Equations (1) to (5).

2.2 Chemicals, reagents, and herbal materials

A comprehensive set of 14 reference standards was procured, including catalpol (≥ 98%), gastrodin (≥ 98%), ligustrazine (≥ 98%), geniposide (≥ 98%), paeoniflorin (≥ 98%), ferulic acid (≥ 98%), ecdysone (≥ 98%), salvianolic acid (≥ 98%), quercetin (≥ 98%), paeonol salvianolic acid (≥ 98%), icariin (≥ 98%), aurantio-obtusin (≥ 98%), chrysophanol (≥ 98%), and tanshinone IIA (≥ 98%). These standards were purchased from Shanghai Yuanye Biological Technology Co., Ltd., China. UPLC-grade acetonitrile was purchased from Tedia Company Inc., USA, and all other reagents were of analytical grade. Ultrapure water was prepared using a Milli-Q Plus water purification system (Millipore Corporation, USA).

DXG is composed of 12 CMMs, i.e., Danshen (Salviae Miltiorrhizae Radix et Rhizoma, the root and rhizome of *Salvia miltiorrhiza* Bge.), Niuxi (Achyranthis Bidentatae Radix, the dried root of *Achyranthes bidentata* Bl.), Tianma (Gastrodiae Rhizoma, the dried tuber of *Gastrodia elata* Bl.), Mudanpi (Moutan Cortex, the dried root bark of *Paeonia suffruticosa* Andr.), Chishao (Paeoniae Radix Rubra, the dried root of *Paeonia lactiflora* Pall.), Chuanxiong (Chuanxiong Rhizoma, the dried root of *Ligusticum*

chuanxiong Hort., Dihuang (Rehmanniae Radix, dried tuberous roots of *Rehmannia glutinosa* Libosch.), Yinyanghuo (Epimedii Folium, the dried leaves of *Epimedium brevicornum* Maxim.), Sangjisheng [Taxilli Herba, the dry leafy stem and branches of *Taxillus chinensis* (DC.) Danser], Zhizi (Gardeniae Fructus, the ripe fruits of *Gardenia jasminoides* Ellis), Juemingzi (Cassiae Semen, the dried ripe seeds of *Cassia obtusifolia* L.), and Huomaren (Cannabis Fructus, the seeds of *Cannabis sativa* L.). The 12 CMMs were originated from three different geographical regions (labeled as S1 – S5, S6 – S10, and S11 – S15, respectively), with five batches from each region. All the raw herbal materials were authenticated by Professor SHI Jilian (School of Pharmacy, Hunan University of Chinese Medicine, Changsha, China). Additionally, 15 batches of DXG were obtained from Jiuzhitang Co., Ltd. with the batch numbers of 20190305, 20190120, 20190105, 20190122, 20190401, 20190606, 20190503, 20190817, 20190705, 20190909, 20190907, 20190920, 20190921, 20191212, and 20190704. The above 15 batches of DXG were marked as S1 – S15, respectively.

2.3 Preparation of samples and standard solutions

2.3.1 DXG solution The DXG was ground into fine powder. An amount of 0.25 g of this powder was weighed out and sonicated at 300 W and 40 kHz for 30 min in a conical flask with 5 mL of 50% methanol. Prior to sonication, the initial weight of the sample was recorded. After the extraction process, weight loss was compensated with additional 50% methanol, and the mixture was centrifuged at 12 000 r/min for 15 min at 4 °C. The supernatant was subsequently filtered through a 0.22 μm microporous membrane to obtain the filtrate.

2.3.2 Single herb solution According to the *Chinese Pharmacopoeia* (2020 Edition), each single herb, including Danshen (Salviae Miltiorrhizae Radix et Rhizoma), Niuxi (Achyranthis Bidentatae Radix), Mudanpi (Moutan Cortex), Chuanxiong (Chuanxiong Rhizoma), and Huomaren (Cannabis Fructus) was ground into coarse powder and accurately weighed. A total of 10 g of each herb was then extracted using 10 times its weight in 70% ethanol, performing the extraction process twice for 1 h each time. The five herb residues and the remaining seven herbs, including Tianma (Gastrodiae Rhizoma), Chishao (Paeoniae Radix Rubra), Dihuang (Rehmanniae Radix), Yinyanghuo (Epimedii Folium), Sangjisheng (Taxilli Herba), Zhizi (Gardeniae Fructus), and Juemingzi (Cassiae Semen) were extracted with 10 times their weight in water, conducted twice with the first extraction lasting for 1.5 h and the second for 1 h. The extracted solution from each herb was filtered, pooled, and concentrated separately to a clear paste at a concentration of 1.5 g/mL (equivalent to a raw drug concentration at 60 °C). Subsequently, 2 mL of this concentrated solution was evaporated to dryness, re-dissolved in 50% methanol, and then diluted to a final volume of 5 mL. This solution was then sonicated for 10 min and centrifuged at 12 000 r/min for 15 min at 4 °C. The supernatant of each herb was filtered through a 0.22 μm microporous membrane, and the filtrate was collected to create the test solution of the single herb.

2.3.3 Mixed standards solution The standard stock solutions were prepared by dissolving appropriate amounts of paeoniflorin, tanshinone IIA, tannic acid B, asparagine, ferulic acid, ecdysone, tannin, orange caspianin, chuanxiongzin, rhubarb phenol, epimedium, gardenia glucoside, quercetin, and catalpol in 5 mL of methanol, mixing them thoroughly and then diluting with methanol to obtain the final mixed solutions with concentrations of 0.081, 0.115, 0.123, 0.547, 0.043, 0.108, 0.263, 0.255, 0.448, 0.408, 0.255, 0.333, 0.127, and 0.553 mg/mL, respectively. The solution was filtered through a 0.22 μm microporous membrane before analysis.

2.4 Apparatus and chromatographic conditions

UPLC fingerprints were acquired on an Acquity UPLC H-Class system (Waters Corporation) equipped with a quaternary solvent manager, a PDA eλ detector, and a sample manager FTN-H at a detection wavelength of 240 nm and an injection volume of 3 μL. The chromatographic separation of samples was achieved by an Acquity UPLC HSS column (2.1 mm × 100 mm, 1.8 μm). The column was maintained at 30 °C with a 0.4 mL/min flow rate. The mobile phase consisted of A (0.1% phosphoric acid aqueous solution) and B (acetonitrile). The gradient elution program was optimized as follows: 0 – 4 min, 4.5% B; 4 – 8 min, 4.5% – 9% B; 8 – 10 min, 9% – 11% B; 10 – 12 min, 11% – 14% B; 12 – 16 min, 14% – 18% B; 16 – 19 min, 18% – 22% B; 19 – 24 min, 22% – 27% B; 24 – 29 min, 27% – 32% B; 29 – 40 min, 32% – 45% B; 40 – 45 min, 45% – 63% B; 45 – 50 min, 63% – 81% B; 50 – 53 min, 81% – 90% B; 53 – 55 min, 90% – 0% B. The test solutions were analyzed under these chromatographic conditions to obtain the fingerprints of DXG and its raw herbal materials.

