

Two Chinese medicine species constants and the accurate identification of Chinese medicines

Huabin zou

School of chemistry and chemical engineering of Shandong University, Jinan 250100, P.R.China

Correspondence:huabinzou@126.com

Abstract: Since the ancient times, all over the world, the identification of herbal medicines have to be based on empirical knowledge. In this article two species constants of traditional Chinese medicines(TCM) were discovered relying on the maximum information states of Dual index information theory equation, or common heredity and variation information theory. The two species constants,common peak ratios $P_g = 61\%$ and $P_g = 70\%$, which corresponding to symmetry and asymmetry variation states,respectively, were used as two absolute quantitative criteria to identify complex biology systems—TCM. Considered the influences of many other factors on components and experiment processes, the practical theoretical identification standards should be established $P_g \cong 58\sim 64\%$ and $P_g \cong 67\sim 73\%$, within the relative error within -3% and $+3\%$ of information value around the maximum information states. Combining the maximum number of effective sample optimum method with this two theoretical standards, the optimized classification of a TCM sample set can be carried out correctly. 42 samples belonging to four species of combination Chinese medicines were tested. The infrared (IR) fingerprint spectra (FPS) of their powder were measured and analyzed by means of the approach provided above. Among the six pairs of four Chinese medicine species, five of them follow the species constant $P_g = 61\%$, one of them obeys the $P_g = 70\%$. The correct recognition ratio of samples was 95.2%, and that of species was 100%.

Keywords : species constant, identification, classification, traditional Chinese medicines, information theory,heredity and variation, biological pattern recognition, fingerprint spectra.

Introduction

The species of herbal medicines determine their efficacy, thus the accurate identification of species is very significant for herbal medicines, and for traditional Chinese medicines whose major part is composed of herbs and animals. In fact,the identification of Chinese medicines belongs to that of biology species. However, until recently there is no one unquestionably scientific definition or quantitatively standard relative to biology species[1,2,3]. This also leads to classification of complex organisms is in chaos which hampers biology conservation heavily presently [4]. Most importantly, in biology, species is the fundamental of biological diversity, and is the ultimate goal of systems biology [5]. In practical the classification or pattern recognition of Chinese medicines as well as biology is performed depending on some macro properties, such as macro characters, biological functions, geographic distributions,

ecological environments and so forth. In the procedure, the criteria were established only rest on subjective deduce by researchers. Recently, all definitions about biology species lack of principle theory,were not based on experimental science and mathematical principles, and also are short of the quantitative standards. That is, the identification of biology and Chinese medicines are totally depending on their shapes, sizes, colors, tastes and the like, and empirical knowledge[6]. Even if in modern numerical taxonomy, the criteria are determined on the basis of a lot of learning-based experience.

As a kind of complex biology systems, whether TCMs are of some intrinsically accurate laws and pose of some precisely quantitative criteria, that is, some constants,like in physics and chemistry systems. Thus by applying some accurately quantitative criteria we can determine the TCM species precisely without experience or prior knowledge. This is a greatly significant scientific subject need to be investigated seriously. Author Zou has paid a great effort to searching for the mathematical theory for identifying traditional Chinese medicines and biological systems for long time, grounded on intrinsic properties and elemental principle in nature. He proposed and established independently the common and variant peak ratio dual index sequence analysis method during his doctorate work[7-13]. This theory has been adopted by many researchers to identify herbs,plants, foods and animals, etc, and more than 50 papers using this method have been published so far,such as papers [14-26]. Based on all these works, author Zou and coworkers established the dual index grade sequence individualized pattern recognition approach [27-34]. The Chines medicines' species can be identified accurately relying on this method by selecting suitable similarity scales $P_g \cong \bar{P}_g + x S$, where $-3 \cong x \cong 3$,without any help

of empirical knowledge and learning samples. In the procedure, by change x smoothly to obtain each sample's characteristic sequence. Samples in core characteristic sequences of each class construct an independent set, and different sets represent different classes. However, this approach can't offer any absolute quantitative criterion for Chines medicine species.

In paper [35] a theory—the Dual index information theory or heredity and variation information theory equation was proposed. According to this theory, there are two intrinsically variation mechanism, which are symmetry and asymmetry variations. Based on the two variation mechanisms, two common peak ratios $P_g= 0.61$ and $P_g= 0.70$, corresponding to the maximum information states, are defined as two constants, in which the constant related to symmetry variation was proved to fit to classify some combination Chinese medicines [35,36,37] successfully, by means of standard $P_g \cong 61\%$. These complex biology systems include components extracted from Guifu Dihuang Pills, Jinkui Shenqi Pills with absolute ethanol[35], the powder of Mingmu Dihuang Pills, Zhibai Dhuang Pills, and Maiwei Dihuang Pills[36], the components of Buzhong Yiqi Pills and Shiquan Dabu Pills extracted with chloroform[37]. While there is no report about the other standard $P_g \cong 70\%$ applied to identification biology systems. This article focuses on to investigate the relationship between Chinese medicine species pairs, in order to find whether some Chinese medicines species

