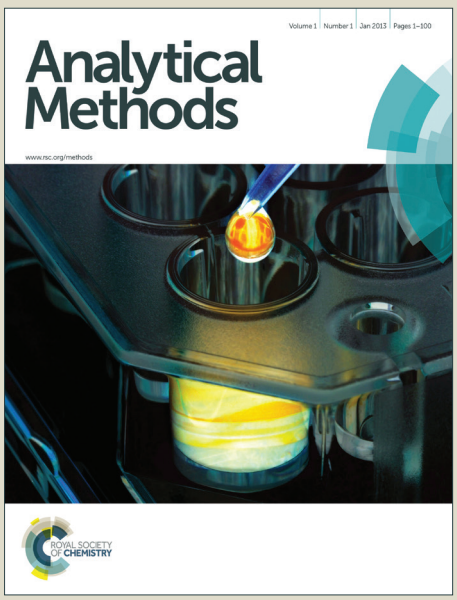
Analytical Methods

Accepted Manuscript



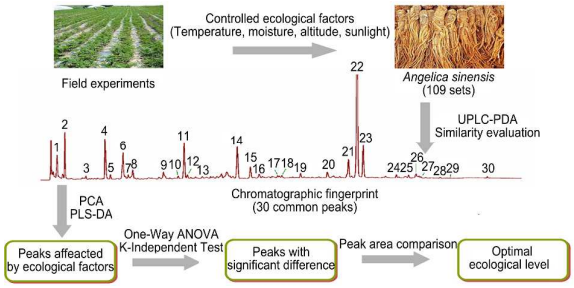
ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Multivariate statistical analysis based on chromatographic fingerprint for the evaluation of important ecological factors on the quality of *Angelica sinensis*

Xin-Yue Song[a, c], Ling Jin[b], Yan-Ping Shi[a], Ying-Dong Li[b] and Juan Chen[a]

A holistic strategy combing chromatographic fingerprint and multivariable statistical analysis was developed to evaluate the effects of ecological factors on the quality of A.*sinensis* and then chose the optimum one.



Multivariate statistical analysis based on chromatographic fingerprint for the evaluation of important ecological factors on the quality of *Angelica sinensis*

Xin-Yue Song[a, c], Ling Jin[b], Yan-Ping Shi[a], Ying-Dong Li[b*] and Juan Chen[a*]

Abstract: Multiple components in traditional Chinese medicines (TCMs) have synergistic action on affecting therapeutic effects of TCMs and their contents may vary substantially with environment changing. In this study, an ultra-performance liquid chromatographic (UPLC) fingerprint was established to choose the optimum ecological level for the cultivation of *Angelica sinensis (A. sinensis)*. Optimum separation was achieved on a C_{18} column (50×2.1 mm i.d., 1.7 μm particle) through a 25 min gradient. And then the developed method was applied to establish the chromatographic fingerprint of *A. sinensis* by analyzing 109 samples cultivated under controlled ecological factors. Representative standard fingerprint chromatogram was obtained by the professional software in which 30 common peaks were marked. The common peaks information of all samples was submitted to principal component analysis (PCA) with consequent partial least squares discriminant analysis (PLS-DA) to screen out the peaks relating specific ecological factor. Peaks with their peak areas having significant differences in different ecological levels were screened out and used to obtain the optimum ecological level. This novel method integrating the advantages of chromatographic

fingerprint and multivariate statistical analysis offer integral characteristics of an herbal medicine. Consequently, it is a more comprehensive and scientific method, providing a technical safeguard for the cultivation of herbal medicines.

Keywords: chromatographic fingerprint, *Angelica sinensis*, ecological factors, UPLC-PDA, common peaks.

1. Introduction

Traditional Chinese medicines (TCMs) have a long therapeutic history over thousands of years, and currently it is still attracting ever-increasing attention worldwide.[1] It is well known that TCMs, either presenting as a single herb or as a collection of herbs in composite formulae, are complex mixture containing hundreds of chemically different constituents which are usually responsible for the therapeutic effects.[2] And the type and amount of these chemical constituents may vary substantially with its growing environment. Therefore, developing sensitive and effective methods to evaluate the effects of the main ecological factors and then choose the optimum one were meaningful. Currently, there are two widely-used methods to evaluate the effects of ecological factors on the quality of herb medicines. One is a common practice among natural products analysts to select one or more compounds as either active compounds or "markers" for purposes of factors assessment. However, therapeutic effects of TCMs are based on the synergic effect of their bioactive compounds, which are totally different

from those of chemical drugs, so just several components could not reflect the holistic nature of a certain herb medicine.[3] Therefore, much attention should be paid on the comprehensive information of TCMs to evaluate the effects of ecological factors. Chromatographic fingerprint technology might be an alternative method since it can offer integral characteristics of TCMs with a quantitative degree of reliability. Thus, chromatographic fingerprint analysis of herbal drugs presents a comprehensive qualitative approach for the purpose of species authentication, evaluation of quality, and ensuring the consistency and stability of herbal drugs and their related products.[4] Based on this suppose, chromatographic fingerprint has gained more and more attention and been internationally accepted as a feasible means for the quality control of TCMs. In 2004, Chinese manufacturers are also required by Chinese State Food and Drug Administration (SFDA) to standardize injections made from TCMs and their raw materials using chromatographic fingerprint.[5] So far, multiple patterns of chromatographic fingerprints have been developed such as multiple chromatographic fingerprints, bio- and meta-fingerprints.[6]

To establish an efficient chromatographic fingerprint, an appropriate analytical method is necessary. Among various chromatographic techniques, ultra-performance liquid chromatography (UPLC) is a preferred method and has earned great attention since it came out in 2004.[7,8] The special stationary phase with sub-2 μm particles endows this technique with some major advantages including increased peak capacity, improved resolution, shorter retention time, and

less solvent consumption.[9] The key to the successful application of the multi-compounds fingerprinting techniques is the selection of suitable compounds that would reflect the link with ecological factors and thus have discriminating potential for the ecological factors. Furthermore, professional software and statistical analysis techniques, *i.e.* principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), hierarchical cluster analysis (HCA) can assist to acquire meaningful information from a mass of data to obtain an effective fingerprint analysis. PCA could be used to identify outliers and innate clustering trends using SIMCA-P software (Umetrics, Umeå, Sweden). After that, PLS-DA was performed to generate a model which could discriminate each ecological level from others in the specific ecological factor and to identify compounds responsible for such discrimination.