2.5 Method validation

2.5.1 Precision Precision of the instrument was confirmed by the TQSM parameters, information entropy, and information amount of the fingerprints. To evaluate the precision, the sample solution of the same brand (S1 of DXG) was tested six times within a single day.

2.5.2 Repeatability Repeatability of the method was evaluated by calculating the TQSM parameters, information entropy, and information amount of the fingerprints using six independently extracted sample solutions (S1 of DXG).

2.5.3 Stability Stability of the sample solution was assessed using a single sample solution (S1 of DXG) stored

at room temperature for 0, 2, 4, 6, 8, 12, and 24 h, by calculating the TQSM parameters, information entropy, and information amount of the fingerprints.

2.6 Data analysis

The chromatographic parameters before integration were obtained using the Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine (version 2012A, Beijing, China). The TQSM parameters, information entropy, and information amount of fingerprints were calculated using Equations (1) – (5). The critical values of matching frequencies and retention time for dividing component families were computed using Equations (6) – (8). The new retention time and integrated component family peaks were calculated using Equations (9) and (10). The differences in these parameters before and after integration were assessed by Student's t test or one-way analysis of variance (ANOVA). $P < 0.05$ was considered statistically significant.

3 Results

3.1 Validation of the UPLC

3.1.1 Precision Analysis of six replicate injections showed ideal precision with RSD values of peak number (0.47%), AUC_T (0.35%), MRT_T (0.16%), VRT_T (0.46%), information entropy (0.09%), and information amount (0.39%).

3.1.2 Repeatability The method demonstrated good repeatability with RSD values of peak number (2.92%), AUC_T (1.47%), MRT_T (1.28%), VRT_T (1.12%), information entropy (0.66%), and information amount (1.25%).

3.1.3 Stability Sample solutions remained stable over 24 h, as evidenced by RSD values of peak number (1.49%), AUC_T (0.24%), MRT_T (0.16%), VRT_T (0.94%), information entropy (0.28%), and information amount (0.36%).

3.2 Establishment of chromatographic fingerprints of DXG and its raw herbal materials

UPLC analysis generated chromatographic fingerprints of DXG from 15 distinct batches (Figure 2). In the DXG fingerprint, a total of 389 peaks were detected, with each peak exhibiting an appearance frequency ranging from 1 to 15. Peaks that appeared with a matching frequency of 15 represented common peaks, resulting in 57 common peaks were identified in the chromatographic fingerprints of DXG (Figure 2A and 2B). Among these, 14 peaks were confirmed by matching their retention times with reference standards, which included catalpol, gastrodin, ligustrazine, geniposide, paeoniflorin, ferulic acid, ecdysone, salvianolic acid B, quercetin, paeonol, icariin, aurantio-obtusin, chrysophanol, and tanshinone IIA

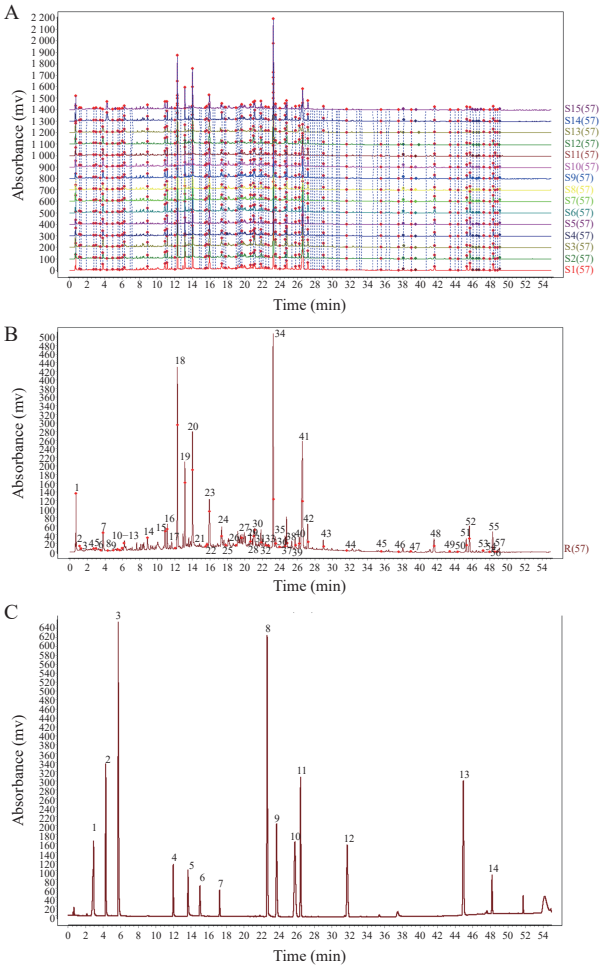

Figure 2 The UPLC fingerprints of DXG

A, similarity matching of UPLC fingerprints of 15 batches of DXG. B, referential chromatogram of DXG. C, UPLC fingerprints the mixed standards peaks. 1 – 14 represent different reference compounds: 1, catalpol; 2, gastrodin; 3, ligustrazine; 4, geniposide; 5, paeoniflorin; 6, ferulic acid; 7, ecdysone; 8, salvianolic acid B; 9, quercetin; 10, paeonol; 11, icariin; 12, aurantio-obtusin; 13, chrysophanol; 14, tanshinone IIA.

(Figure 2C). Similarly, the chromatographic fingerprints of raw herbal materials of DXG (15 batches) were generated (Figure 3). The TQSM parameters, information entropy, and information amount were calculated using Equations (1) to (5).

3.3 Application of the MFSM method for establishing integrated chromatographic fingerprints of DXG and its raw herbal materials

The chromatographic fingerprints of 15 DXG samples were utilized to illustrate how to reconstitute the integrated fingerprints by dividing and compressing the original fingerprints and how to calculate their new chromatographic parameters using the MFSM method (Table 1).