obey the two absolute quantitative criteria, especially the constant $P_g \geq 70\%$. In this paper, 6 pairwise of four species of Chinese medicines: Guifu Dihuang Pills, Jinkui Shenqi Pills, Mingmu Dihuang Pills, Zhibai Dihuang Pills were investigated subtly. The infrared fingerprint spectra of 42 samples' powders belonging to the four species were measured and analyzed. Compared to components extracted from Chinese medicines with some kind of solvent, compounds existed in Chinese medicine powder can fully reflect their properties or features. The Infrared fingerprint spectra of their powder are better than that of their extracted components for verifying theory. All six pairwise of these four Chinese medicines were tested, the results showed among them, five pairs were successfully classified based on $P_g \geq 61\%$, and only one pair Guifu Dihuang Pills and Zhibai Dihuang Pills obeys the criterion $P_g \geq 70\%$. These 42 samples and the six Chinese medicine pairs were distinguished perfectly. The accuracy of sample recognition was 95.2%. These results proved that there really exist the two constants in some complex organisms, and they can be defined as Chinese medicine species constants, which exhibit the intrinsic quality characteristic. Depending on these two Chinese medicine species constants, some biology systems can be precisely identified, pattern recognized, without help of any prior knowledge relative to sample set.

The Infrared fingerprint spectra of these 42 samples' powder were also analyzed by means of the dual index grade sequence individualized pattern recognition method [27-34]. When selected the similarity scale to be $P_g \geq \bar{P}_g + 1.1S$ and $P_g \geq \bar{P}_g + 1.3S$, these 42 samples were accurately classified with the correct ratio being 90.5% and 95.2%, respectively. The results gained rest on the two theoretical approaches are equivalent to each other. While the two Chinese medicine species constants are the absolute quantitative criteria, they make the identification of biology systems and Chinese medicines being of the most simplicity, clear physics and biology meanings. On the other hand, this work, to put forward two Chinese medicine species constants may be a breakthrough from ancient times.

1 Two TCM species constants

According to the Dual index information theory equation-the heredity and variation information theory equation[35], the information of any two biological samples or any two evolution stages of one biological system can be calculated by means of the follow equation:

$$I = -(P_g \ln P_g + P_a \ln P_{va} + P_b \ln P_{vb}) \quad (1)$$

The meaning of every parameters /variables and its definition are showed as follows, as in paper[6-10].

$$P_g = \frac{N_g}{N_g + n_a + n_b} \times 100\% = \frac{N_g}{N_d} \times 100\% \quad (2)$$

$$P_{va} = \frac{n_a}{N_g} \times 100\% \quad (3)$$

Here P_g is common peak ratio of fingerprint spectra a and b . Its value is the ratio of number of common peaks N_g to that of total kinds of peaks N_d in a and b , N_d is defined as independent peaks, that is, N_d is the kinds of peaks. N_g are the peaks existed both in fingerprint spectra a and b .

Interestingly, this parameter is the same as Jaccard parameter, Sneath and Sokal parameter[38]. n_a and n_b are the number of variation peaks only existed in fingerprint spectra a and b , respectively.

$$P_{va} = \frac{n_a}{N_g} \times 100\% \quad (4)$$

$$P_{vb} = \frac{n_b}{N_g} \times 100\% \quad (5)$$

$$P_a = \frac{n_a}{N_d} \times 100\% \quad (6)$$

$$P_b = \frac{n_b}{N_d} \times 100\% \quad (7)$$

P_a is the existed probability or ratio of n_a to N_d . P_b is that of n_b to N_d . P_{va} and P_{vb} are defined as variation peak ratio of n_a and n_b to N_g , respectively.

The relationships among different peaks are listed here.

$$N_d = N_g + n_a + n_b, \quad (8)$$

$$N_a = N_g + n_a \quad (9)$$

$$N_b = N_g + n_b \quad (10)$$

In order to describe the symmetric degree of among n_a and n_b , that is among system a and b , a new parameter—the symmetric degree α is defined as

$$\alpha = \frac{P_b}{P_a} = \frac{P_{vb}}{P_{va}} = \frac{n_b}{n_a}, \quad 0 \leq \alpha \leq 1. \quad \text{When } \alpha=1, \text{ It means perfectly symmetric variation, and}$$

$\alpha=0$, It means extremely asymmetric variation. There are two maximum information

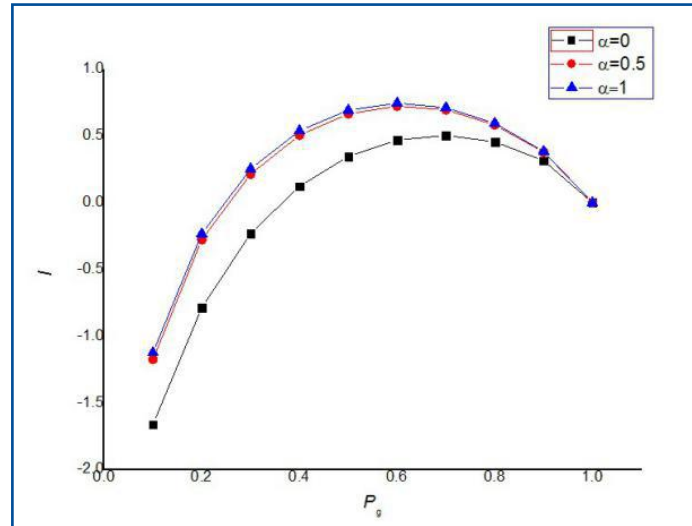


Fig.1 the $I \sim P_g$ curves of dual index information equation

values corresponding to the two states, the $\alpha=1$, $P_g=0.61$, and $\alpha=0$, $P_g=0.70$. These states can be displayed in figure 1.

