

Quality comparison of *Flos Lonicerae Japonicae* by several dissimilarity methods

Guoxiang Sun¹, Wenjing Song², Xiangyu Deng³, Zhongbo Liu¹

¹Shenyang Pharmaceutical University, Shenyang 110016, China; ²School of Pharmacy, Hainan Medical University, Haikou 571199, P.R. China;

³Novartis AG, Suzhou, P.R. China

Corresponding to: Guoxiang Sun, Ph.D, Professor and Vice Dean of Pharmacy College. Shenyang Pharmaceutical University, 103 Wenhua Road, Shenhe District, Shenyang 110016, P.R. China. Email: gxswmwys@163.com or gtcnfp@163.com.

Abstract: Computing similarity is extremely important in modern scientific research such as computational biology and data mining. In this paper, we put forward the concept of quantitative dissimilarity (QDS) and define a series of formulations to measure how similar or dissimilar between two objects are in quality and content for their n -dimensional properties, which is based on a simple mathematical model. Through establishing the HPLC fingerprints (HPLC-FPs) of *Flos Lonicerae Japonicae* (FLJ) to obtain n -dimensional vectors, in which each peak area serves as an element, the authentic quality control of FLJ were successfully implemented by several QDSs. The presented dissimilarity methods can accurately and quantitatively assess all objects differences in n -dimensional properties.

Key Words: Quality comparison; *Flos Lonicerae Japonicae*; quantitative dissimilarity



Submitted Dec 15, 2012. Accepted for publication Jan 18, 2013.

doi: 10.3978/j.issn.2305-5839.2013.01.02

Scan to your mobile device or view this article at: <http://www.atmjournals.org/article/view/1419/1983>

Introduction

The similarity phenomenon existing universally in nature, science, society and other fields has been playing a great role in evaluating the difference among things. Aristotle, the Greek philosopher, thought “In philosophy, it is usually correct to think over the similarities, although they are far away from each other”. Planck, the German physicist, said “On the surface the images of the nature vary greatly, but the similar simple principle often exists in incoherent fields”. When humans face the world, looking for similarities is their instincts, and they always understand the world through the similarities between the objects. In 1869, Mendeleev, the Russian chemist, had successfully discovered periodic law of elements by studying on the similarities between atomic weights and chemical properties of the elements; and computing similarity plays a more significant role in modern scientific research, such as the diversity and distribution proportion of biological genes or proteins, chemical composition and content distribution of

plant medicines, popular ranking of universities.

In ancient study on similarity, people depended mainly on experience and expertise. However, with the development of science and technology, the requirement of the research becomes higher and higher, so that it is hard to precisely measure similarity depending only on experience and expertise. Therefore mathematical tools gradually were introduced into research to give rise to the multivariate data analysis. Such by means of vectors in n -dimension space to study the similarity degrees between objects is called similarity science that mainly contains qualitative similarity and quantitative similarity. Qualitative similarity merely lays stress on the “quality” of the objects, and quantitative similarity investigates their similarity degrees in respect of “quantity”, so they are unified and mutual complementarity. Qualitative similarity adequate is the premise to assessing the quantitative similarity, for quantitative similarity is blind and meaningless without an adequately qualitative similarity; also quantitative similarity would make qualitative similarity more accurate and scientific in measuring for the

multivariable objects. Wherever the universal law followed by the similarity may possess similar form. The approach to such form and its law is also the final goal that quantitative dissimilarity (QDS) efforts to achieve.

The resemblance magnitude can be measured by both similarity and QDS, in which similarity describe how larger of resemblance degree between two objects, and the more alike objects the larger similarity. Usually, similarity is positive and typically within 0 to 1 to indicate various degrees of resemblance, where 0 means that two objects are not similar at all and 1 reflects maximal similarity. Dissimilarity represents how big is the difference degree, in which the opposite to similarity, the more similar the two objects the lower the QDS. Dissimilarity may varies within [0, 1] or from $-\infty$ to ∞ . The development of QDS has principally involved three stages: similarity research for binary variable, similarity research for multivariate vectors and similarity research based on fuzzy sets. The variable has two attributes (0 or 1) is called binary variable, whose similarity can be measured by simple matching coefficient (SMC) (1,2) and Jaccard coefficient (3,4). This stage rests on qualitative similarity and still has been applied in information retrieval now. Until 1930s, the objects had turned into multivariate vectors, whose similarity measurements mainly included all kinds of coefficients, such as Tanimoto coefficient (5), cosine coefficient, Pearson's correlation coefficient (6), Dice coefficient (7), Bray-Curtis coefficient (8), Overlap measure and so on. Nei (9,10) has used Nei coefficient to characterize the genetic similarity between individuals when he researched the diversity of genes. Also these methods don't have pretty quantitative properties. Dissimilarity is generally measured by distances, for example the popular Euclidean distance served as a tool to determine the distance between two vectors, in addition Minkowski distance (11,12), Manhattan distance, Chebyshev distance, Mahalanobis distance, Aitchison distance (13,14), Hamming distance, Bhattacharya's distance (14), Hausdorff distance (15) and Dimension Root similarity (16), Gowda and Diday's QDS measure (17), Ichino and Yaguchi's QDS measure (12), Canberra QDS (18) etc. Although they can quantify in some extent, the visualized and comprehensive percent results of quantitative evaluation cannot be given directly. Since 1965 Zadeh (19) proposed fuzzy set theory, fuzzy mathematics had been widely used and rapid developed, fuzzy similarity became an important tool for comparing two groups or two elements (20-24). Its main idea is establishing fuzzy data according to fuzzy set theory and then evaluating their similarities using various coefficients or distances as mentioned above. Due to late start, this method is far from

perfect up to now and is used broadly in image processing and recognition.

This paper presented several dissimilarities through establishing a mathematical model to give the qualitative dissimilarities and QDSs, in which the quality comparison based on HPLC-FPs of FLJs was successfully performed. These methods were also suitable for evaluations of the diversities and distribution proportions of biological genes and proteins, evaluations of chemical composition and contents distributions of herbal medicines and computing the popular ranking of universities.