3.3.1 Finding the marker peak using the boundary value of matching frequencies Building upon the findings outlined in section 3.2, the total number of peaks, their

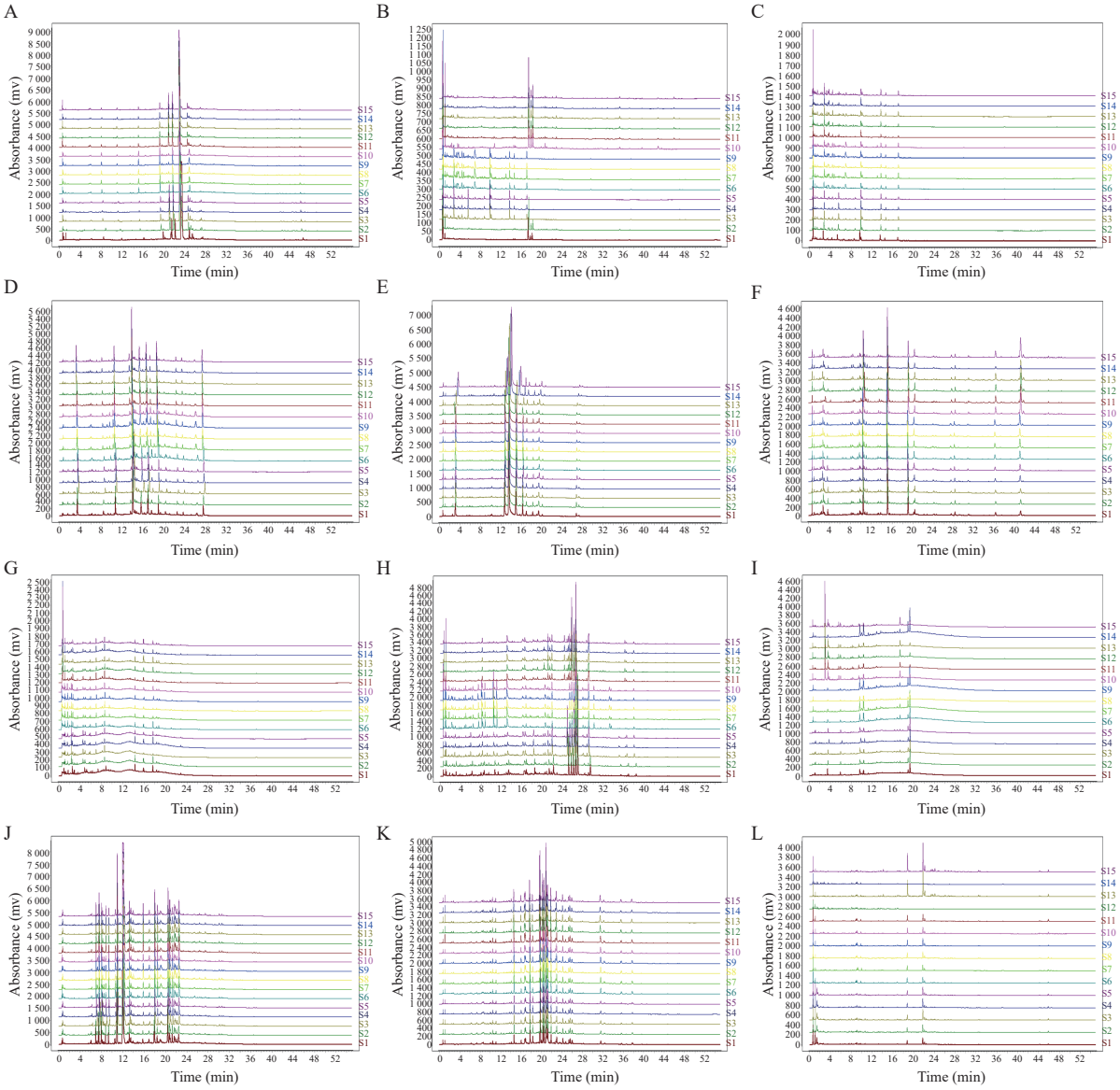

Figure 3 The fingerprints of the raw herbal materials of DXG

A, Danshen (Salviae Miltiorrhizae Radix et Rhizoma). B, Niuxi (Achyranthis Bidentatae Radix). C, Tianma (Gastrodiae Rhizoma). D, Mudanpi (Moutan Cortex). E, Chishao (Paeoniae Radix Rubra). F, Chuanxiong (Chuanxiong Rhizoma). G, Dihuang (Rehmanniae Radix). H, Yinyanghuo (Epimedii Folium). I, Sangjisheng (Taxilli Herba). J, Zhizi (Gardeniae Fructus). K, Juemingzi (Cassiae Semen). L, Huomaren (Cannabis Fructus).

retention times, peak areas, and matching frequencies were obtained (Table 1). Statistical analysis of the matching frequencies of the chromatographic peaks was then performed according to methodology described in section 2.2.1: (i) the $\overline{M_p}$ and its SD were 8.09 and 4.14, respectively; (ii) the confidence level $(1 - \alpha)$ was set at 0.95 and the degree of freedom (v) at 3 146 – 1= 3 145; accordingly, $t_{(\alpha,v)}$ was given to be 1.96; (iii) the critical value of the matching frequency for dividing component families was computed as 16.20 (M_{pc}) by inputting the mean and SD of the matching frequencies into Equation (6). Theoretically, the critical value of the matching frequencies should be no more than 15, the total number of batches. Therefore, the critical value was corrected to be 15 rather than the aforementioned 16.20. Based on the actual critical value, the chromatographic peaks with a matching frequency greater than or equal to 15 were regarded as the marker peaks for dividing component families. The next step involves determining the retention time boundary to segregate peak families, utilizing the two marker peaks as references to ensure that this boundary falls between them.

3.3.2 Dividing and integrating component families using the boundary value of retention time The boundary value of retention time for splitting two component families was calculated using Equation (7) or (8). For example, t_{Rpc1} for dividing N_{pc1} and N_{pc2} was calculated to be 0.85. This value was positioned between the retention time of N_{P4} and N_{P5}. Therefore, a division line for t_{Rpc1} was drawn on the chromatogram to separate the peaks N_{P4} and N_{P5} (Table 1).