In accordance with $I \sim P_g$ figure, one can find that the top points of I are at $P_g=0.61=61\%$, when $\alpha=0.5, 1$. While that of I is at $P_g=0.695=0.70=70\%$, when $\alpha=0$. Moreover, there is an flat interval existed near $P_g=0.61$ and 0.70 . In each interval I changes little. P_g corresponding to the maximum information intervals, and the change ratios of information corresponding to these P_g intervals, at different α , are shown in table 1.

Table 1 P_g corresponding to the maximum information intervals, and the change ratios of information corresponding to these P_g intervals at different α

α	P_g corresponding to the maximum information intervals, and the variant ratios of information corresponding to these P_g intervals						Variant ratios of information and their P_g interval (1)	Variant ratios of information and their P_g interval (2)
	P_g							
	0.55	0.61 ^a	0.65	0.66	0.70 ^b	0.75		
0			0.497 ^c	0.500	0.504	0.491	1.34%(0.65~0.70)	2.58%(0.75~0.70)
0.5	0.706	0.724	0.719	0.716			2.49%(0.55~0.61)	1.10%(0.66~0.61)
1	0.731	0.746	0.739	0.735			2.01%(0.55~0.61)	1.47%(0.66~0.61)

a , when symmetric degree $\alpha=0.5, 1$, the maximum information corresponding to $P_g=0.61$; b , when $\alpha=0$, the maximum information is at $P_g=0.70$. c , the information values corresponding to P_g upper.

The theoretical analysis offered above point out, when $\alpha=0.5 \sim 1$, the maximum information are all at $P_g=0.61$. When the P_g interval region from 0.55 to 0.66, the information change little, only within $\pm 3\%$. When $\alpha=0$, the maximum information is

at $P_g=0.70$, and when P_g interval region is from 0.65 to 0.75, the information also varies within $\pm 3\%$. These results indicated when P_g changes within $\pm 5\%$, from P_g corresponding to the maximum information, the I vary within $\pm 3\%$. This region may be considered as the critical characteristic zone of TCM species.

The higher the common peak ratio between samples, the larger the similarity between them. This means their quality is more similar to each other. Therefore, relying on table 1, $P_g=61\%$ and $P_g=70\%$ can be regarded as or defined as two TCM species constants. Depending on the two absolute theory criteria, we can reasonably set practical theory standards $P_g \cong (61 \pm 3)\%$ and $P_g \cong (70 \pm 3)\%$ as two optimizing and judging TCM species, considering all kinds of factors affecting experiment results.

2 material

2.1 Regents and Chinese medicine samples

KBr (AR) were purchased from Tianjing national reagent company (China), and 42 samples belonging to four species of combination Chinese medicines: Guifu Dihuang pills, Jinkui Shenqi Pills, Mingmu Dihuang Pills and Zhibai Dihuang Pills are shown in Table 2.

Table 2 The sources of samples of TCM

Samples	Species	Sources	Date	in	Batches
			produced		
S1	Gf ^a	Hefei shenlu shuangheJiuhua Pharmaceutical Co.,Ltd	2003.04.14		0303493
S2	Gf	Hefei shenlu shuangheJiuhua Pharmaceutical Co.,Ltd	2003.09.01		0308413
S3	Gf	HenanWanxi Pharmacy Co.,Ltd	2003.12.21		031204
S4	Gf	HenanWanxi Pharmacy Co.,Ltd	2004.07.19		040702
S5	Gf	HenanWanxi Pharmacy Co.,Ltd	2005.02.21		050203
S6	Gf	HenanWanxi Pharmacy Co.,Ltd	2006.01.05		060101
S7	Gf	Wuhu Zhang Hengchun business Co., Ltd	2004.08.12		20040802
S8	Gf	Wuhu Zhang Hengchun business Co., Ltd	2004.08.12		20040802
S9	Gf	Anqing HuifengPharmacy Co.,Ltd	2004.05.25		040520
S10	Gf	Suzhou Changjia Pharmacy Co.,Ltd	2005.05.17		0505171
S11	Jk ^b	Beijing Tongrentang technologies	2003.03.11		3030440
S12	Jf	Beijing Tongrentang technologies	2003.03.11		3030440
S13	Jf	Beijing Tongrentang technologies	2004.03.30		4030045
S14	Jf	Beijing Tongrentang technologies	2004.08.02		4032954
S15	Jf	Beijing Tongrentang technologies	2004.08.23		4032972
S16	Jf	Beijing Tongrentang technologies	2005.07.07		5032371
S17	Jf	Beijing Tongrentang technologies	2005.07.07		5032372
S18	Jf	Beijing Tongrentang technologies	2005.08.10		5032879
S19	Mm ^c	Huangsan Tianmu Pharmacy Co.,Ltd	2006-05-26		060526
S20	Mm	Huangsan Tianmu Pharmacy Co.,Ltd	2006-05-26		060526
S21	Mm	Huangsan Tianmu Pharmacy Co.,Ltd	2006-05-26		060526