The dual qualitative similarity (DQLS)

If sample vector (SV) is denoted as $\overline{OX} = \overline{X} = (x_1, x_2, \dots, x_n)$ and the referential vector (RV) as $\overline{OY} = \overline{Y} = (y_1, y_2, \dots, y_n)$, the cosine of the angle (θ) between them is defined as qualitative similarity (S_F), as shown in Eq.[1], which clearly reveals the resemble degree between the coordinates in distribution proportion of numerical magnitude. Unfortunately, it does not possess any quantitative properties, for instance, the S_F of $\overline{a} = [1, 2, 3, 4, 5]$ and $\overline{b} = [10, 20, 30, 40, 50]$ is 1.0, whereas the content of each element in \overline{b} is 10 times as that in \overline{a} . Furthermore, x_i and y_i contribute differently to S_F , respectively, which suffers from the drawbacks that the larger elements seriously mask the smaller ones. If RV and SV are denoted as $\overline{P}_0 = [1, 1, \dots, 1]$ and $\overline{P}_s = (\frac{x_1}{y_1}, \frac{x_2}{y_2}, \dots, \frac{x_n}{y_n}) = (r_1, r_2, \dots, r_n)$, respectively, then qualitative ratio similarity ($S_{F'}$) is defined by calculating the cosine between \overline{P}_0 and \overline{P}_s , as listed in Eq.[2]. Because of the tendency divided by y_i , x_i shows equal weights to $S_{F'}$. However, $S_{F'}$ seriously ignores the contributions of the larger elements to system and overemphasizes those of the smaller elements. Therefore S_F and $S_{F'}$ are combined to qualitatively evaluate the similarity between SV and RV and called **DQLS**, in which they can simultaneously monitor the contributions of both larger elements and smaller elements to system with the characteristics of accuracy and not laying particular stress on any elements. The resemblance of \overline{X} and \overline{Y} is judged qualified when S_F and $S_{F'}$ no less than 0.9 are servered as the necessary condition to quantitatively evaluate them. Otherwise, it would be meaningless for quantitative similarity evaluations, because the chemicals number attribution and congener content degree does not meet the requirements (25).

$$S_F = \cos\theta = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad [1]$$

$$S_F' = \cos\theta' = \frac{\sum_{i=1}^n r_i}{\sqrt{n \sum_{i=1}^n r_i^2}} = \frac{\sum_{i=1}^n \frac{x_i}{y_i}}{\sqrt{n \sum_{i=1}^n (\frac{x_i}{y_i})^2}} \quad [2]$$

The dual quantitative similarity (DQTS)

The ratio of module length of \bar{X} to module length of \bar{Y} (W) well reflects the relationship of macroscopic content between SV and RV, see Eq.[3], where W is also strongly influenced by the larger elements and subjects to a problem of the smaller elements being masked by the larger ones, either. In addition, θ can also affect W . Apparent content similarity (R) is defined as the sum of the element values of SV divided by that of RV, and its formulation is shown in Eq.[4]. Because of all the elements being summed, the content property of a sample to reference reflected by R is accurate. Nevertheless, in SV, the larger elements and smaller elements may crossly compensate to bring about errors. W and R constitutes the first level dual quantitative similarities (1L-DQTS), which can not only reduce the masking effects of the larger elements in some extent but also equally treat the contributions of the smaller elements to system. Consequently, they are capable of monitoring sensitively the contributions of all elements to the 1L-DQTS (26).

$$W = \frac{|\bar{X}|}{|\bar{Y}|} \times 100\% = \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{\sum_{i=1}^n y_i^2}} \times 100\% \quad [3]$$

$$R = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \times 100\% \quad [4]$$

Defined the percentage of projection length of \bar{X} to \bar{Y} relative to the module length of \bar{Y} as projection content similarity (C), as shown in Eq.[5]. Like W , the larger elements dramatically affect it for they masking the smaller elements. But C is more accurate than W , since it is the ratio of the length of SV projection on RV to that of RV. When θ trending to zero ($C=WS_F$) $C=W$. Correct R with S_F to yield quantitative similarity (P), see Eq.[6], where P successfully eliminates the cross compensation of R . C and P are used together to constitute the second level dual quantitative similarities (2L-DQTS). They also treats all elements as the same weights and correct their distribution proportion, so are of the most significant and obviously being the correction values of the 1L-DQTS by S_F . The better qualified conditions of C and P are within 90-110%

(or broaden to 70-130% as fine) and the discrepancy between them is no more than 15%, so then the content of each element of \bar{X} is equivalent to 90-110% that of \bar{Y} (or 70-130%) (25).

$$C = \frac{|\bar{X}| S_F}{|\bar{Y}|} \times 100\% = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^2} \times 100\% = WS_F \quad [5]$$

$$P = R S_F \quad [6]$$

The QDSs from the tetra-pyramid model

Proposed \bar{X} and \bar{Y} intersect at O , the cosine (γ_X) of the angle (ψ) between $\bar{X} = (x_1, x_2, \dots, x_n)$ and $\bar{a} = [1, 1, \dots, 1]$, the cosine (γ_Y) of angle (ρ) between $\bar{Y} = (y_1, y_2, \dots, y_n)$ and $\bar{a} = [1, 1, \dots, 1]$, are calculated to well represent the similar degrees between \bar{a} and \bar{X} or \bar{Y} , respectively, thus they are called homogenizing coefficient. The ratio of DQTSs within the 1L-DQTS or 2L-DQTS is just the ratio of γ_X to γ_Y , see Eq.[7]. We regulate the differences between R and W , P and C not exceeding 15% in order to ensure the distinction of homogenizing coefficient between \bar{X} and \bar{Y} less than 15%. In this way, numerical homogeneity degree of any elements of the two vectors is confined to resemblance. Eq.[8] can be derived from Eq.[5] and Eq.[6], which indicates that in general the qualitative similarity is the ratio of two quantitative similarities and merely be used to classify by attribution rather than quantity.