Angelica sinensis (A. sinensis), named as female's ginseng, is predominantly renowned for treating gynecological diseases, constipation, cardiovascular disease and hepatic fibrosis.[10] To obtain high-quality of *A. sinensis*, some studies on its growing environment have been carried out.[11-14] However, most work focuses on using a few "markers" such as volatile flavor components, phenolic compounds, coumarins and phthalide, *etc*, to compare *A. sinensis* from different regions in China. However, these methods could not reflect its integrative properties. In this work, we developed a simple and reliable procedure for establishing a characteristic UPLC chromatographic fingerprint of *A. sinensis*. Furthermore, multivariate statistical analysis assisted the established chromatographic

fingerprint to evaluate the effects of main ecological factors on the quality of *A. sinensis*. As we know, it is the first time for chromatographic fingerprint technique combining with the proposed multivariable statistical analysis to use for the selection of the optimum ecological level for herbal medicines. And the obtained optimum ecological level would be scientifically used to instruct scientific growth of *A. sinensis*. Consequently, the proposed method was an alternative to evaluate the effects of ecological factors on the quality of TCMs and crops which was of great importance to establish a standardized plant cultivation method.

2. Materials and methods

2.1. Chemicals and materials

Ferulic acid was purchased from the National Institutes for Food and Drug Control (Beijing, China). Four ecological factors such as temperature, moisture, sunlight and altitude were set different levels, respectively. 109 batches of *A. sinensis* grew up under the special ecological level and then were collected. All of them were identified by Dr Huan-Yang Qi of our Laboratory. Chromatographic grade methanol was obtained from Merck Co. (Darmstadt, Germany). Other chemicals of analytical grade were provided by Tianjin Chemical Reagent Co. (Tianjin, China). Deionized water was prepared by using OKP purification system (Model: Exceed-AC-16, Shanghai Laikie Instrument Co. Ltd., China) and then used to prepare mobile phase and sample solution.

2.2. Apparatus and chromatography

UPLC was performed using a Waters Acquity Ultra Performance LC^{TM} system (Milford, MA, USA) equipped with a binary solvent delivery pump, an autosampler, a thermostated column compartment, a photodiode array detector and a Masslynx 4.1 workstation. Chromatographic separation of the analytes was performed on a Waters Acquity BEH C_{18} column (50 × 2.1 mm i.d., 1.7 μm particle). Chromatographic separation was achieved through a 25 min gradient delivery of a mixture of water including 1.0% aqueous acetic acid (A) and methanol (B) at a flow rate of 0.3 mL min^{-1}. The gradient schedule was: (a) 0-5 min, B 5-30%; (b) 5-6.5 min, B 30-35%; (c) 6.5-8.5 min, B 35-50%; (d) 8.5-14 min, B 50-80%; (e) 14-15 min, B 80-100%; (f) 15-20 min, B 100%; (g) 20-20.5 min, B 100-5%; (h) 20-25 min, B 5%. The absorption spectra of the compounds were recorded in the range of 190-400 nm, and the detection wavelength was set at 270 nm. The temperature of the column and sample manager room was maintained at 30^{o}C and 20^{o}C, respectively. The injection volume was 1 μL.

2.3. Program of field experiments

The experimental field with a total area of about 500 m^2 is located in Baitupo village, Min county, Gansu province, China. Every ecological level set three test plots and each plot had an area of 12 m^2 and an interval of 0.5 m. The seedlings of *A. sinensis* were purchased from a local farmer, planted at the end of March and grew under controlled growth condition. Specifically, the temperature level was controlled by using different cultivation modes including greenhouse cultivation, plastic film mulching cultivation and outdoor cultivation. In the mulching

cultivation, the ridge with the width of 0.8 m was prepared and covered by black plastic film. After that, hole was made in the plastic film for seedlings to be planted in the soil. As for greenhouse cultivation, seedlings were first planted as the above plastic film mulching technique and then grew in greenhouse before sprouting. We used sunshade net to design four light intensities, namely 100%, 75%, 50% and 25% of sunlight. The sunlight should be controlled before seedlings sprouting. In the moisture control experiment, five different treatments, namely no-irrigation, irrigated with 1 L m^{-2}, 2 L m^{-2}, 5 L m^{-2}, 10 L m^{-2}, respectively, were performed. The irrigation started on the tenth day after the seedlings sprouting. The altitude level ranged from 2300 to 3100 m with the interval of 100 m. The samples were collected at the end of October and their roots were air dried.

2.4. Preparation of standard solution

The stock solution of ferulic acid (0.1004 mg mL^{-1}) was prepared in 50% methanol and stored at 4^{o}C before use. Working solution was freshly prepared by appropriate dilution of the stock solution with 50% methanol to adjust to UPLC analysis.

2.5. Preparation of samples solution of *A. sinensis*

5.00 g ground power of *A. sinensis* sample was accurately weighted and introduced into a conical flask with stopper. The sample was extracted with 50 mL of 50% methanol in an ultrasonator for 45 min. After making the reduced weight up, the extract was filtered through a filter paper and a 0.22 μm filter membrane to be ready for analysis.

The effect of extraction method, solvent, time and sample quantity were studied by adopting a mono-factorial experimental design. The evaluation parameter involved two extraction methods (reflux and ultrasonication), twelve extraction solvents (methanol, 30, 40, 50, 60 and 80% methanol; 30, 40, 50, 60, 80 and 95% ethanol), four levels of sample quantity (2.5, 5, 7.5 and 10 g) and three extraction time (30, 45 and 60 min). The other extraction conditions were the same as for the above procedure.