S22	Mm	Huangsan Tianmu Pharmacy Co.,Ltd	2006-11-27	061127
S23	Mm	Huangsan Tianmu Pharmacy Co.,Ltd	2006-11-27	061127
S24	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-05-29	060507
S25	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-05-29	060507
S26	Mm	HenanWanxi Pharmacy Co.,Ltd	2005-05-14	051002
S27	Mm	HenanWanxi Pharmacy Co.,Ltd	2005-05-14	051002
S28	Mm	HenanWanxi Pharmacy Co.,Ltd	2005-05-14	051002
S29	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-04-15	060403
S30	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-04-15	060403
S31	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-10-27	061005
S32	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-10-27	061005
S33	Mm	HenanWanxi Pharmacy Co.,Ltd	2006-10-27	061005
S34	Zb ^d	Hubei Duanyao Pharmaceutical Co.,Ltd	2005-08-11	050801
S35	Zb	Hubei Duanyao Pharmaceutical Co.,Ltd	2005-11-09	051101
S36	Zb	Hubei Duanyao Pharmaceutical Co.,Ltd	2005-11-09	051101
S37	Zb	Anhui Huatuo ChinesePharmaceutical	2006-01-17	20060101
S38	Zb	Beijing Tongrentang technologies	2006-01-04	6030426
S39	Zb	Beijing Tongrentang technologies	2006-01-04	6030426
S40	Zb	Hubei Qingda Pharmaceutical science and technology Co.,Ltd	2005-10-20	051002
S41	Zb	Hubei Qingda Pharmaceutical science and technology Co.,Ltd	2006-05-09	060501
S42	Zb	Hubei Qingda Pharmaceutical science and technology Co.,Ltd	2006-05-09	060501

a .Gf, Guifu Dhuang Pills, *b* . Jinkui Shenqi Pills, *c* . Mingmu Dihuang Pills, *d* . Zhibai Dihuang Pills.

2.2 Composition of four TCM species [39]

The composition of four species of TCM investigated in this paper are as follows.

Guifu Dhuang Pills: Radix Rehmanniae, Poria, Cortex Moutan, Rhizoma Dioscoreae, Fructus Corni, Rhizoma Alismatis, Radix Aconiti Later alis Preparata, Cortex Cinnamomi.

Jinkui Shenqi Pills:Radix Rehmanniae, Poria, Cortex Moutan, Rhizoma Dioscoreae, Fructus Corni, Rhizoma Alismatis, Radix Aconiti Later alis Preparata, Cortex Cinnamomi, Radix Achyranthis Bidentatae, Plantain seed, Cassia twig.

Mingmu Dihuang Pills: Radix Rehmanniae, Poria, Cortex Moutan, Rhizoma Dioscoreae, Fructus Corni, Rhizoma Alismatis, Radix Paeoniae Alba, Radix Angelicae Sinensis, Fructus Lycii, Fructus Tribuli, Flos Chrysanthemi, Concha Haliotidis.

Zhibai Dihuang Pills:Radix Rehmanniae, Poria, Cortex Moutan, Rhizoma Dioscoreae, Fructus Corni, Rhizoma Alismatis, Cortex Phellodendri, Rhizoma Anemarrhenae.

The list showed all four species of TCM contain common six kinds of herbs, Radix Rehmanniae, Poria, Cortex Moutan, Rhizoma Dioscoreae, Fructus Corni, Rhizoma

Alismatis. Thus these four species of TCM are composed of a great deal of the same or similar compounds, which make it very difficult to accurately identify them or classify them.

3 Results and discussion

3.1 Detection of infrared fingerprint spectra

In keeping with the experiment conditions listed in methods, the infrared fingerprint spectra of 42 samples' powder belonging to four species of traditional Chinese medicines, were measured, showed in figure 1~4. From these figures, one can notice they are very analogous, and are of high complexity. It is difficult to distinguish them by direct observation. Their accurate identification must be carried out depending on mathematical theory to analyze these infrared fingerprint spectra.