$$\frac{\gamma_X}{\gamma_Y} = \frac{\cos\psi}{\cos\rho} = \frac{\sum_{i=1}^n x_i}{\sqrt{n \sum_{i=1}^n x_i^2}} \left(\frac{\sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n y_i^2}} \right)^{-1} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \times \frac{\sqrt{\sum_{i=1}^n y_i^2}}{\sqrt{\sum_{i=1}^n x_i^2}} = \frac{R}{W} = \frac{P}{C} \quad [7]$$

$$S_F = \frac{C}{W} = \frac{P}{R} \quad [8]$$

By the respective characteristics, the two level DQTSs compose a tetra-pyramid $OXYVZ$, see *Figure 1*. Then change it into a tetra-pyramid $OVYX'Z'$ with a rectangular undersurface ($VX'=YZ'$) and $OV=OX'=OY=OZ'=100\%$. Make $XN \perp OY$ and $ZN' \perp OV$, connect the lines shown in *Figure 1*. Module length quantitative QDS (ΔW) is deduced on $W+XX'=OX'=100\%$, projection content QDS (ΔC) from $C+NY'=OY=100\%$, overall content QDS (ΔR) from $R+ZZ'=OZ'=100\%$ and quantitative QDS (ΔP) according to $P+N'V=OV=100\%$, see Eq.[9]-[12]. Above four quantitative similarities W , C , R and P , the higher the errors the bigger the dissimilarities.

$$\Delta W = W - 100\% = -XX' = \frac{\sqrt{\sum_{i=1}^n x_i^2} - \sqrt{\sum_{i=1}^n y_i^2}}{\sqrt{\sum_{i=1}^n y_i^2}} \times 100\% \quad [9]$$

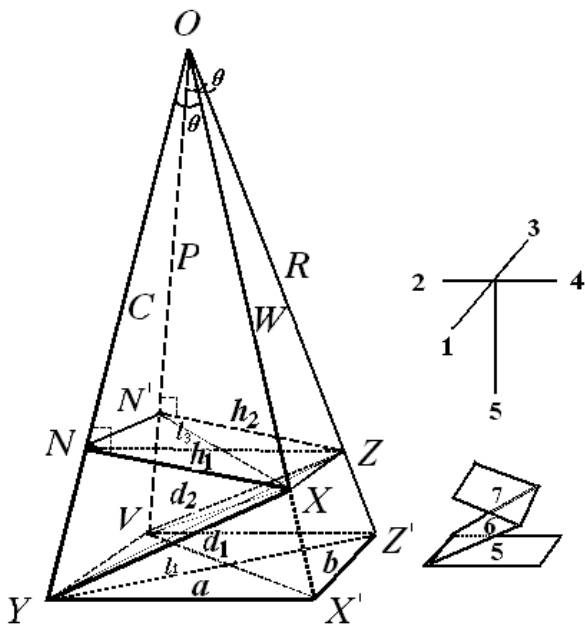


Figure 1 Tetra-pyramid constructed by the first and second level DQTS based on vectors in *n*-dimension space of \bar{X} and \bar{Y}

broaden the quantitative evaluating scope of vectors in *n*-dimension space.

$$d_1 = \frac{|\bar{X} - \bar{Y}|}{|\bar{Y}|} \times 100\% = \frac{\sqrt{\sum_{i=1}^n (x_i - y_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2}} \times 100\% = \sqrt{1 + W^2 - 2WS_F} = \sqrt{1 + W^2 - 2C} \quad [13]$$

$$h_1 = W \sin \theta = \sqrt{W^2 - C^2} = \sqrt{d_1^2 - \Delta C^2} \quad [14]$$

In the same way, in ΔVOZ and $\Delta ZON'$, because of $ZN' \perp OV$, *Euclidean*-liked QDS (d_2) and end vertical distance quantitative similarity (h_2) are acquired according to R and P . Eq.[15] and Eq.[16] are their formulations ($h_2 < d_2$).

$$d_2 = \sqrt{1 + R^2 - 2RS_F} = \sqrt{1 + R^2 - 2P} \quad [15]$$

$$h_2 = R \sin \theta = \sqrt{R^2 - P^2} = \sqrt{d_2^2 - \Delta P^2} \quad [16]$$

d_1, d_2, h_1, h_2 could quantitatively reflect the QDS between \bar{X} and \bar{Y} very well ($d_i > h_i$). When they are equal to zero, \bar{X} and \bar{Y} are thoroughly equal. Obviously, tangent value shown as Eq.[17] can qualitatively explain the QDS of \bar{X} and \bar{Y} , and is named tangent QDS. The more $tg\theta$ is close to zero, the more \bar{X} and \bar{Y} are similar in content.

$$tg\theta = \frac{h_1}{C} = \frac{h_2}{P} \quad [17]$$

$$\Delta C = C - 100\% = -NY = \frac{\sum_{i=1}^n (x_i - y_i)y_i}{\sum_{i=1}^n y_i^2} \times 100\% \quad [10]$$

$$\Delta R = R - 100\% = -ZZ' = \frac{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i}{\sum_{i=1}^n y_i} = \frac{\sum_{i=1}^n (x_i - y_i)}{\sum_{i=1}^n y_i} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n y_i} \quad [11]$$

$$\Delta P = P - 100\% = -N'V \quad [12]$$

It is generally considered that *Euclidean* distance is the most famous and convenient to describe the differences between vectors. However, it cannot point out how much in percentage they differ relatively to \mathbf{RV} . So *Euclidean* QDS (d_1) is proposed to describe its magnitude relative to \mathbf{RV} module length, in which d_1 is percentage of $\bar{X} - \bar{Y}$ module length to that of \bar{Y} to quantitatively reveals how much \bar{X} and \bar{Y} differ from each other and with Eq.[13]. The ratio of vertical distance from the vertex of \bar{X} to \bar{Y} and the \bar{Y} module length is defined as end vertical QDS (h_1) as shown in Eq.[14], where it reflects the QDS between \bar{X} and \bar{Y} ($h_1 \leq d_1$).