Table 1 Examples of dividing original fingerprints of DXG to integrate its novel fingerprints

Peak (family) number	Peak area or peak family area												...		
	1	**2**	**3**	**4**	Npc_1	**5**	**6**	Npc_2	**7**	**8**	**9**	**10**	...	**389**	Npc_{57}
S1	0	142.61	376.92	0	519.53	0	81.75	81.75	158.31	0	153.41	0	...	1	7.55
S2	0	145.16	411.93	0	557.08	0	87.42	87.42	192.69	0	152.23	9.12	...	0	52.98
S3	0	96.67	370.77	0	467.44	0	68.2	68.2	192.28	0	116.38	0	...	0	54.32
S4	0	144.75	356.83	0	501.57	0	87.73	87.73	157.03	24.99	111.62	0	...	0	66.70
S5	0	158.12	445.23	0	603.35	0	100.21	100.21	187.73	22.86	186.67	0	...	0	47.79
S6	0	121.92	408.06	0	529.98	0	68.43	68.43	171.13	0	112.77	0	...	0	41.70
S7	1.46	152.51	536.05	0	690.03	0	94.76	94.76	189.79	35.84	238.51	0	...	0	131.54
S8	0	0	467.84	112.12	579.96	0	70.74	70.74	139.77	23.59	115.39	0	...	0	191.84
S9	0	0	463.17	103.17	566.34	0	63.71	63.71	122.31	21.36	126.89	0	...	0	175.52
S10	0	94.03	395.41	0	489.44	0	95.78	95.78	155.63	23.64	111.07	0	...	1.81	132.96
S11	0	0	413.14	77.74	490.87	0	88.12	88.12	136.68	24.72	101.91	0	...	0	141.45
S12	0	90.91	373.05	0	463.96	0	91.93	91.93	167.18	1.65	110.57	0	...	0	128.40
S13	0	124.26	441.49	0	565.75	0	101.74	101.74	201.36	0	142.53	0	...	0	84.76
S14	0	132.44	365.05	0	497.49	42.81	49.83	92.65	143.38	41.54	0	125.08	...	0	66.69
S15	0	0	446.13	118.43	564.57	0	69.39	69.39	149.2	0	171.72	0	...	0	12.97
Matching frequency	1	11	15	4		1	15		15	9	14	2	...	1	
Retention time (min)	0.05	0.62	0.69	0.73	0.65	1.02	1.13	1.12	1.29	1.62	1.91	1.97	...	54.05	51.32

Retention time pertains to the duration that chromatographic peaks or peak families persist in a chromatogram. Number refers to the count of these chromatographic peaks or peak families. Npc_1 refers to the 1st peak family. Npc_2 refers to the 2nd peak family. Npc_{57} refers to the 57th peak family.

$$t_{Rpc1} = \sum_{j=3}^{6} N_{P,j} t_{R,j} \Big/ \sum_{j=3}^{6} N_{P,j} =$$
$$(0.69 \times 15 + 0.73 \times 4 + 1.1 \times 1 + 1.13 \times 15) /$$
$$(15 + 4 + 1 + 15) = 0.89 \text{ min}$$

Similarly, the t_{Rpc2} for dividing N_{pc2} and N_{pc3} was drawn directly between N_{P6} and N_{P7}, because the matching frequencies of the two peaks were both 15.

With the same method, t_{Rpc3} was determined to be 2.11, and t_{Rpc56} was calculated between N_{pc56} and N_{pc57}.

Using these boundary values of retention time, a total of 389 chromatographic peaks were identified from the chromatographic fingerprints of the 15 DXG batches and re-integrated into 57 peak families, each containing of a common peak (Table 1).

3.3.3 Calculation of the new chromatographic parameters of component families The calculation of new chromatographic parameters, including the new retention time and the integrated peak family area, was performed using Equations (9) and (10), respectively. For example, the new retention time for N_{pc1} in sample S1 (t_{Rp1}) was computed as follows.

$$t_{Rp1} = \sum_{j=1}^{4} N_{P,j} t_{R,j} \Big/ \sum_{j=1}^{4} N_{P,j} =$$
$$(0.05 \times 1 + 0.62 \times 11 + 0.69 \times 15 + 0.73 \times 4) /$$
$$(1 + 11 + 15 + 4) = 0.65 \text{ min}$$

Here, N_{pc1} refers to all the components eluted from 0 to 0.65 min. The total integrated peak area for component families can be calculated by addition. For example, the new peak family area for N_{pc1} in sample S1 (A_{Rpi}) was determined as follows.

$$A_{Rp1} = \sum_{j=1}^{4} A_{P,j} = (0 + 142.61 + 376.92 + 0) = 519.53 \ \mu\text{v} \cdot \text{s}$$

Similarly, the new peak family areas for N_{pc1} in S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, and S15 were 557.08, 467.44, 501.57, 603.35, 529.98, 690.03, 579.96, 566.34, 489.44, 490.87, 463.96, 565.75, 497.49, and 564.57 μv·s, respectively. In summary, the new chromatographic parameters of the 57 peak families in the integrated fingerprints of DXG were obtained with the use of the MFSM method.

Additionally, the chromatographic parameters of the integrated fingerprints of its raw herbal materials were obtained, with the following counts of component families identified for each material: 10, 33, 29, 20, 22, 28, 33, 30, 32, 75, 88, and 45 for Danshen (Salviae Miltiorrhizae Radix et Rhizoma), Niuxi (Achyranthis Bidentatae Radix), Tianma (Gastrodiae Rhizoma), Mudanpi (Moutan Cortex), Chishao (Paeoniae Radix Rubra), Chuanxiong (Chuanxiong Rhizoma), Dihuang (Rehmanniae Radix), Yinyanghuo (Epimedii Folium), Sangjisheng (Taxilli

Herba), Zhizi (Gardeniae Fructus), Juemingzi (Cassiae Semen), and Huomaren (Cannabis Fructus), respectively.

3.4 Validation and evaluation for the MFSM method

3.4.1 Comparison of the properties between original and integrated fingerprints of DXG and its raw herbal materials The peak number, three TQSM parameters (AUC_T, MRT_T, and VRT_T), information entropy, and information amount for the peaks or peak families of UPLC fingerprints of DXG and its 12 raw herbal materials were calculated and compared before and after integration (Figure 4). Before integration, the ranges of the six parameters of DXG and its raw materials were 95.07 – 209.73, 9 390 – 183 064 μv·s, 5.928 – 21.33 min, 22.62 – 106.69 min^2, 4.230 – 6.539, and 50 530 – 974 186 μv·s, respectively. After integration, the ranges of these parameters were 10.00 – 88.00, 9 390 – 183 064 μv·s, 5.951 – 22.02 min, 22.27 – 104.73 min^2, 2.223 – 5.277, and 38 159 – 807 200 μv·s, respectively. The TQSM parameters were not statistically significant after integration (all $P > 0.05$), whereas the number of peaks, information entropy, and information amount significantly decreased after integration (all $P < 0.05$). The results showed that the MFSM method significantly reduced the number of analytical units for the material basis by consolidating the chromatographic information, while maintaining the original TQSM parameters. In addition, both information entropy and information amount decreased as a result of the reduced total peak number. A clear correlation between these parameters was established, as detailed in Equations (4) and (5). The TQSM parameters are commonly used to characterize the functional properties of chromatographic fingerprints. In line with literature findings [14-17], information entropy and information amount were employed to express the information of chromatographic fingerprints. The results demonstrated that the analytical units of the material basis could be reduced while maintaining their properties of original fingerprints.