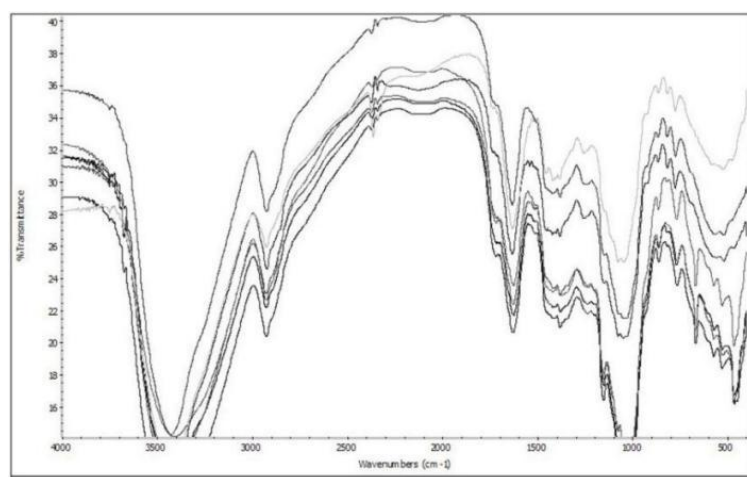


Fig.1 the overlapped infrared fingerprint spectra of Guifu Dihuang pills and Jingkui Shenqi Pills' powders. They were S1,S3,S8,S10(Guifu Dihuang pills), S11,S14,S18 (Jingkui Shenqi Pills) at near 1600cm^{-1} from bottom to top.

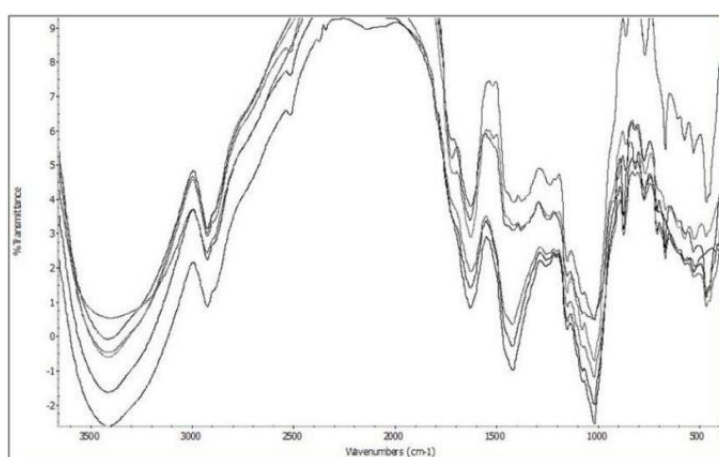


Fig.2 the overlapped infrared fingerprint spectra of Mingmu Dihuang pills and Zhibai Dihuang Pills' powders. They were S24,S26,S32 (Mingmu Dihuang pills), S34, S38, S42 (Zhibai Dihuang Pills) at near 1600cm^{-1} from bottom to top.

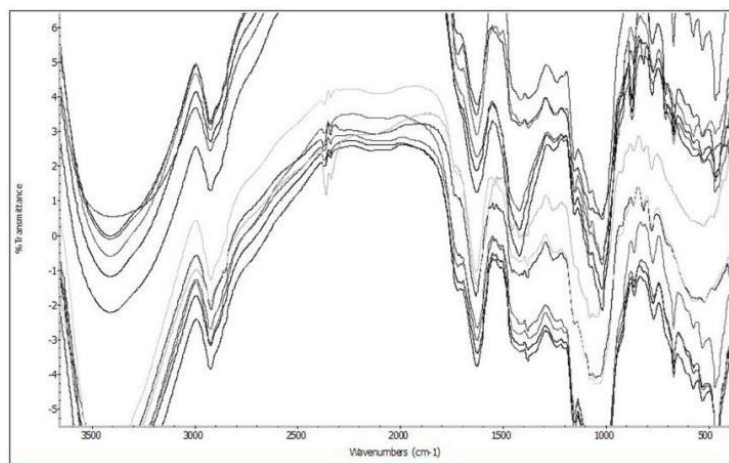


Fig.3 the overlapped infrared fingerprint spectra of Guifu Dihuang pills, Jingkui Shenqi Pills, Mingmu Dihuang pills and Zhibai Dihuang Pills' powders. They were S1, S3,S8,S10(Guifu Dihuang pills), S11,S14,S18(Jingkui Shenqi Pills), S24,S26,S32(Mingmu Dihuang pills),S34, S38, S42(Zhibai Dihuang Pills), at near 1600cm^{-1} from bottom to top.

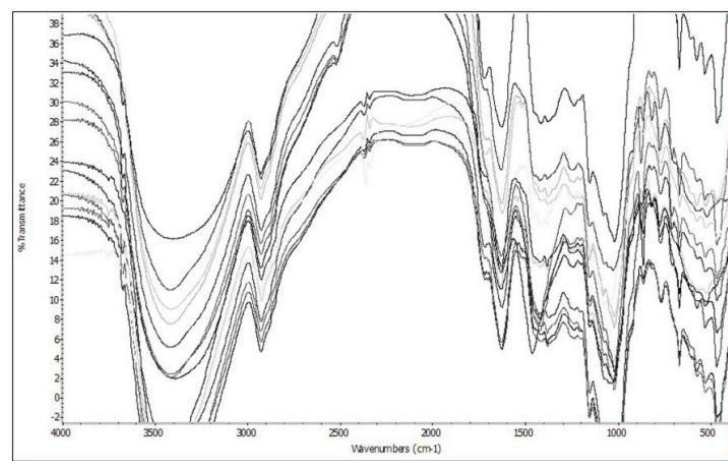


Fig.4 the overlapped infrared fingerprint spectra of Guifu Dihuang pills, Jingkui Shenqi Pills, Mingmu Dihuang pills and Zhibai Dihuang Pills' powders. They were S1, S3, S8, S10 (Guifu Dihuang pills), S11, S14, S18 (Jingkui Shenqi Pills), S19, S24, S26,S32 (Mingmu Dihuang pills),S34, S37, S38, S42 (Zhibai Dihuang Pills), at near 2900cm^{-1} from bottom to top.