The tetra-pyramid consists of 4 triangles [1-4] and 3 quadrilaterals [5-7], in which W and R construct the **1L-DQTS** in the 4th triangle, C and P construct the **2L-DQTS** in the 2nd triangle. What is more, the lengths of all lines (in total 17 except 3 replicates) in the hexahedron $VYX'Z'NN'XZ$ can better reflect the quantitative QDS between \bar{X} and \bar{Y} , which

Characteristics of auxiliary rectangular undersurface (the 5th surface)

In tetra-pyramid, $\angle VOY = \angle XOZ = \zeta$ ($0 \leq \zeta \leq \pi$) and $\angle VOX = \angle YOZ = \lambda$ ($\theta \leq \lambda \leq \pi$), from the rectangular undersurface $\square VYX'Z'$ and cosine theorem, the length of hemline (a) is computed as Eq.[18], the width of hemline (b) is calculated as Eq.[19], and the diagonal length of the undersurface (l_1) can be seen in Eq.[20]. Without a doubt, let $0 \leq \zeta \leq \theta$, then a, b and l_1 all could reflect the difference degree between SV and RV . When they all trend to zero, the qualitative similarity is the most satisfactory. Let $\mu_1 = \cos \zeta$, Eq.[21] is derived from Eq.[20], we use $S_c = \cos \lambda$ to reflect the magnitude of similarity between P and W , or between C and R . Because $\mu_1 = \cos \zeta$ is an uncertain value, when $\zeta = 0, b = 0$ and $a = l_1$; when $\zeta = \theta, a = b$, and l_1 as displayed in Eq.[22].

$$a = X'Y = VZ' = \sqrt{2(1 - S_F)} \quad [18]$$

$$b = VY = XZ' = \sqrt{2(1 - \cos \zeta)} = \sqrt{2(1 - \mu_1)} \quad [19]$$

$$l_1 = YZ' = VX' = \sqrt{VZ'^2 + XZ'^2} = \sqrt{4 - 2(\cos \theta + \cos \zeta)} = \sqrt{2 - 2\cos \lambda} \quad [20]$$

$$\cos \lambda = \cos \zeta + S_F - 1 = \mu_1 + S_F - 1 \quad [21]$$

$$l_1 = 2\sqrt{1 - \cos \theta} = 2\sqrt{1 - S_F} \quad [22]$$

Characteristics of quadrilateral undersurface XYVZ (the 6th surface)

In quadrilateral XYVZ, $XY=d_1$, $VZ=d_2$, $VY=b$ and $XZ=c_1$. According to cosine theorem, the lengths of other sides and the diagonal lengths are respectively measured as shown in Eq.[23]-[25], which also represent the QDS between \bar{X} and \bar{Y} . In addition, $VX (YZ) > d_1 (d_2)$.

$$c_1 = XZ = \sqrt{R^2 + W^2 - 2RW\cos\zeta} \quad [23]$$

$$l_2 = VX = \sqrt{1 + W^2 - 2W\cos\lambda} \quad [24]$$

$$l'_2 = YZ = \sqrt{1 + R^2 - 2R\cos\lambda} \quad [25]$$

Characteristics of quadrilateral section N'NXZ (the 7th surface)

In quadrilateral N'NXZ, $NX=h_1$, $N'Z=h_2$, $XZ=c_1$ and $NZ (N'X) > h_1 (h_2)$. On the basis of cosine theorem, the lengths of other sides and the diagonal lengths are calculated as displayed in Eq.[26]-[28], which also can disclose the QDS of \bar{X} and \bar{Y} .

$$c_2 = NN' = \sqrt{C^2 + P^2 - 2CP\cos\zeta} \quad [26]$$

$$l_3 = NX = \sqrt{P^2 + W^2 - 2PW\cos\lambda} \quad [27]$$

$$l'_3 = NZ = \sqrt{C^2 + R^2 - 2CR\cos\lambda} \quad [28]$$

In basic condition of $\zeta=0$, $\lambda=\theta$ the simple and practical formulars were listed as follow:

$$c_1 = XZ = \sqrt{W^2 + R^2 - 2RW} = W - R \quad [29]$$

$$l_2 = VX = \sqrt{1 + W^2 - 2C} \quad [30]$$

$$l'_2 = YZ = \sqrt{1 + R^2 - 2P} \quad [31]$$

$$c_2 = NN' = \sqrt{C^2 + P^2 - 2CP} = C - P \quad [32]$$

$$l_3 = NX = \sqrt{P^2 + W^2 - 2CP} \quad [33]$$

$$l'_3 = NZ = \sqrt{C^2 + R^2 - 2CP} \quad [34]$$

In a word, there are 15 parameters for QDS evaluation derived from both 1L-DQTS (W and R) and 2L-DQTS (C and P) to quantitatively assess the difference between \bar{X} and \bar{Y} , one of which is closely relevant to *Euclidean* distance. Their specific applications depend on the existing problems, and must control $0 \leq \zeta \leq \theta$, in which these QDSs can be denoted by percentage. If QS_i stands for each sample with the total content similarity, then how much \bar{X} and \bar{Y} are quantitatively resemble is represented by **QDS**, shown in Eq.[35], where tangent DS is the ratio of two quantitative values also belonging to quantitative instinct, and certainly

Eq.[35] will give a new QDS method with $tg\theta$. These above new QDSs were summarized in *Table 1*.

$$QS_i = 100\% - QDS_i \quad [35]$$

Application of above QDSs in FLJ quality comparisons

FLJ is the dry flower bud of *Lonicera japonica* Thunb., with or without young flower (27), which contains chlorogenic acid (CGA), isochlorogenic acid, flavonoids, saponins and volatile oil, etc.. According to the latest statistics by China Economic Forest Association, the annual output of FLJ is 800 million kg in China and its domestic and abroad market demands could be up to 2,000 million kg. Especially during SARS epidemic in 2003, because of good effect of heat-clearing and detoxifying, FLJ had become a shining star of antiSARS and the inventories were once empty. For saling FLJ in the biggest herbal market named Anguo city in Hebei province, China, there were several herbal merchants to become the millionaires over one night, nearly every day emerging one at the period. Though FLJ price had risen over 33 dollars from 3, it is very difficult to buy it. Being so important, we established FLJ-HPLC-FPs to regard the every fingerprint peak area respectively as each element of n -dimensional vector to effectively compared their quality differences by applying above QDSs.