2.6. Data analysis

2.6.1 Establishment of the reference chromatographic fingerprint

Data analysis was carried out by professional software named Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine, which was recommended by SFDA. This software was used for evaluating similarities of different chromatograms by calculating the correlative coefficient and/or cosine value of vectorial angel.[15-17] When correlative coefficients among the collected samples in the same group were higher than 0.9, it was considered that the samples were highly related and the group had high repeatability. The information of the matched peaks could be obtained from the similarity evaluation software and the common peaks would be chosen out according to the matched information. The common peaks should be able to exist in 80% of samples. Meanwhile, they should account for 90% of the total peak areas and could be separated from the adjacent peaks in the chromatograms well.[18]

2.6.2 Principal component analysis

As an unsupervised method, PCA is a method used to reduce data dimensionality and provide an overview of class separation, clustering and outliers.[19] Generally, PCA compresses the original data, and a new set of orthogonal variables, called principal components (PCs), are obtained. These PCs are linear combinations of the original variables.[20] PCA provides a loading plot which could reveal the contribution of the original variables. In this study, the PCA was performed with SIMCA-P software. The peak of ferulic acid was chosen as the reference peak since ferulic acid is the main active component in *A. sinensis* and has also been selected as the evaluation marker for *A. sinensis* in the *Chinese Pharmacopoeia* (2010 edition). The relative peak areas (RPAs) of the common peaks to the reference peak in every sample were calculated and input into the dataset of SIMCA-P software. In the dataset, the first column was the serial number of samples and the row contained the RPAs of common peaks of the corresponding sample. The data was mean-centered and unit variance-scaled. Then, the PCs were extracted and the two PCs of greatest contribution were used to build the loading plot to reflect the contribution of each common peak.[21]

2.6.3 Partial least squares discriminant analysis

PLS-DA in SIMCA-P software was used to build a model which would discriminate one ecological level from others and screen out ecological factor-related compounds. The predictabilities and reliabilities of the models were evaluated by using 7-fold cross-validation test. The PLS-DA model parameters, R^2X and R^2Y, showed the explanative ability for the X matrix and the stability of

the constructed model, respectively.[21] The closer the values were to 1, the better the meaning expressed. In addition, the success of the classification was recalculated by comparison with the known. The Variable Importance in the Projection (VIP) value was used to evaluate the variable contribution. Usually, the compounds with VIP values larger than 1.0 were considered responsible for the model classification.

2.6.4 Significance tests

Significant tests could check whether the differences of RPAs of common peaks among different ecological levels were statistically meaningful. In this paper, significant tests including One-Way ANOVA (One-way analysis of variance) and K-Independent Test in Nonparametric tests were performed with screened peaks with VIP>1 to examine whether the differences of their RPAs were significant among different ecological levels at $p = 0.05$ level (SPSS software, Version 16.0, SPSS Inc., Chicago, IL, USA). For peaks which RPAs meet the requirement of the normality and homogeneity of variance, One-Way ANOVA was used while K-Independent Test was used to analyze the remaining peaks. Finally, the optimum ecological level of every factor was chosen out by comparing the peak areas of the peaks with significant differences. Fig. 1 shows schematic illustration of the proposed strategy for screening the optimal ecological level for *A. sinensis*.

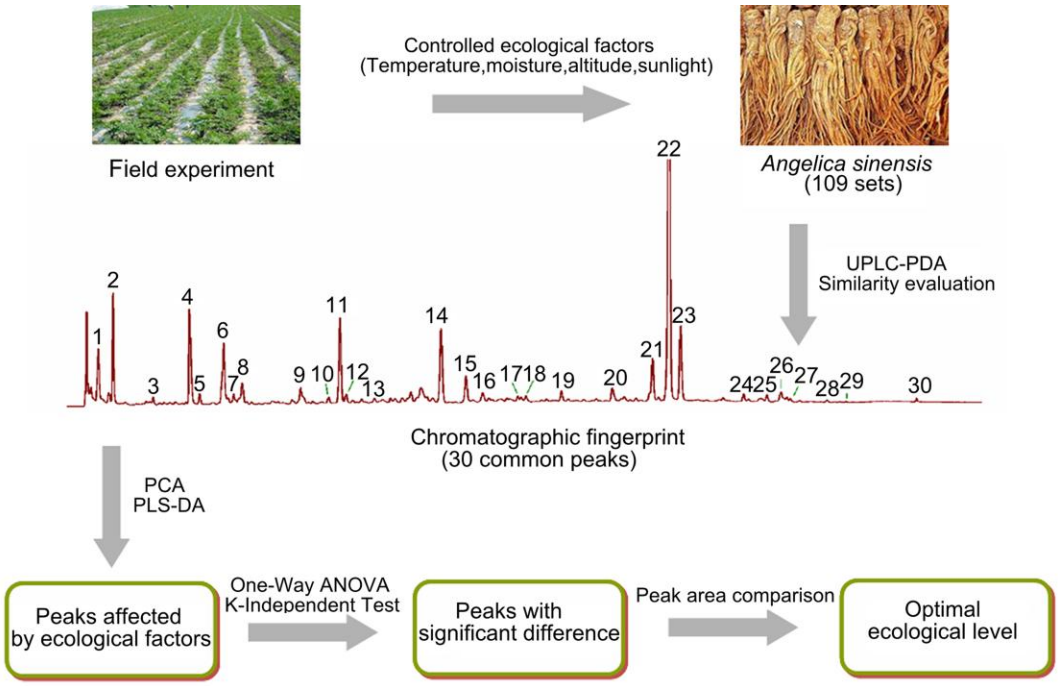


Fig.1 Schematic illustration of the proposed strategy for screening the optimal ecological level of *Angelica sinensis*.

3. Results and discussion

3.1. Optimization of chromatographic conditions

The UV spectra between 190 and 400 nm were recorded using a photodiode array detector. As shown in Fig. 2, 270 nm was selected as detection wavelength in the subsequent study based on the baseline stability, the number and resolution of characteristic peaks.