3.4.2 Comparison of the variability between original and integrated fingerprints of DXG and its raw herbal materials To compare the variability of fingerprints between DXG and its 12 raw herbal materials, the RSD of peak number, TQSM parameters, information entropy, and information amount was calculated and compared before and after integration, respectively (Figure 5). As Figure 5A displays, before integration, the RSD of peak number of DXG and its raw herbal materials ranged from 2.12% to 9.15%; after integration, it was 0.00%. As shown in Figure 5B – 5F, before integration, the RSD ranges for the three TQSM parameters, information entropy, and information amount were 6.04% – 49.78%, 1.15% – 23.10%, 3.97% – 25.79%, 1.49% – 19.86%, and 6.64% – 51.20%, respectively; after integration, those ranges were 6.04% – 49.87%, 1.73% – 23.02%, 3.84% – 26.85%, 1.17% – 16.54%, and 6.40% – 48.59%, respectively.

Moreover, *t* test was employed to statistically evaluate the differences and similarities of the aforementioned parameters before and after integration. As shown in Table 2, the RSD of peak number for DXG and its raw herbal materials significantly decreased following integration ($P < 0.05$). In contrast, there were no significant differences in the RSD values of the three TQSM parameters, information entropy, and information amount (all $P > 0.05$), indicating that these parameters remained stable after integration.

The RSD of the three TQSM parameters, information

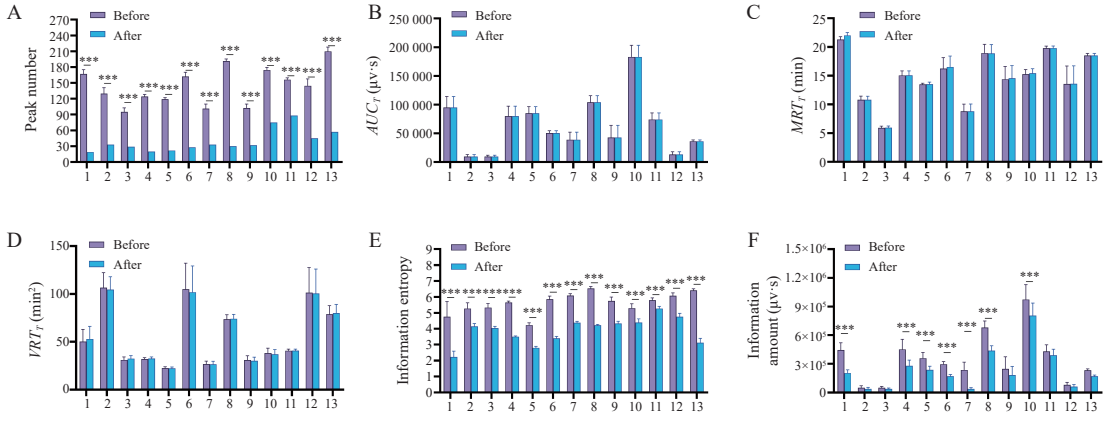

Figure 4 Comparison of peak number, TQSM parameters, information entropy, and information amount of the UPLC fingerprints before and after integration

A, peak number. B, AUC_T. C, MRT_T. D, VRT_T. E, information entropy. F, information amount. 1 – 13 represents Danshen (Salviae Miltiorrhizae Radix et Rhizoma), Niuxi (Achyranthis Bidentatae Radix), Tianma (Gastrodiae Rhizoma), Mudanpi (Moutan Cortex), Chishao (Paeoniae Radix Rubra), Chuanxiong (Chuanxiong Rhizoma), Dihuang (Rehmanniae Radix), Yinyanghuo (Epimedii Folium), Sangjisheng (Taxilli Herba), Zhizi (Gardeniae Fructus), Juemingzi (Cassiae Semen), Huomaren (Cannabis Fructus), and DXG, respectively. Data were represented as mean ± SD ($n = 15$). $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$, comparison of these parameters before and after integration.

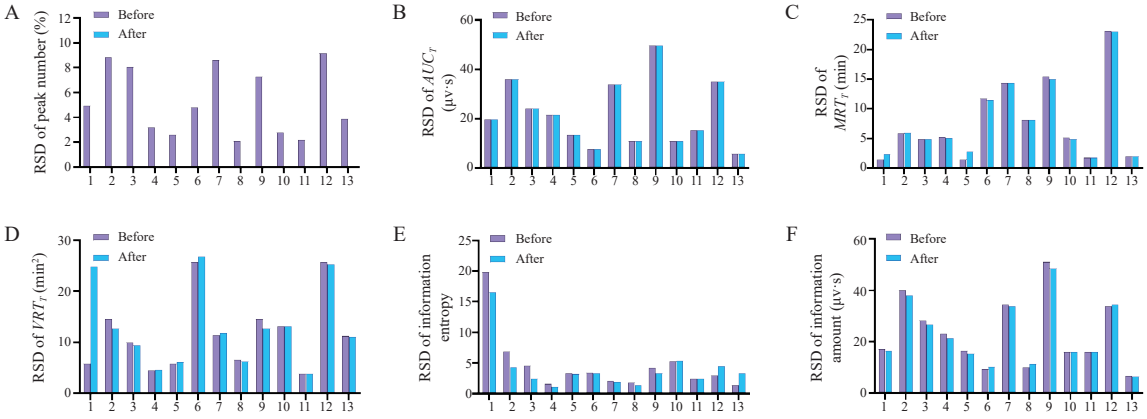

Figure 5 The RSD of peak number, TQSM parameters, information entropy, and information amount of UPLC fingerprints before and after integration

A, the RSD of peak number. B, the RSD of AUC_T. C, the RSD of MRT_T. D, the RSD of VRT_T. E, the RSD of information entropy. F, the RSD of information amount. 1 – 13 represents Danshen (Salviae Miltiorrhizae Radix et Rhizoma), Niuxi (Achyranthis Bidentatae Radix), Tianma (Gastrodiae Rhizoma), Mudanpi (Moutan Cortex), Chishao (Paeoniae Radix Rubra), Chuanxiong (Chuanxiong Rhizoma), Dihuang (Rehmanniae Radix), Yinyanghuo (Epimedii Folium), Sangjisheng (Taxilli Herba), Zhizi (Gardeniae Fructus), Juemingzi (Cassiae Semen), Huomaren (Cannabis Fructus), and DXG, respectively.