Experimental

Apparatus and reagents

The analysis was performed with an Agilent 1100 HPLC series (Hewlett Packard, CA, USA), consisting of DAD detector, low pressure quaternary pumps, online degasser and an autosampler. All data acquired were proceeded by ChemStation workstation (Agilent technology Inc.). KQ-50B ultrasonic bath (Kunshan ultrasonic instrument company limited, China) and Sartorius-BS110S analytic scale (Saiduolisi scale company limited, Beijing, China) were used during the analysis process. Methanol, acetonitrile (purchased from Yuwang industry limited company, Shandong, China) and glacial acetic acid (provided by Kemiou chemicals developing center, Tianjin, China) were all of HPLC grade. Tetrahydrofuran (THF) of chromatographic grade and ethanol of analytical grade were from Concord Technology Co., Ltd (Tianjin, China). The other reagents were of analytical grade. Deionized water was used for the preparation of all samples and solution. Chlorogenic acid

(CGA), Caffeic acid (CFA), and Luteolin (LTL) used as control substances were obtained from National Institute for the Pharmaceutical and Biological Products (Beijing, China) for markers. 3,5-O-dicaffeoylquinic acid (DCQ) was self-made, whose purity was more than 98.6% determined by HPLC. The habitats of 14 batches of raw herbs of FLJs were shown as follows: S1 Linyi, Shandong province; S2 Julu, Hebei province; S3 Hanzhong, Shanxi province; S4 Shangluo, Shanxi province; S5 Baoji, Shanxi province; S6 Xinxiang, Henan province; S7 Pingyi, Shandong province; S8 Feixian, Shandong province; S9 Shandong province, S10 Fengqiu, Henan province; S11 Zhashui, Shanxi province; S12 Wanrong, Shanxi province; S13 Longhui, Hunan; S14 Henan province. All of them were identified by professor Shiyi XU who researches on botany at Shenyang Pharmaceutical University (China) as *Lonicera japonica* Thunb. except S4, S13 and S14, which were *Lonicera confusa* DC.

Solution preparation

Control solution

120 $\mu\text{g}\cdot\text{mL}^{-1}$ of CGA was prepared by mixing accurately weighed CGA with 20% methanol, and 150 $\mu\text{g}\cdot\text{mL}^{-1}$ of CFA was also prepared by same method. 680 $\mu\text{g}\cdot\text{mL}^{-1}$ of DCQ was prepared by dissolving accurately weighed in methanol, and 230 $\mu\text{g}\cdot\text{mL}^{-1}$ of LTL was also prepared, all which were stored in refrigerator before use.

Test solution of sample

After being dried under 60 °C for 40 min and crushed into powder, 5.0 g of FLJ was weighted accurately. Heat reflux extraction was performed twice with 100 and 60 mL water for 2 h, respectively. The combined filtrate were concentrated to about 20 mL under reduced pressure. Then, ethanol was added to the extracts until its concentration about 80% (v/v). After being stood away from light and kept cool 4 °C for 24 h, ethanol was removed completely from the filtrate using a rotary evaporator at 50 °C. After repeating the alcohol precipitation, the residue was diluted to 50 mL with water and filtered through 0.45 μm microporous membrane before injection.

HPLC conditions

The chromatographic separation was performed on a Century SIL C18 AQ column (250 mm \times 4.6 mm, 5 μm) (Dalian Johnson Separation Science & Technology Corporation, China) at column temperature of (30 \pm 0.15) °C. The mobile phase consisted of binary mixture of solvent

A (water, containing 0.5% acetic acid) and solvent B (acetonitrile, containing 5 % THF and 0.5% acetic acid) with a linear gradient program as follows: 0-8 min, 2-5% B; 8-30 min, 5-20% B; 30-60 min, 20-40% B; 60-75 min, 40-50% B. The flow rate was 1.0 $\text{mL}\cdot\text{min}^{-1}$ and DAD detector was set at 254 nm, where UV spectra and 3D-plots were recorded between 200 and 400 nm. The injection volume of test solution and control solution were both 5 μL , and eluting time was 75 min.

System suitability tests and selection of reference peak

Five microliters of CGA, CFA, DCQ, LTL and test solution were injected into the chromatograph, respectively, and their chromatograms were recorded as shown in *Figure 2*. The components thus determined were: peak 9- CGA, peak 11- CFA, peak 21- DCQ. Almost no LTL was detected for its low content in test solution. Because of stronger signal, baseline resolution with the adjacent peaks and moderate retention time, peak 9 was selected as the referential peak (RP), whose theoretical plate number were above 140,000.

Injection precision, sample stability test and method repeatability

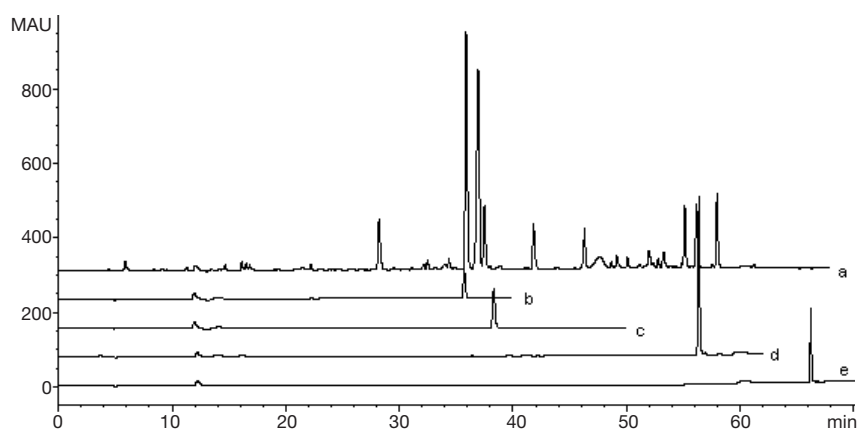
Replicately injected the test solution of S10 for 5 times and loaded the test solution of S10 at 0, 1, 5, 15 and 25 h after fresh preparation, respectively, and analyzed 5 independently prepared test solution of S10 to assess the analysis method. All the relative standard deviation (RSD) of relative retention time and of relative peak area were all less than 1% and 3%, respectively, which indicated this method was pretty precise and good in producibility, and stable for test solution within 25 h.