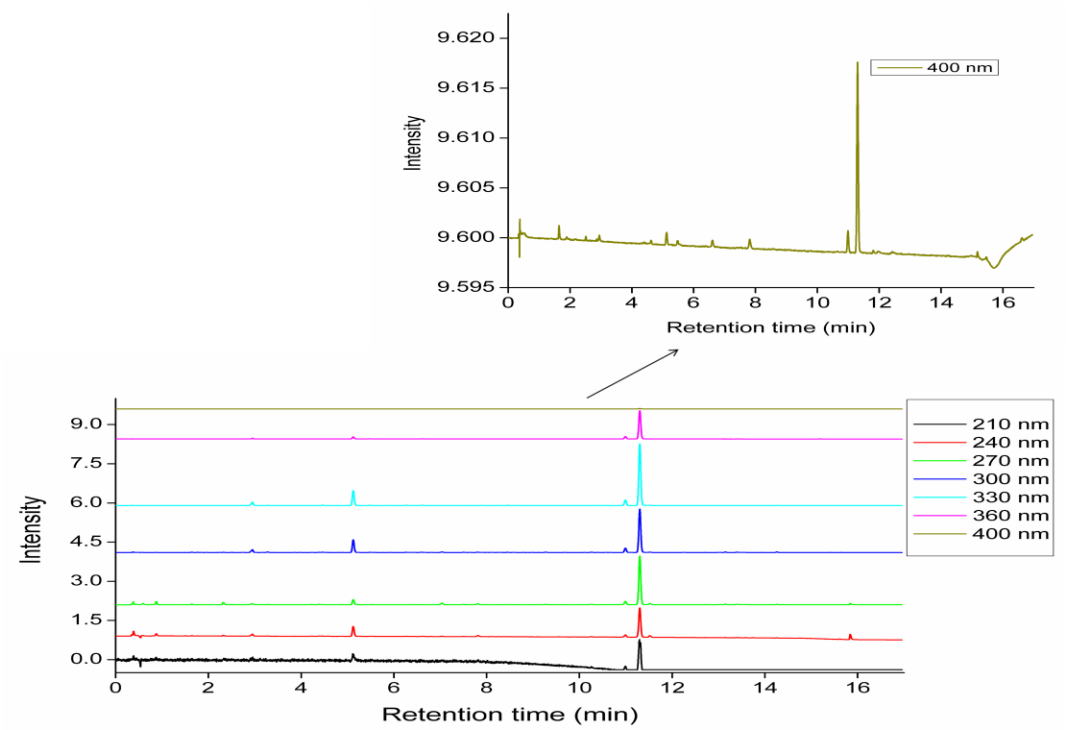


Fig.2 Chromatograms of extraction solution of an *Angelica sinensis* sample at different detection wavelengths. Mobile phase: water containing 1.0% aqueous acetic acid (A) and methanol (B); flow rate: 0.3 mL min^{-1}; gradient: (a) 0-5 min, B 5-30%; (b) 5-6.5 min, B 30-35%; (c) 6.5-8.5 min, B 35-50%; (d) 8.5-14 min, B 50-80%; (e) 14-15 min, B 80-100%; (f) 15-20 min, B 100%; (g) 20-20.5 min, B 100-5%; (h) 20-25 min, B 5%; column temperature: 30°C; the temperature of sample manager room: 20°C; the injection volume: 1 μL.

The selection of the mobile phase was guided by the requirement for obtaining chromatograms with stable baseline, good resolution of adjacent peaks, and as many characteristic peaks as possible within a relatively short analysis time, so different mobile phases consisting of acetonitrile or methanol and water with some modifiers including formic acid, acetic acid and phosphoric acid were investigated. The results showed that methanol and water containing 1.0% acetic

acid as mobile phase could provide efficient separation.

In the study, the parameters regarding to gradient elution and the initial mobile phase composition were taken into consideration. After several types of gradient of different duration were tried, the optimum gradient in section 2.2 was selected.

The effect of column temperature in the range of 25-45^{o}C on separation was also investigated. The increase in temperature often produces a compromising effect between the column efficiency and the peak-to-peak separation. According to the rate equation of chromatography analysis (van Deemter equation), when temperature increases in a certain range, diffusion coefficients of analytes in mobile phase are enhanced which would consequently improve the column efficiency. From the point of thermodynamics, the increase in temperature would decrease the retention factor which is defined as the amount ratio of the analytes in stationary phase and mobile phase, leading to the lessening of the peak-to-peak separation. When a low column temperature was employed, relatively long analysis time was observed due to the slow diffusion coefficients of analytes in mobile phase. As the result shown, the optimum column temperature was 30^{o}C. In addition, other important chromatographic conditions like flow rate, injection volume were also optimized. After a series of tests, excellent separation was achieved within 25 min in a single run.

3.2. Optimization of extraction conditions

In order to acquire a stable and reproducible fingerprint, a sufficient extraction of

the main compounds should be guaranteed, so extraction variables involved in the extraction procedure, for example, sample quantity, extraction method, solvent and time were optimized by comparing the sum peak numbers, peak areas and resolution. Two main extraction methods, reflux and ultrasonication were used separately to extract compounds form *A. sinensis* samples. Extraction efficiency for ultrasonication was equal to that for reflux. So, ultrasonificaton was selected owing to its simplicity, feasibility and less human error. In addition, the effect of extraction time was also evaluated. The results showed that 45 min was enough to extract the desired compounds completely. In the following step, methanol, ethanol, aqueous methanol and ethanol were evaluated as the extraction solvents. As the results shown in Table 1, 50% methanol had the highest extraction efficiency for most of the compounds. So, 50% methanol was chosen as the extraction solvent in the subsequent experiment. Sample quantity was also studied. Actually, increasing sample quantity would make the peak areas of main chromatographic peaks increase, while the increase would become unapparent when the sample quantity was more than 5.00 g. So, sample quantity was set at 5.00 g.