3.2 Analysis on data

To determine common and variant peaks of infrared fingerprint spectra of these 42 samples based on the Shapiro-Wilk W-testing method[40]. To select any sample as a reference, and calculate the common peak ratios of the rest samples to the reference. Then to place these rest sample marks, together with their common peak ratios in the order :common peak ratio from high to low to form a sequence.

Based on the two traditional Chinese medicine species constants $P_g = 0.61$ and $P_g = 0.70$, and combining the maximum number of effective samples method in characteristic sequence of sample set [36].

$$Y = \sum_{i=1}^M (N_{ci} - N_{ri}) \quad (11)$$

Y : the maximum effective samples in characteristic sequences of sample set.

N_{ci} : the number of samples in the i th samples' characteristic sequence.

N_{ri} : the number of samples in i th samples' related sequence.

M : the number of total samples.

Y reflects degree of effective classification of sample set. The larger the Y , the more clear the classification, and the much less the samples in related sequences. This means that the more samples in core characteristic sequences, the classification is more ideal.

To optimize classification relying on $P_g \cong (61 \pm 3)\%$ and $P_g \cong (70 \pm 3)\%$, and to get integrally optimized characteristic sequences. Then to carried out the concisely identification and classification of these 42 samples.

When $P_g \cong 61\%$ and $P_g \cong 72\%$, the optimized characteristic sequences of the 42 samples belonging to four TCM species were showed as follows.

Characteristic sequences

Core characteristic sequences	<u>related sequences</u>
Guifu Dihuang Pills(Gf)	
S1: <i>(S1) S2 S3 S4 S5 S6 S7 S8 S9 S10</i> ^a	<u>S34 S35 S36 S37</u> S40 S41 S42 ^b
S2: <i>S1 (S2) S3 S4 S5 S6 S7 S8 S9</i> ^c S10	<u>S34</u> S35 S36 S37 S40 S41 S42
S3: <i>S1 S2 (S3) S4 S5 S6 S7 S8 S9 S10</i>	<u>S37 S40 S41 S42</u>
S4: <i>S1 S2 S3 (S4) S5 S6 S7 S8 S9 S10</i>	<u>S34 S35 S36 S37 S40 S42</u>
S5: <i>S1 S2 S3 S4 (S5) S6 S7 S8 S9 S10</i>	<u>S34</u> S35 S36 S37 S40 S41 S42
S6: <i>S1 S2 S3 S4 S5 (S6) S7 S8 S9</i>	<u>S34 S35 S36 S40 S41 S42</u>
S7: <i>S1 S2 S3 S4 S5 S6 (S7) S8 S9 S10</i>	<u>S34 S35 S36 S40 S41 S42</u>
S8: <i>S1 S2 S3 S4 S5 S6 S7 (S8) S9 S10</i>	<u>S34 S35 S36 S37 S40 S42</u>
S9: <i>S1 S2 S3 S4 S5 S6 S7 S8 (S9) S10</i>	<u>S35</u> S34 S37 S41 S42
S10: <i>S1 S2 S3 S4 S5 S7 S8 S9 (S10)</i>	<u>S37</u>
Jingkui Shenqi Pills(Jk)	
S11: <i>(S11) S12 S13 S14 S15 S16 S17 S18</i>	
S12: <i>S11 (S12) S13 S14 S15 S16 S17 S18</i>	
S13: <i>S11 S12 (S13) S14 S15 S16 S17 S18</i>	
S14: <i>S11 S12 S13 (S14) S15 S16 S17</i>	
S15: <i>S11 S12 S13 S14 (S15) S16 S17 S18</i>	
S16: <i>S11 S12 S13 S14 S15 (S16) S17 S18</i>	
S17: <i>S11 S12 S13 S14 S15 S16 (S17) S18</i>	
S18: <i>S11 S12 S13 S15 S16 S17 (S18)</i>	
Mingmu Dihuang Pills(Mm)	
S19: <i>(S19) S20 S21 S22 S23</i>	<u>S41</u>
S20: <i>S19 (S20) S21 S22 S23</i>	
S21: <i>S19 S20 (S21) S22 S23</i>	