Table 2 RSD of TQSM parameters, information entropy, and information of UPLC fingerprints of DXG and its 12 raw herbal materials before and after integration

Parameter	Degree of freedom (ν)	*t*	*P*
RSD of peak number	12	6.937 1	0.001 0
RSD of AUC_T	12	1.000 0	0.337 0
RSD of MRT_T	12	0.493 4	0.630 7
RSD of VRT_T	12	0.795 9	0.441 6
RSD of information entropy	12	1.271 9	0.227 5
RSD of information amount	12	1.778 4	0.100 7

$t_{(0.05,12)} = 2.179$; $t_{(0.01,12)} = 3.055$; $\alpha = 0.05$ (0.01). Typically, the confidence level $(1 - \alpha)$ was set at 95% or 99%, which implied that α was set at 2.5% or 0.005%, respectively, on one side. And $t(\alpha,\nu)$ was the t critical value under the confidence coefficient $(1 - \alpha)$ and the degree of freedom (ν). When the t value is greater than the t critical value, it indicates significant difference, vice versa.

entropy, and information amount are traditionally employed to reflect the variability in the chromatographic fingerprints. All the results demonstrated that the fluctuation range of analytical units was markedly reduced after integration, while the variability of the original fingerprints among samples was maintained.

3.4.3 Comparison of the peak area variability of the original and integrated fingerprints of DXG and its raw herbal materials To further investigate the variability of the peak area before and after integration, the RSD of average peak area, the RSD variance of peak area, the maximum RSD of peak area, and minimum RSD of peak area were compared and analyzed in this study. The results are shown in Figure 6. As shown in Figure 6A and 6B, the RSD of average peak area and RSD variance of peak area decreased overall after integration. In particular, as shown in Figure 6C and 6D, all the maximum RSD values of peak area of DXG and its herbal materials were reduced after integration (before integration, 140.1% – 281.22%; after integration, 67.61% – 202.97%), and all the minimum RSD values of peak area were elevated (before integration, 0% – 1.34%; after integration, 5.89% – 22.97%).

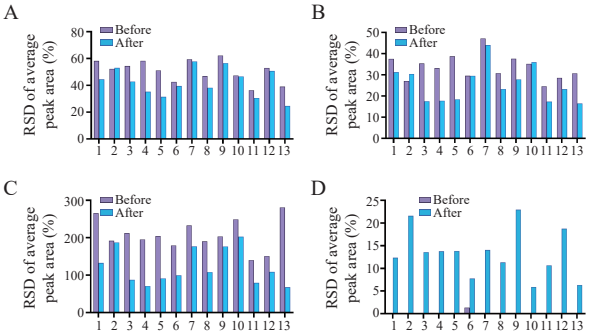

Figure 6 Comparison of the RSD values of the chromatographic peak areas of the UPLC fingerprints of DXG and its 12 raw herbal materials before and after integration

A, the RSD of average peak area. B, the RSD variance of peak area. C, the maximum RSD of peak area. D, the minimum RSD of peak area. 1 – 13 represents Danshen (Salviae Miltiorrhizae Radix et Rhizoma), Niuxi (Achyranthis Bidentatae Radix), Tianma (Gastrodiae Rhizoma), Mudanpi (Moutan Cortex), Chishao (Paeoniae Radix Rubra), Chuanxiong (Chuanxiong Rhizoma), Dihuang (Rehmanniae Radix), Yinyanghuo (Epimedii Folium), Sangjisheng (Taxilli Herba), Zhizi (Gardeniae Fructus), Juemingzi (Cassiae Semen), Huomaren (Cannabis Fructus), and DXG, respectively.

To statistically evaluate the differences in the aforementioned parameters, t test was performed on the data before and after integration (Table 3). The results indicated a significant reduction in the RSD of both average peak area and the variance of peak area (all $P < 0.05$), suggesting a more consistent peak area after integration. Furthermore, the range of RSD values for peak area (defined as the difference between the maximum and minimum

Table 3 RSD of the chromatographic peak areas of UPLC fingerprints of DXG and its 12 raw herbal materials before and after integration

Parameter	Degree of freedom (v)	t	P
RSD of average peak area	12	3.2548	0.0069
RSD variance of peak area	12	4.0661	0.0016
Maximum RSD of peak area	12	5.2066	0.0002
Minimum RSD of peak area	12	8.8055	0.0001

$t_{(0.05,12)} = 2.179$; $t_{(0.01,14)} = 3.055$; $\alpha = 0.05$ (0.01). Typically, the confidence level $(1 - \alpha)$ was set at 95% or 99%, which implied that α was set at 2.5% or 0.005%, respectively, on one side. And $t(\alpha,v)$ was the t critical value under the confidence coefficient $(1 - \alpha)$ and the degree of freedom (v). When the t value is greater than the t critical value, it indicates significant difference, vice versa.

RSD values) narrowed significantly after integration (all $P < 0.001$). The results showed that the variability in the chromatographic peak areas was notably reduced after integration.

4 Discussion

Unlike chemical drugs, the chemical composition and content of CMM are characterized by their variability and dynamism, influenced by a variety of factors. This variability complicates the quality evaluation of CMM. To mitigate this complexity, chromatographic fingerprinting techniques are often employed to identify characteristics or common components. However, these methods do not fully encapsulate the holistic quality of the herbal materials, as common components may not always be the active medicinal components, and non-common components could potentially be medicinal as well [19]. Relying solely on common components for comparison may not accurately represent the material's overall medicinal basis [20]. In light of this, the MFSM method was developed. This approach integrates chromatographic peaks into peak families without omitting any information, with the goal of minimizing the number of analytical units required for evaluation on material basis.

The study results revealed that the newly integrated fingerprint exhibited a significant reduction in the analytical units of material basis, information entropy, and information amount. In contrast, the TQSM parameters, which are essential for quality assessment, remained stable. Additionally, the RSD of the TQSM parameters, information entropy, and information amount were not significantly different before and after integration, and the RSD of both the number and area for the integrated analytical units decreased markedly. In essence, even though the integration process significantly reduced the number of analytical units and the amount of chromatographic fingerprint information, the TQSM properties of the chromatographic fingerprint and sample-to-sample variability were preserved. Components that were grouped into the same "imprinting template" family exhibited similar parent nuclei, chemical structures, comparable physical and chemical properties, and similar chromatographic behaviors. The retention time aligned closely with the component exhibiting structural characteristics most akin to the imprinting template, and the likelihood of modifying it to match these characteristics increased, accompanied by a high frequency of occurrence. The likelihood of modifying it into a component far away from the structural characteristics of the imprinting template was low, with a low frequency of occurrence. Finally, in chromatography, the distribution exhibited clustering centered around the component most akin to the imprinting template structural characteristics, accompanied by a convex-concave pattern in the retention time based on chromatographic peaks. The number of imprinting templates was determined through statistical analysis of the frequency of recurring chromatographic peaks in the fingerprint spectra.