Table1. Effect of extraction solvent on the peak areas of common peaks in *Angelica sinensis*

peak	retention	methanol content						ethanol content					
no.	time (min)	30%	40%	50%	60%	80%	100%	30%	40%	50%	60%	80%	95%
1	0.581	3013	1861	1933	348	415	450	2217	1228	---	64	---	---
2	0.854	5440	3857	3910	300	279	472	4441	----	207	275	82	---
3	1.600	98	180	1093	1899	564	396	427	540	174	247	273	48
4	2.470	249	266	576	753	1565	509	386	349	385	---	169	62
5	3.094	155	155	181	94	126	221	93	82	51	58	---	---
6	3.254	757	717	636	359	576	554	484	391	284	290	422	281

7	4.342	1288	426	760	489	432	402	330	376	---	384	303	259
8	4.856	210	191	218	139	163	50	133	111	83	65	69	---
9	5.070	3260	4204	9076	8576	10794	5676	7110	7212	6576	6193	5968	3392
10	5.183	229	230	290	167	273	204	74	69	56	---	---	62
11	5.705	220	189	171	130	138	86	147	116	91	243	107	---
12	6.946	3110	3006	3574	956	905	4746	3983	3529	2656	2979	4274	3914
13	7.408	672	630	852	412	483	1065	734	579	358	451	986	837
14	7.716	216	527	2283	3736	4343	258	82	---	113	79	---	---
15	8.363	21	63	153	136	186	130	---	---	11	---	---	18
16	8.520	231	206	223	131	158	262	189	169	143	154	189	23
17	8.808	53	51	69	38	41	63	62	58	47	43	54	34
18	8.900	187	213	229	124	102	---	---	---	10	14	21	41
19	9.178	252	923	776	394	385	548	422	419	339	343	404	320
20	10.145	627	729	986	714	839	1012	591	665	669	727	954	83
21	10.863	1820	3028	6297	6786	8914	7808	2181	2956	3850	4422	6170	6526
22	11.168	20251	37917	94423	100738	126839	127218	29937	45694	61491	70291	92769	103672
23	11.383	142	522	2113	930	1218	4132	882	1466	2162	2504	3321	4031
24	12.551	28	---	---	---	---	---	---	---	---	---	---	---
25	12.984	80	194	1110	1033	1105	1497	104	586	1019	1057	1250	---
26	13.351	127	200	729	803	968	1250	---	---	43	754	909	984
27	13.415	12	7	116	---	75	---	---	30	76	48	41	---
28	14.098	74	133	611	468	450	674	61	371	568	526	578	87
29	14.252	74	133	632	464	510	666	61	371	566	530	578	---
30	15.747	1656	1658	1667	1697	1725	1678	1530	1528	1570	1582	1690	1815

Each value in the table is the mean value of three replicate, all RSDs were less than 5.0%.

3.3. Precision and stability

The precision of the instrument was evaluated by performing intra-day and inter-day assays by replicate injection of the same real sample solution. Intra-assay precision was measured for injections at 3h intervals during the same day, and the obtained relative standard deviation (RSD) values of relative retention time (RRT) and RPAs of the common peaks to the reference peak were in the range of 0.06-5.12% and 0.22-3.28%, respectively. Inter-assay precision was measured on three consecutive days with RSD values of RRT and RPAs in the range of 0.08-5.45% and 0.23 to 6.62%, respectively. The precision of the method was

evaluated by the injections of six different sample solution of *A. sinensis* which was prepared in parallel by the same sample preparation procedure. The RSD values of RPAs ranged between 4.46 and 5.01%. For the stability test, one of six sample solutions was analyzed at 0, 4, 8, 12, 24, 48 h, respectively, during storage at 20^{o}C. The sample solution can be regarded as stable within 48 h because the obtained RSD values of RPAs were all less than 4.76%.

3.4. Chromatographic fingerprint analysis

The developed method was applied to analyze 109 sets of samples which grew under different ecological factors. The obtained chromatograms were first input into the "Similarity Evaluation System for Chromatographic Fingerprint of TCM" (2004 A edition), and then the peak areas were normalized to eliminate injection error. The identical peaks in every chromatogram were matched after correction for the drift in retention time. Finally, the reference chromatogram based on median would be output.[18] The chromatographic fingerprint based on UPLC-PDA was shown in Fig. 3. 30 peaks were determined to be the common characteristic peaks. Peak 11 of ferulic acid was selected as the reference peak. Each sample was analyzed by the proposed UPLC five times under the optimum condition and the average of the RRT and RPAs with RSDs less than 5.0% were used for analysis.

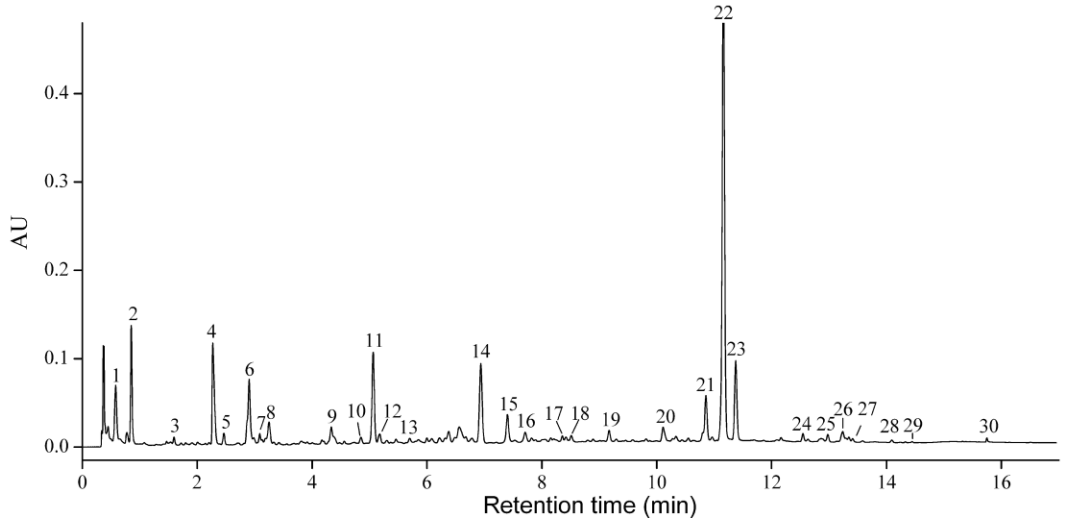


Fig.3 Chromatographic fingerprint of *Angelica sinensis* based on UPLC-PDA. The marked peaks were the common peaks. The peak marked 11 is ferulic acid.