Development of FL-HPLC fingerprint

Respectively, test solutions of S1 to S14 were analyzed and their chromatograms were recorded as shown in *Figure 3*, in which all fingerprint peaks were eluted within 65 min. Supposed that all the common peaks could be observed in every chromatogram, there were 22 co-possessing peaks marked in the fingerprints. Besides S4, S13 and S14, integral signal of chromatograms was introduced into the professional software named Super-information Characteristic Digitalized Evaluation System for Chromatographic Fingerprints of Traditional Chinese Medicine (Version 4.0), invented by Guoxiang Sun, to yield the referential fingerprint (RFP) by means of average method that synthysized the mean model from S1-S3 and S5-S12 (excluding S4, S13 and S14 due to their lower

Table 1 The fifteen quantitative dissimilarities and one similarity for comparison among fingerprints

No.	QDS_i	Formular	Name & physical means	No.	QDS_i	Formular	Name & physical means
1	ΔW	$W-1$	Module QDS	10	c_1	$W-R$	WR QDS
2	ΔR	$R-1$	Apparent QDS	11	c_2	$C-P$	CP QDS
3	ΔC	$C-1$	Projection QDS	12	l_1	$\sqrt{2(1-S_F)}$	Full QDS
4	ΔP	$P-1$	Elemental QDS	13	l_2	$\sqrt{1+W^2-2C}$	WC cross QDS
5	d_1	$\sqrt{1+W^2-2C}$	Euclidean QDS	14	l_2'	$\sqrt{1+R^2-2P}$	RP cross QDS
6	d_2	$\sqrt{1+R^2-2P}$	RP cross QDS	15	l_3	$\sqrt{W^2+P^2-2PC}$	WPC cross QDS
7	h_1	$\sqrt{W^2-C^2}$	WC square QDS	16	l_3'	$\sqrt{R^2+C^2-2CP}$	RCP cross QDS
8	h_2	$\sqrt{R^2-P^2}$	RP square QDS	17	$tg \theta$	$\frac{h_1 \text{ or } h_2}{C \text{ or } P}$	Tangent DS
9	a	$\sqrt{2(1-S_F)}$	Full QDS	18	b	$b=0$ when $\zeta=0, \lambda = \theta$	Basic condition

**Figure 2** The HPLC chromatograms of *Flos Lonicerae* (a), CGA (b), CFA (c), DCQ (d) and LTL (e)

qualitative similarities), shown in *Figure 3* (RFP).

The chromatograms of S4, S13 and S14 significantly differed from those of the rest 11 batches of FLJ. The various similarities between samples of S1 to S14 and the RFP were calculated as shown in *Table 2*. S_F and S_F' of the three batches of *Lonicera confusa* DC. (S4, S13, S14) were far less than 0.90, thus they were considered as defective goods and could not be regarded as *Lonicera japonica* Thunb. Although S_F of S9 and S12 were more than 0.90, their S_F' were less than 0.90 and neither regarded as qualified. Defined $\Delta S_F = S_F - S_F'$ and that of S9 and S12 were all above 0.10, which also indicated a bigger discrepancy in the homogenizing coefficient (indicating the uniformity of fingerprint content distribution) γ_i of the above two samples that were significantly higher than that of RFP. For the 11 batches of FLJs, its $|C-P| < 12\%$ demonstrated rigorously that they were qualified for their distribution

proportion of chemical composition contents with that of RFP. If the mean of the W, R, C, P was denoted as \bar{P} that was controlled within 70-130%, then the contents of S3, S9, S10 and S12 were outliers that all had the bigger QDSs more than 24. Though the QSs of S4, S13 and S14 were pretty qualified, the qualitative similarities had identified they were not belong to authentic FLJs with the much larger QDSs. According to QDS and DS of samples, S1, S2, S5, S6, S7, S8 and S11 were completely qualified for having the lower QDSs and adequate QSs, in which we calculated the mean of 11 kinds of QDSs in terms of the formular [36] to assess the sample differences. Calculated the mean of d_1 and d_2 ; h_1 and h_2 ; l_2 and l_2' ; l_3 and l_3' to be represented by d, h, l_2 and l_3 , respectively, to draw *Figure 4* with l_1, a and $tg\theta$ in order to acquire which is the best QDS parameter.

$$\overline{QDS} = \frac{1}{6}(d_1 + d_2 + h_1 + h_2 + l_1 + l_2 + l_2' + l_3 + l_3' + tg\theta + a) \quad [36]$$