S22: S19 S20 S21 (S22) S23	<u>S40 S41 S42</u>
S23: S19 S20 S21 S22 (S23)	
S24: (S24) S25 S26 S27 S28 S29 S30 S31 S32 S33	
S25: S24 (S25) S26 S27 S28 S29 S30 S31 S32 S33	
S26: S24 S25 (S26) S27 S28 S29 S30 S31 S32 S33	
S27: S24 S25 S26 (S27) S28 S29 S30 S31 S32 S33	
S28: S24 S25 S26 S27 (S28) S29 S30 S31 S32 S33	
S29: S24 S25 S26 S27 S28 (S29) S30 S31 S32 S33	
S30: S24 S25 S26 S27 S28 S29 (S30) S31 S32 S33	
S31: S24 S25 S26 S27 S28 S29 S30 (S31) S32 S33	
S32: S24 S25 S26 S27 S28 S29 S30 S31 (S32) S33	
S33: S24 S25 S26 S27 S28 S29 S30 S31 S32 (S33)	
Zhibai Dihuang(Zb)	
S34: (S34) S35 S36 S37 S40 S41S42	<u>S1 S2 S4 S5 S6 S7 S8 S9</u>
S35: S34 (S35) S37 S40 S41	<u>S1 S2S4 S5S6 S7 S8 S9</u>
S36: S34 (S36) S37 S40 S42	<u>S1 S2 S4 S5 S6 S7 S8</u>
S37: S34S35 S36 (S37)S40	<u>S1S2 S3 S4 S5 S8 S9 S10</u>
S38: (S38) S39	
S39: S38 (S39)	
S40: S34 S35 S36 S37 (S40) S41 S42	<u>S1 S2 S3 S4 S5 S6 S7 S8 S22</u>
S41: S34S35S40 (S41) S42	<u>S1 S2 S3 S5 S6 S7 S9 S19 S22</u>
S42: S34 S36 S40 S41 (S42)	<u>S1S2 S3 S4 S5 S6 S7 S8 S9 S22</u>

a , the core characteristic sequences ; b , the related sequences, samples in this part not belonging to the class to that the samples in core characteristic sequences corresponded. characteristic sequences is made up of core characteristic sequences and related sequences. c , the samples represented by **black-itaics** are the characteristic sequences of 42 samples, when choose $P_g \cong 72\%$. While the whole sequences are the characteristic sequences when $P_g \cong 61\%$.

When $P_g \cong 61\%$, five of the six pairwise, constructed from four TCM species, could be classified accurately according to each samples' characteristic, listed ahead. They were Gf-Jk, Gf-Mm, Jk-Mm, Jk-Zb, Mm-Zb. While there existed heavy overlap between Gf-Zb samples' characteristic sequences exhibited above. This indicated when $P_g \cong 61\%$, Gf and Zb samples could not be classified.

When $P_g \cong 72\%$, for S35, S37, samples in related sequences are more than that in core characteristic sequences. Thus S35, S37 could not be corrected recognition. However, other 40 samples could be identified clearly. The correct ratio was $40/42=95.2\%$. Under this condition, all six pairs were identified and classified perfectly. Based on the TCM species constant $P_g=70\%$, together with the maximum effective sample optimum method[34], to change P_g according to $P_g \cong (70 \pm x)\%$, $x=-1, 0, 1, 2, 5$, the number of maximum effective samples were listed in table 3.

From table 3, when $P_g \cong 69\sim 75\%$, all samples could be distinguished completely. However, when $P_g \cong 72\%$, the number of the effective samples was the largest one. The result was most reasonable. Further, when $P_g \cong 75\%$, the correct ratio was 100%, which is larger than 95.2%, but the number of the effective samples was the lowest,

only was 243. In this case, the over classification was carried out, the results were not reasonable.

Table 3 The number of effective samples based on $P_g \cong (70 \pm x)\%$

$P_g\%$	Optimized results	
	The maximum number of effective samples	Correct recognition ratio (%)
$\cong 69$	249	92.9
$\cong 70$	254	92.9
$\cong 71$	254	92.9
$\cong 72$	256	95.2
$\cong 75$	243	100

x^* , $x = -1, 0, 1, 2, 5$.

Analysis on the data from the IR FPS of 42 sample powders, belonging to four TCM species, showed that the identification criteria of these four TCM species are $P_g \cong 61\%$ and $P_g \cong 75\%$, the results were listed in table 4.

Under the similarity scales showed in table 4, all six pairs among four TCM species were classified correctly. The correct ratio of samples was 95.2%. In this case, the number of effective samples was the largest. The correct ratio of TCM species was 100%.

On the other hand, the IR FPS could be analyzed by means of the dual index grade sequence individualized pattern recognition method. When similarity scale is $P_g \cong \bar{P}_g + 1.3S$, S35 was not classified successfully. S37 could not be recognized because the number of samples in its related sequence is equal to that in its core characteristic sequence. The correct recognition ratio is $40/42 = 95.2\%$, which is the same as that obtained by means of the Two TCM species constants. These results proved that this new method is reliable.