3.5. Effect of temperature factor

In the present study, each level of temperature factor represented for one specific group. The similarity of parallel samples in every group was analyzed by similarity evaluation system and the result was shown in Table 2. The RPAs of the common peaks were input into the SIMCA-P software which was used to set up the optimum model. The PCA analysis auto-fitted three PCs which could interpret 70.0% of the total common peaks information, consequently, the three PCs concentrated the multidimensional information into 3-D dataset to sort the samples. The separation of the three groups could be achieved with the PLS-DA model parameters R^2X=0.755, R^2Y=0.896. On the basis of the VIP threshold (>1), a total of 20 peaks (Table 2) were selected, which was also proved to have important influence on the discrimination of the samples in the loading plots (Fig. 4). Subsequently, the RPAs of these components in every sample were input into the SPSS software to validate whether they had significant difference among the three

groups. The one-way ANOVA test found the peaks 5, 6, 16, 24, 25, 26, 28, 30 had significant difference in the three different groups while the K-Independent test in Nonparametric tests found the peaks 8, 10, 13, 19, 21, 22, 23 had significant difference. It meant that the temperature influenced them greatly, so their contents were significantly different when planted under the three different temperatures. After comparing the peak areas of these peaks in the three groups, plastic film mulching cultivation was found to be the best, which was in accordance with the previous research.[11]

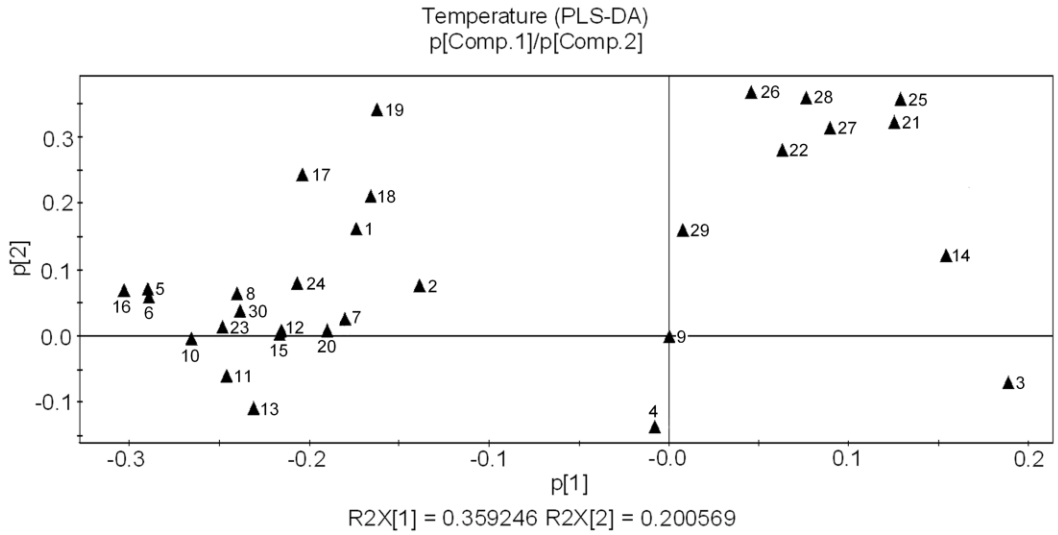


Fig.4 Loadings plot of variables corresponding to samples in the three groups of the temperature control. Every point in the plot was marked with the serial number of peak in the chromatogram.

Table 2. Information of similarity analysis and difference evaluation

Ecological factor	special level	sample numbers	similarity analysis	No. peak with VIP>1	No. peak with significant difference among groups	optimum level
temperature	greenhouse cultivation	9	≥0.978	5, 6, 8, 10, 11, 12, 13, 14, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30	5, 6, 8, 10, 13, 16, 19, 21, 22, 23, 24, 25, 26, 28, 30	plastic film mulching cultivation
	plastic film mulching cultivation	6	≥0.953			
	outdoor cultivation	6	≥0.942			

altitude	2300 m	3	=1.000	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 16, 17, 18, 19, 20, 22, 23, 30	2700 m
	2400 m	3	=1.000			
	2500 m	3	≥0.999			
	2600 m	3	≥0.991			
	2700 m	3	≥0.992			
	2800 m	3	≥0.998			
	2900 m	3	≥0.998			
	3000 m	3	≥0.999			
	3100 m	3	≥0.999			
moisture	no water	9	≥0.966	1, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 22, 23, 25, 27, 28, 30	1, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 23, 25, 27, 28, 30	irrigated with 2 L m^{-2}
	irrigated with 1 L m^{-2}	9	≥0.962			
	irrigated with 2 L m^{-2}	9	≥0.958			
	irrigated with 5 L m^{-2}	9	≥0.961			
	irrigated with 10L m^{-2}	9	≥0.981			
sunlight	full sunlight	4	≥0.989	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 29, 30	1, 6, 8, 15, 16, 18	50% sunlight
	75% sunlight	4	≥0.989			
	50% sunlight	4	≥0.990			
	25% sunlight	4	≥0.993			

1 The correlation coefficient of the median was considered as the index of similarity.

3.6 Effect of altitude factor

The altitude factor was studied from 2300 to 3100 m. Satisfactory similarities were obtained in every group with correlation efficiencies higher than 0.991 (shown in Table 2). Then the RPAs of common peaks were analysis by the SIMCA-P software. The PCA analysis obtained three principal components which could interpret 79.5% of variable information while the PLS-DA (PCs, 5; R^2X=0.872; R^2Y=0.492) could discriminate each group from the others. The important compounds with VIP>1 which were also proved to be important in the loading plots (Fig. 5) were screened out and then used to validate whether they had significant differences among groups. The results showed that 19 compounds had significant contents difference in the nine groups and the optimum altitude was 2700 m. Wang and Qiu group also found that increasing altitude in suitable range

would be advantageous to the conversion and accumulation of active compounds in *A. sinensis* while both height and boling rate of *A. sinensis* would decrease significantly in high altitude.[11,12]