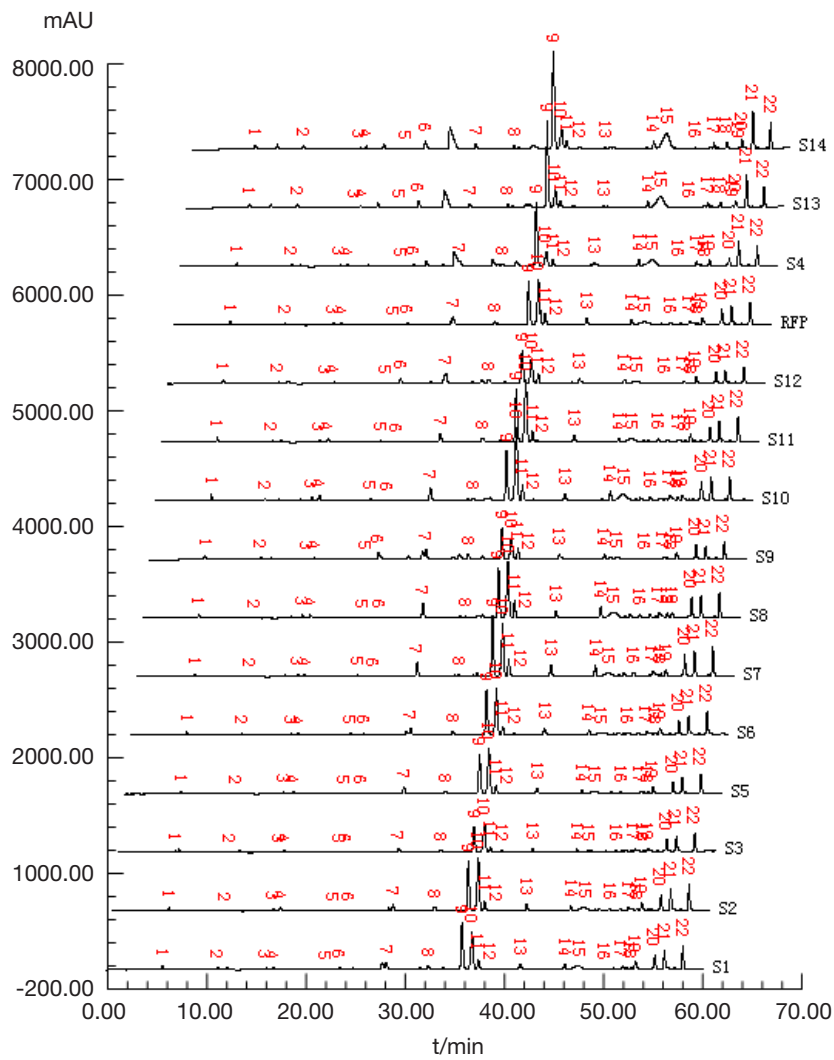


Figure 3 HPLC-FPs of *Flos Lonicerae Japonicae* (S1 to S14) and the RFP

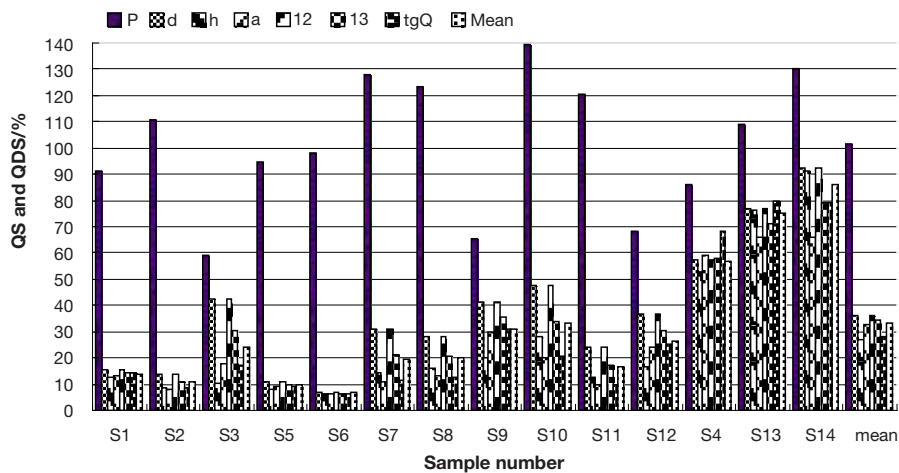


Figure 4 The comparison among the QS and various QDSs of *Flos Lonicerae Japonicae* (S1 to S14)

Table 2 The qualitative quantitative similarities and dissimilarities of FLJ from different places

Para.	S1	S2	S3	S5	S6	S7	S8	S9	S10	S11	S12	S4	S13	S14	Mean
S_F	0.991	0.997	0.984	0.996	0.998	0.994	0.991	0.955	0.980	0.995	0.971	0.827	0.781	0.783	0.946
S_F'	0.965	0.969	0.900	0.963	0.980	0.947	0.931	0.730	0.943	0.962	0.858	0.852	0.782	0.756	0.896
ΔS_F	0.03	0.03	0.08	0.03	0.02	0.05	0.06	0.23	0.04	0.03	0.11	-0.03	0.00	0.03	0.050
γ	0.565	0.533	0.549	0.509	0.538	0.580	0.583	0.632	0.538	0.508	0.615	0.525	0.502	0.509	0.55
W	91.5	113.8	60.1	99.1	100.5	126.1	121.6	63.1	144.1	126.8	66.2	97.1	129.1	153.3	106.6
R	92.2	108.3	58.9	90.1	96.4	130.5	126.6	71.2	138.2	115.1	72.6	90.9	115.6	139.2	103.3
C	90.6	113.4	59.2	98.7	100.3	125.3	120.6	60.3	141.2	126.2	64.2	80.3	100.9	120.1	100.1
P	91.4	108.0	58.0	89.8	96.2	129.7	125.5	67.9	135.4	114.4	70.5	75.2	90.3	109.0	97.2
\bar{P}	91.4	110.9	59.1	94.4	98.4	127.9	123.6	65.6	139.7	120.6	68.4	85.9	109.0	130.4	101.8
ΔW	-8.5	13.8	-39.9	-0.9	0.5	26.1	21.6	-36.9	44.1	26.8	-33.8	-2.9	29.1	53.3	6.6
ΔR	-7.8	8.3	-41.1	-9.9	-3.6	30.5	26.6	-28.8	38.2	15.1	-27.4	-9.1	15.6	39.2	3.3
ΔC	-9.4	13.4	-40.8	-1.3	0.3	25.3	20.6	-39.7	41.2	26.2	-35.8	-19.7	0.9	20.1	0.1
ΔP	-8.6	8.0	-42.0	-10.2	-3.8	29.7	25.5	-32.1	35.4	14.4	-29.5	-24.8	-9.7	9.0	-2.8
ΔD_P	-8.6	10.9	-41.0	-5.6	-1.7	27.9	23.6	-34.4	39.7	20.6	-31.6	-14.1	9.0	30.4	1.8
d_1	15.6	16.0	42.2	8.6	6.3	28.7	26.1	44.0	50.3	29.3	39.2	58.0	80.6	97.4	38.7
d_2	14.9	11.4	43.2	12.6	7.3	33.0	30.5	38.6	44.9	19.2	34.2	56.8	72.8	87.0	36.2
\bar{d}	15.4	13.9	42.7	10.8	6.8	31.0	28.1	41.2	47.6	24.1	36.7	57.4	76.7	92.2	35.9
h_1	12.8	9.5	10.4	8.9	6.3	14.2	15.6	18.6	28.8	12.3	16.1	54.6	80.5	95.3	27.4
h_2	12.1	8.1	10.3	7.3	6.2	14.4	16.7	21.4	27.7	12.7	17.3	51.1	72.2	86.6	26.0
\bar{h}	12.5	8.8	10.4	8.1	6.3	14.3	16.2	20.0	28.3	12.5	16.7	52.9	76.4	91.0	26.7
a	13.4	7.7	17.9	8.9	6.3	11.0	13.4	30.0	20.0	10.0	24.1	58.8	66.2	65.9	32.9
c_1	-0.7	5.5	1.2	9	4.1	-4.4	-5	-8.1	5.9	11.7	-6.4	6.2	13.5	14.1	3.3
c_2	-0.8	5.4	1.2	8.9	4.1	-4.4	-4.9	-7.6	5.8	11.8	-6.3	5.1	10.6	11.1	2.9
\bar{c}	-0.8	5.5	1.2	9.0	4.1	-4.4	-5.0	-7.9	5.8	11.8	-6.3	5.6	12.1	12.6	3.1
l_1	13.4	7.7	17.9	8.9	6.3	11.0	13.4	30.0	20.0	10.0	24.1	58.8	66.2	65.9	32.9
l_2	15.9	16.4	42.1	9.0	6.3	29.0	25.8	43.8	50.2	29.0	39.3	58.0	80.5	97.4	36.7
\dot{l}_2	14.9	11.4	43.2	12.6	7.3	33.0	30.5	38.6	44.9	19.2	34.2	56.8	72.8	87.0	35.1
\bar{l}_2	15.4	13.9	42.7	10.8	6.8	31.0	28.1	41.2	47.6	24.1	36.7	57.4	76.7	92.2	35.9
l_3	12.8	11.0	10.4	12.6	7.5	14.8	16.3	20.1	29.3	17.1	17.3	54.8	81.2	95.9	36.8
\dot{l}_3	12.1	9.7	10.3	11.5	7.4	15.1	17.4	22.7	28.3	17.3	18.4	51.3	72.9	87.3	35.1
\bar{l}_3	12.5	10.3	10.4	12.1	7.5	15.0	16.8	21.4	28.8	17.2	17.9	53.1	77.1	91.6	35.9
tg	14.1	8.4	17.5	9.0	6.3	11.3	12.9	30.8	20.4	9.8	25.2	68.0	79.8	79.3	28.1
Mean	13.8	10.7	24.1	10.0	6.7	19.6	19.9	30.8	33.2	16.9	26.3	57.0	75.1	85.9	33.3