The MFSM method facilitated the correlation between analytical units and medicinal effects, enabling the discovery of additional medicinal effects and further unveiling novel medicinal components. At present, spectrum-effect correlation primarily focuses on known components, while often neglected. Recognizing that the pharmacological activity of a CMM stems from all its components, our method enables the screening of the most bioactive components as markers, encompassing both known and unknown components. Spectrum-effect correlation involves identifying the correlation between the common or characteristic peaks in the fingerprint and pharmacodynamic indicators to locate effective components. However, this approach can inadvertently overlook non-common peaks with high activity but low content, thereby compromising the holistic nature of CMM. Recently, scholars have established robust models to associate bio-activity with specific chromatographic regions by dividing the entire chromatographic profiles into distinct segments. These models, such as partial least squares (PLS) and uninformative variable elimination PLS regression, multiple linear regression and artificial neural network, have been employed to predict the region features responsible for bio-activity. Subsequently, a quantitative structure-activity relationship approach has been introduced [21-23]. However, previous studies divided chromatographic regions solely based on retention time. In contrast, our study divided chromatographic regions considering both retention time and the matching frequency of chromatographic peaks.

The established MFSM method not only facilitates the transformation of CMM quality evaluation from a

characteristic component pattern to a component family pattern, but also provides an efficient pathway for discovering potential drug leads within CMM. Our team has secured a patent based on the aforementioned MFSM method, which divided and integrates CMM component families into a novel fingerprint characterized by component families as the unit [24]. This fingerprint retains the total statistical moment characteristics and inter-sample variability of the original fingerprint while significantly reducing the number of chromatographic peaks [24]. Furthermore, we have successfully applied this method to identify common component families in pungent herbs and screen for key components that cause sensitization in Houttuynia cordata injection [25, 26]. Our research team has established mathematical models that explore the correlation between component families and efficacy, akin to the spectrum-effect correlation but with a reduced number of analytical units [27]. In the future, we aim to further separate component families using preparative chromatography, correlate them with pharmacodynamics, and then perform spectrum-quantity correlation and spectrum-efficacy correlation on CMM fingerprints to discover more effective or toxic components. However, this method also has its limitations. For example, variations in analytical conditions, including chromatographic columns and conditions, can result in the identification of different component families.

5 Conclusion

The study demonstrated that the MFSM method effectively integrated chromatographic peaks, reducing analytical units while maintaining the original fingerprint properties and variability. Furthermore, after integration, the number and area of peak families became more consistent among batches. Therefore, the proposed MFSM method might serve as a particularly appropriate and feasible tool for analyzing complex chromatographic fingerprints, simplifying evaluation of CMM quality and facilitating the link between efficacy and component families. Additionally, this paper establishes a foundation for transforming CMM quality evaluation from a pattern based on indicator components, including quality markers, to a pattern centered on component families.

Fundings

Natural Science Foundation of Hunan province (2022JJ30453 and 2024JJ6362), and the Key Research and Development Program of Hunan Province (2022SK2014).

Competing interests

The authors declare no conflict of interest.

References

[1] CAI ZC, WANG CC, CHEN CH, et al. Quality evaluation of Lonicerae Japonicae Flos and Lonicerae Flos based on simultaneous determination of multiple bioactive constituents combined with multivariate statistical analysis. Phytochemical Analysis, 2021, 32(2): 129–140.

[2] XU HY, ZHANG YQ, WANG P, et al. A comprehensive review of integrative pharmacology-based investigation: a paradigm shift in traditional Chinese medicine. Acta Pharmaceutica Sinica B, 2021, 11(6): 1379–1399.

[3] PAN X, LI HY, CHEN DF, et al. Comparison of essential oils of *Houttuynia cordata* Thunb. from different processing methods and harvest seasons based on GC-MS and chemometric analysis. International Journal of Analytical Chemistry, 2021, 2021: 8324169.

[4] LAN LL, SUN WY, CHANG Q, et al. Comprehensive evaluation of Licorice extract by five-dimensional quantitative profiling. Journal of Chromatography A, 2021, 1644: 462105.

[5] ZHANG JL, GONG DD, LAN LL, et al. Comprehensive evaluation of Loblolly fruit by high performance liquid chromatography four wavelength fusion fingerprint combined with gas chromatography fingerprinting and antioxidant activity analysis. Journal of Chromatography A, 2022, 1665: 462819.

[6] HOU JJ, CAO CM, XU YW, et al. Exploring lipid markers of the quality of coix seeds with different geographical origins using supercritical fluid chromatography mass spectrometry and chemometrics. Phytomedicine, 2018, 45: 1–7.

[7] JIANG M, HAN YQ, ZHOU MG, et al. The screening research of anti-inflammatory bioactive markers from different flowering phases of Flos Lonicerae Japonicae. PLoS One, 2014, 9(5): e96214.

[8] NOVIANA E, INDRAYANTO G, ROHMAN A. Advances in fingerprint analysis for standardization and quality control of herbal medicines. Frontiers in Pharmacology, 2022, 13: 853023.

[9] LIU CL, ZUO ZT, XU FR, et al. Authentication of herbal medicines based on modern analytical technology combined with chemometrics approach: a review. Critical Reviews in Analytical Chemistry, 2023, 53(7): 1393–1418.

[10] CHATTERJEE NS, CHEVALLIER OP, WIELOGORSKA E, et al. Simultaneous authentication of species identity and geographical origin of shrimps: untargeted metabolomics to recurrent biomarker ions. Journal of Chromatography A, 2019, 1599: 75–84.

[11] YANG WZ, ZHANG YB, WU WY, et al. Approaches to establish Q-markers for the quality standards of traditional Chinese medicines. Acta Pharmaceutica Sinica B, 2017, 7(4): 439–446.

[12] WANG MC, HE QJ, ZHOU YZ, et al. Study on one-time steady-state feeding quantity of UPLC fingerprint of composition of function components of Danxi Granules. Chinese Traditional and Herbal Drugs, 2022, 53(3): 704–711.

[13] FAN QM, LIU WL, YANG YT, et al. A new similarity method for assessment of pharmacokinetic interaction between flucloxacillin and midazolam. Die Pharmazie, 2019, 74(7): 397–405.

[14] HAIYING LI, MEIFENG XIAO, XUE PAN, et al. Comparative

polypharmacokinetics of nine anti-inflammatory components of Jinyinhua (Lonicerae Japonicae Flos) and Shanyinhua (Lonicerae Flos) in mice with p-xylene-induced ear edema. Digital Chinese Medicine, 2023, 6(1): 73–85.