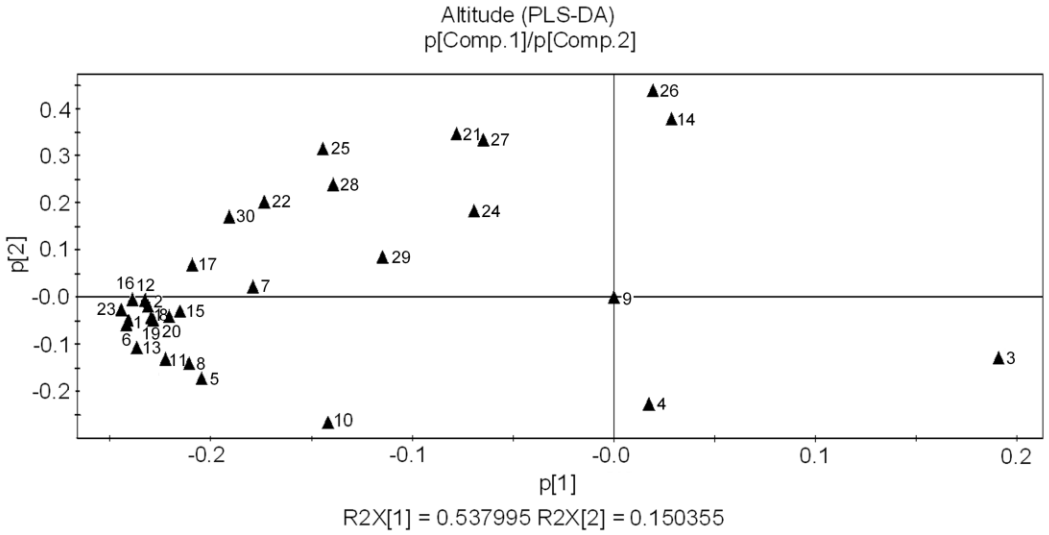


Fig.5 Loadings plot of variables corresponding to samples in the nine groups of the altitude control. Every point in the plot was marked with the serial number of peak in the chromatogram.

3.7. Effect of moisture factor

Water is one of the most important factors affecting crop growth, yield and quality, and too high or too low would be adverse condition.[13] In the present study, we designed five levels about the moisture by controlling the irrigation volume. The RPAs of the common peaks were used to perform PCA analysis (PCs, 3; R^2X=0.717) and PLS-DA analysis (PCs, 3; R^2X=0.681; R^2Y=0.603). VIP compounds were found and validated in the loading plot (Fig. 6), among which 18 peaks were confirmed to have significant differences among the five groups. Finally, the optimum moisture condition (2 L m^{-2}) was chosen by comparing the

peak areas of the 18 peaks in every group. The corresponding results were shown in Table 2.

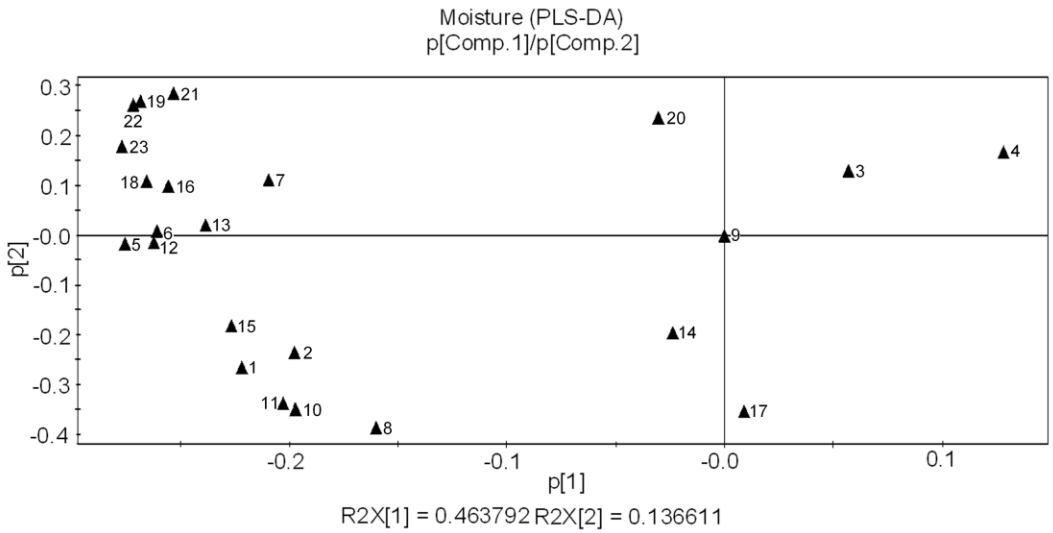


Fig.6 Loadings plot of variables corresponding to samples in the five groups of the moisture control. Every point in the plot was marked with the serial number of peak in the chromatogram.

3.8. Effect of sunlight factor

In this experiment, four levels of light intensities were designed to evaluate the effects of sunlight. The correlation coefficients of the four groups were higher than 0.989, which proved they were highly related. Subsequently, the VIP peaks were screened out by SIMCA-P (PCA: PCs, 3; R^2X=75.9; PLS-DA: PCs, 7; R^2X=0.961; R^2Y=0.947, loading plot, Fig. 7) and the ones with significant differences analyzed by the SPSS software were used to choose the optimum value. As the results in Table 2 shown, 50% of sunlight would be the best for the growth of *A. sinensis* since suitable shade would decrease the likelihood of photoinhibition of photo-synthesis and water evaporation.[13]

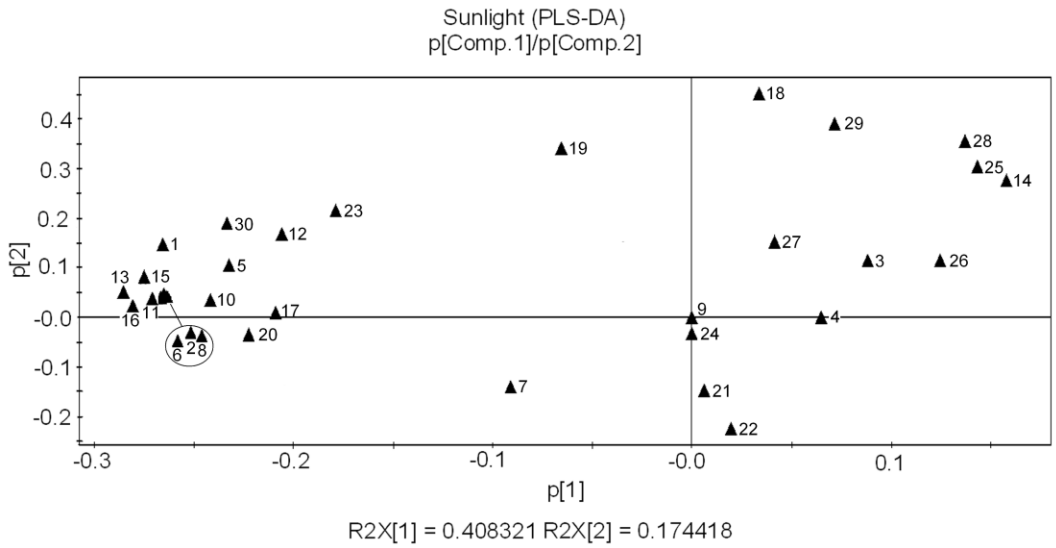


Fig.7 Loadings plot of variables corresponding to samples in the four groups of the sunlight control. Every point in the plot was marked with the serial number of peak in the chromatogram.