Conclusions

Dissimilarity comparison based on vectors in n -dimension space involves overall qualitative similarity and overall quantitative similarity i.e. Qs and QDs. The d and \bar{l}_2 were much near furthermore a and $tg\theta$ had the closely values. In the present work, several new calculation

methods for QDs had been first set up, which could substitute *Euclidean distance* to quantitatively reflect the DSs between two n -dimension vectors. Their remarkable characteristics stressed on the differences and compared to Qs in diversity reflected by them were more benefit to classification and evaluation for all things in

n -dimensional variances.

The better combination of both the DQLS and the 1L-DQTS or 2L-DQTS could exactly solve the problem of both overall qualitative evaluations and overall quantitative evaluations of all objects with vectors in n -dimension space, in which there 15 QDSs to display the disparity. In a word, the novel QDSs will be getting more important in 21 century for nature and human.

Acknowledgements

This research work was financially supported by National Natural Science Foundation of China (Accession No.90612002) belong to the Important Research Plan "Study on the Scientific Action Circumstances Based on the Internet", and by Great Projects of the Tenth Five Years Plan granted by Ministry of Science and Technology of China. We also thank National Codex Committee of China for the Funding of the Research Plan of Traditional Chinese Medicine Injection Fingerprints.

Disclosure: The authors declare no conflict of interest.

References

- Sierra G. A simple method for the detection of lipolytic activity of micro-organisms and some observations on the influence of the contact between cells and fatty substrates. *Antonie Van Leeuwenhoek* 1957;23:15-22.
- Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 1958;38:1409-39.
- Jaccard P. Nouvelles recherches sur la distribution florale. *Bulletin Société Vaudoise des Sciences Naturelles* 1908;44:223-70.
- Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des Sciences Naturelles* 1901;37:241-72.
- Rogers DJ, Tanimoto TT. A Computer Program for Classifying Plants. *Science* 1960;132:1115-8.
- Pearson K. Mathematical contributions to the theory of evolution - on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc R Soc London* 1897;60:489-502.
- Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1948;26:297-302.
- Bray JR, Curtis JT. An ordination of the upland forest communities of southern wisconsin. *Ecol Monogr* 1957;27:325-49.
- Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 1978;89:583-90.
- Nei M, Li WH. Mathematical model for studying genetical variation in terms of restriction endonucleases. *Proc Natl Acad Sci* 1979;74:5269-73.
- Zwick R, Carlstein E, Budescu DV. Measures of similarity among fuzzy concepts: A comparative analysis. *Int J Approximate Reasoning* 1987;1:221-42.
- Ichino M, Yaguchi H. Generalized Minkowski metrics for mixed feature type data analysis. *IEEE Trans on Sys Man & Cyb* 1994;24:698-708.
- Aitchison J. The statistical analysis of compositional data. *J R Stat Soc Ser B* 1982;44:139-77.
- Bhattacharya A. On a measure of divergence between two multinomial populations. *Sankhyā* 1946;7:401-6.
- Ralescu AL, Ralescu DA. Probability and fuzziness. *Inf Sci* 1984;34:85-92.
- Saraçoğlu R, Tütüncü K, Allahverdi A. A fuzzy clustering approach for finding similar documents using a novel similarity measure. *Exp Sys Appl* 2007;33:600-5.
- Gowda KC, Diday E. Symbolic clustering using a new dissimilarity measure. *Pett Recognit* 1991;24:567-78.
- Stephenson W, Williams WT, Cook SD. Computer Analyses of Petersen's Original Data on Bottom Communities. *Ecol Monogr* 1972;42:387-415.
- Zadeh LA. Fuzzy sets. *Inf Control* 1965;8:338-53.
- Li RJ. Fuzzy method in group decision making. *Comput Math Appl* 1999;38:91-101.
- Ölçer Aİ, Odabaşı AY. A new fuzzy multiple attributive group decision making methodology and its application to propulsion/manoeuvring system selection problem. *Eur J Ope Res* 2005;166:93-114.
- Wang XZ, De Baets B, Kerre E. A comparative study of similarity measures. *Fuz Sets Sys* 1995;73:259-68.
- Yeh CH, Deng H, Chang YH. Fuzzy multicriteria analysis for performance evaluation of bus companies. *Eur J Ope Res* 2000;126:459-73.
- Chang YH, Yeh CH. A survey analysis of service quality for domestic airlines. *Eur J Ope Res* 2002;139:166-77.
- Sun G, Ren P, Bi Y, et al. Fingerprints of ginkgo leaf extract and dipyrindamole injection by double qualitative similarities and double quantitative similarities. *Se Pu* 2007;25:518-23.
- Sun GX, Hou ZF, Zhang CL, et al. Comparison between the qualitative similarity and the quantitative

similarity of chromatographic fingerprints of traditional Chinese medicines. Yao Xue Xue Bao

2007;42:75-80.

27. Chinese Pharmacopoeia. 2005;I;152.

Cite this article as: Sun G, Song W, Deng X, Liu Z. Quality comparison of *Flos Lonicerae Japonicae* by several dissimilarity methods. Ann Transl Med 2013;1(1):7. doi: 10.3978/j.issn.2305-5839.2013.01.02