[15] ZHU JP, ZHANG XL, ZHAO J, et al. The components transitive regularity of three dosage forms of Liuwei Dihuang Fufang. Digital Chinese Medicine, 2022, 5(1): 68–74.

[16] HE FY, ZHOU HH, DENG KW, et al. A new qualitative and quantitative analytical method of chromatographic fingerprints: total quantum statistical moment. Acta Pharmaceutica Sinica, 2008, 43(2): 195–201.

[17] GONG F, LIANG YZ, XIE PS, et al. Information theory applied to chromatographic fingerprint of herbal medicine for quality control. Journal of Chromatography A, 2003, 1002(1/2): 25–40.

[18] LI HY, HE QJ, DENG KW, et al. Constructing a new quality imprinting evaluation system for traditional Chinese medicine Q-Marker based on supramolecular “qi-xi” theory. Chinese Traditional and Herbal Drugs, 2021, 52(16): 4771–4778.

[19] RANJAN S, ADAMS E, DECONINCK E. Multidimensional chromatographic fingerprinting combined with chemometrics for the identification of regulated plants in suspicious plant food supplements. Molecules, 2023, 28(8): 3632.

[20] SHARMA V, CHIB S, KUMARI D, et al. Correction: chromatographic fingerprinting of epiphytic fungal strains isolated from *Withania somnifera* and biological evaluation of isolated okaramine H. Analytical Methods, 2024, 16(38): 6577–6578.

[21] DUMAREY M, SMETS I, VANDER HEYDEN Y. Prediction and interpretation of the antioxidant capacity of green tea from dissimilar chromatographic fingerprints. Journal of Chromatography B, 2010, 878(28): 2733–2740.

[22] XU GL, XIE M, YANG XY, et al. Spectrum-effect relationships as a systematic approach to traditional Chinese medicine research: current status and future perspectives. Molecules, 2014, 19(11): 17897–17925.

[23] CHAU FT, CHAN HY, CHEUNG CY, et al. Recipe for uncovering the bioactive components in herbal medicine. Analytical Chemistry, 2009, 81(17): 7217–7225.

[24] HE FY, WANG MC, HE QJ, et al. Method and product of clustering integration of traditional Chinese medicine imprinting template components based on chromatographic peak matching frequency statistics of fingerprints. Hunan Province: CN114034806A, 2022-02-11.

[25] QIAN XB, ZHANG LQ, XIAO Y, et al. Characteristic analysis of “imprinting template” for pungent herbs based on molecular connectivity index and matching frequency total statistical moment. Chinese. Journal of Experimental Traditional Medical Formulae, 2023, 29(9): 218–224.

[26] LIU X, WANG YC, ZHANG YT, et al. Supramolecular structure characteristics of allergens in Houttuynia cordata injection based on isothermal titration calorimetry. Chinese Traditional and Herbal Drugs, 2023, 54(17): 5530–5540.

[27] PAN X. Study on the division of volatile constituents of *Fritillaria asiatica* into “substance units” and their “quantitative-constructive-effective” relationships. Changsha: Hunan University of Chinese Medicine, 2022.

一种整合中药材色谱指纹图谱分析单元的新方法：匹配频率统计矩法

李海英[a, b†], 潘雪[b, c†], 王敏存[b, c], 李文姣[b, c], 贺鹏[b, c], 黄胜[d], 贺福元[b, c, e, f*]

a. 湖南中医药大学第一附属医院制剂中心, 湖南 长沙 410007, 中国
b. 中药成药性与制剂制备湖南省重点实验室, 湖南 长沙 410208, 中国
c. 湖南中医药大学药学院, 湖南 长沙 410208, 中国
d. 九芝堂股份有限公司, 湖南 长沙 410205, 中国
e. 湖南中医药大学中医药超分子机理与数理特征化实验室, 湖南 长沙 410208, 中国
f. 中药药性与药效国家中医药管理局重点实验室, 湖南 长沙 410208, 中国

【摘要】目的 开发并应用匹配频率统计矩（MFSM）新方法，促进适合于中药（CMM）独特性的质量评价。**方法** 本研究建立 MFSM 方法，并以丹膝颗粒（DXG）及其原料药为例对该方法进行验证。首先，使用超高效液相色谱法（UPLC）获得 DXG 及其原料药的色谱指纹图谱，并采用 MFSM 方法将其压缩和整合为具有较少分析单元的新指纹图谱。然后，通过计算总量统计矩（TQSM）参数、信息熵和信息量及其相对标准偏差（RSD）来表征 DXG 及其原料药原始和整合后指纹图谱的性质和变异性。最后，为了验证 MFSM 的有效性，比较 DXG 及其原料药原始和整合后指纹图谱的 TQSM 参数、信息熵和信息量及其 RSD。**结果** 通过 MFSM 将 DXG 及其 12 种原料药材的色谱峰划分并整合为峰族。在整合之前，DXG 及其 12 种原料药材的 UPLC 指纹图谱峰数、3 种 TQSM 参数、信息熵和信息量的范围分别为 95.07 – 209.73、9 390 – 183 064 μv·s、5.928 – 21.33 min、22.62 – 106.69 min^2、4.230 – 6.539 和 50 530 – 974 186 μv·s。整合后，这些参数的范围分别为 10.00 – 88.00、9 390 – 183 064 μv·s、5.951 – 22.02 min、22.27 – 104.7 $3min^2$、2.223 – 5.277 和 38 159 – 807 200 μv·s。相应地，上述所有参数在整合前的 RSD 分别为 2.12% – 9.15%、6.04% – 49.78%、1.15% – 23.10%、3.97% – 25.79%、1.49% – 19.86% 和 6.64% – 51.20%；整合后，它们分别为 0.00% 、6.04% – 49.87%、1.73% – 23.02%、3.84% – 26.85%、1.17% – 16.54% 和 6.40% – 48.59%。结果表明，在新整合后的指纹图谱中，物质分析单元、信息熵和信息量显著减少（$P < 0.05$），而 TQSM 参数保持不变（$P > 0.05$）。此外，TQSM 参数、信息熵和信息量的 RSD 在整合前后没有显著差异（$P > 0.05$），整合后的分析单元的数目和面积的 RSD 则显著降低（$P < 0.05$）。**结论** 采用 MFSM 方法，可以在保留 DXG 及其原料药自身原有的指纹图谱性质和变异性基础上，减少物质分析单元。因此，MFSM 方法可以作为降低多组分中药分析难度的可行和可靠工具，从而有利于质量评价。

【关键词】色谱指纹图谱；分析单元；匹配频数统计矩法；中药；丹膝颗粒；质量评价