4 Conclusion

In conclusion, the quality of herbal medicine, mainly being embodied in its chemical constituents, is affected by its cultivation environment. In the present study, a new method based on chromatographic fingerprint and statistical analysis to evaluate the effects of ecological factors was developed. Unlike the traditional method that a single or a few specific compounds were used as evaluation markers to choose the optimum ecological level, this new method took the advantages of chromatographic fingerprint which would be more sensitive and comprehensive since it offered the integral information of TCMs. By virtue of powerful software such as similarity evaluation software, SIMCA-P and SPSS, the common peaks were obtained and then the peaks having great important effect on the classification were subsequently chosen out from the common peaks. Finally, the

optimum ecological level was obtained according to the peak areas of the peaks having significant difference in different groups. The obtained optimum level of ecological factors may be used to guide the cultivation of *A. sinensis*. Moreover, the developed method based on chromatographic fingerprint and statistic analysis may serve as a new mode to evaluate the effects of ecological factors on the quality or yield of TCMs and crops.

Acknowledgments

This work was supported by the National Key Technology Research and Development Program of China (No. 2011BAI05B02) and the National Natural Science Foundation of China (No. 21105106)

Notes and references

[a] Key Laboratory of Chemistry of Northwestern Plant Resources and Key Laboratory for Natural Medicine of Gansu Province, Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, Lanzhou 730000, People's Republic of China

[b] Gansu College of Traditional Chinese Medicine, Lanzhou 730000, P. R. China

[c] University of Chinese Academy of Sciences, Beijing 100039, P. R. China

*Correspondence: Juan Chen, Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, Lanzhou 730000, P. R. China, E-mail: chenjuan@licp.cas.cn; Tel.: 86-931-4968208; fax: 86-931-4968094; Ying-Dong Li,

Gansu College of Traditional Chinese Medicine, Lanzhou 730000, P. R. China, E-mail: lydj412@163.com.

1. P. Li, L. W. Qi, E. H. Liu, J. L. Zhou, X. D. Wen, *Trends Anal. Chem.*, 2008, **27,** 66-77.
2. T. H. Xue, R. Roy, *Science*, 2003, **300,** 740-741.
3. J. Han, M. Ye, H. Guo, M. Yang, B. R. Wang, D. A. Guo, *J. Pharm. Biomed. Anal.*, 2007, **44,** 430-438.
4. V. R. Naumovski, W. Tongkao, *Wolrd Sci. Technol.*, 2012, **12,** 99-106.
5. S. L. Li, Q. B. Han, C. F. Qiao, J. Z. Song, C. L. Cheng, H. X. Xu, *Chin. Med.*, 2008, **3,** 7.
6. Y. L. Hua, Y. S. Guo, S. H. He, J. F. Fan, Y. M. Wei, *J. Anim. Vet. Adv.*, 2012, **11,** 191-199.
7. P. R. Li, L. L. Dong, J. X. Huang, *Anal. Chim. Acta*, 2005, **546,** 167-173.
8. S. A. C. Wren, P. Tchelitcheff, *J. Chromatogr. A*, 2009, **1119,** 140-146.
9. K. Yu, D. Little, R. Plumb, B. Smith, *Rapid Commun. Mass Spectrom.*, 2006, **20,** 544-552.
10. X. D. Wen, L. W. Qi, J. Chen, Y. Song, L. Yi, X. W. Yang, P. Li, *J. Chromatogr. B*, 2007, **852,** 598-604.
11. H. Z. Wang, L. Jin, E. H. Zhang, *Chin. Trad. Herbal Drugs*, 2013, **44,** 219-223.
12. D. Y. Qiu, H. M. Lin, Y. Chen, Y. D. Li, F. X. Guo, *Acta Agrestia Sinica*, 2012, **18,** 838-843.

13. B. Liu, B. Y. Rao, C. Qian, Z. J. Wu, *Modern Agric. Technol.*, 2013, **2**, 11-13.

14. C. Qian, B. Y. Rao, S. Q. Luo, Z. J. Wu, *J. Anhui Agri. Sci.*, 2012, **40,** 5613-5617.

15. L. X. Wang, H. B. Xiao, X. M. Liang, K. S. Bi, *Acta Pharm. Sin.*, 2002, **37,** 713-717.

16. F. Gong, Y. Z. Liang, P. S. Xie, F. T. Chau, *J. Chromatogr. A*, 2003, **1002,** 25-40.

17. Y. Jin, T. Liang, Q. Fu, Y. S. Xiao, J. T. Feng, Y. X. Ke, X. M. Liang, *J. Chromatogr. A*, 2009, **1216,** 2136-2141.

18. D. Z. Yang, Y. Q. An, X. L. Jiang, D. Q. Tang, Y. Y. Gao, H. T. Zhao, X. W. Wu, *Talanta*, 2011, **85,** 885-890.

19. M. Mal, P. K. Koh, P. Y. Cheah, E. C. Y. Chan, *Anal. Bioanal. Chem.*, 2012, **403,** 483-493.

20. Y. N. Ni, B. H. Li, S. Kolot, *Anal. Methods*, 2012, **4,** 4326-4333.

21. H. W. Wang, *The extracting method of components in table-Principal Component Analysis in Partial Least-Square Regression-Model and Application*, Springer, 1999.