Table 4 test results of the two TCM species constants

The obeyed species constants P_g (%)				
Species	Guifu Dihuang Pills	Jinkui Shenqi Pills	Mingmu Dihuang Pills	Zhibai Dihuang Pills
Guifu Dihuang Pills		61	61	70
Jinkui Shenqi Pills	61		61	61
Mingmu Dihuang Pills	61	61		61
Zhibai Dihuang Pills	70	61	61	

4 conclusion

Depending on the two TCM species constants $P_g=61\%$ and $P_g=70\%$, obtained from the dual index information theory equation, and on the two theoretical standards

$P_g \cong (61 \pm 3)\%$ and $P_g \cong (70 \pm 3)\%$, established rest on the two TCM species constants, 42 samples and the six pairs combined from four TCM species were perfectly identified and classified, together with the maximum number of effective samples method [36]. These results can support conclusion that the two common peak ratios deduced from the dual index information theory equation, corresponding two maximum information states, which related to two extremely variation states, are able to truly reflect the intrinsic characteristic of complex biology systems. This theory could concisely identify, classify some complex biology systems without any help of prior knowledge, such as some learning samples, some parameters decided subjectively about sample set. These work were achieved only relying on the measured data from sample set and the maximum number of effective sample method. In this approach, the maximum number of effective samples enable researchers to avoid over classifying a sample set, and to achieve an optimum pattern recognition result. This verifies that in biology systems, there exist some precisely science laws, and some constants, like in physics and chemistry. These original work will attract other researchers to discover some precisely science laws in biological systems in the future.

Trace back to several thousand years, it may be a great breakthrough by far to accurately identify TCM species, or biology systems by means of absolute quantitatively theoretical criteria, not experience, prior knowledge. Moreover, this theory is of great simplicity, and may be suit to other biology systems in principle.

5 Methods

5.1 Instruments

FT-IR spectrophotometer Model NICOLET-5700-FT-IR(USA), with spectral range: 4000–400 cm^{-1} , resolving power 4 cm^{-1} ; high speed grinder; tablet press; analytical balance were used in this study.

5.2 Conditions

All samples were powdered. To shift the powders with 80 eye sieve, then to heat them at 60°C for 2 hours, and keep them at below 4°C. The experiments show if the powders are kept at low temperature, the IR FPS obtained remain the same and no degradation is observed.

The infrared fingerprint spectra of these samples were measured by means of KBr tableting method. Each sample was detected for 6 times to obtain 6 infrared fingerprint spectra, based on which a combination numerical fingerprint spectrum was formed. The wavenumbers of each peak in the combination numerical fingerprint spectrum were the mean of that of this peak appeared in the 6 infrared fingerprint spectra. These infrared fingerprint spectra were smoothed by 25 cm^{-1} at sensitivity 80.

5.3 Repeatability and stability

The sample S20 was parallelly measured for 6 times and got 6 infrared fingerprint spectra, and choose any four fingerprint spectra to construct a combination numerical

fingerprint spectra, in which every peak's wavenumbers is the average of the peaks appeared in the same position in four fingerprint spectra. The results showed the minimum P_g was 93% of among these combination numerical fingerprint spectra. Thus the fingerprint spectra were of excellent repeatability. The powders of traditional Chinese medicines were kept below 4°C, and were of good stability.

References

- [1] Wang Jinwu. Spermatophyte taxonomy. Beijing: High Education Press, 2011 (second edition).
- [2] Zhou Changfa, Yang Guang. The status and definition of species, Beijing: Science press, 2011.
- [3] Ernst Mayr. The growth of biological thought—diversity, evolution and inheritance, Chengdu: Sichuan education press, 2010.
- [4] Stephen T. Garnett, Les Christidis. Taxonomy anarchy hampers conservation. *Nature*, 2017, 546:25–27.
- [5] Edward O. Wilson. The diversity of life. Published by arrangement with Harvard University Press, 1992.
- [6] Zou Huabin, Yuan Jiurong, Yuan Hao. The historical development of techniques and methods for controlling Chinese medicines. *Chinese Traditional Patent Medicine*, 2003, 25(9): add 4-6.
- [7] Zou Huabin, Yuan Jiurong, Yuan Hao. Study of HPLC-FPS and 42 Dimensional UV-FPS of Peru's Ginseng healthcare product. *Chinese Traditional Patent Medicine*, 2003, 25(4): 88-92.
- [8] Zou Huabin, Yuan Jiurong, Lv Qingtao, Rong Rong. The Dual Index Sequence Analytical Method of Common Peak Ratio and Variant Ratio for Analysing UV Fingerprint Spectra of Radix Glycyrrhizae. *Journal of Chinese Medicinal Materials*, 2003, 26(9): 625-629.
- [9] Zou Huabin, Yuan Jiurong, Du Aiqin, Sun Linlin. Dual-index sequence analytical method for IR fingerprint spectra of the chloroform extract of Radix Glycyrrhizae. *China Journal of Chinese Materia Medica*, 2005, 30(1): 16-20.
- [10] Zou Hua-bin, Yuan Jiu-rong, Du Ai-qin, Sun Lin-lin, Hassan Y. Aboul-Enein. Dual-Index Sequence Analytical Method for IR Fingerprint Spectra of Ethanolic Extract of Various Glycyrrhizae's Root Species components. *Analytical letters*, 2005, 38 (7): 1167 - 1178
- [11] Huabin Zou, Guosheng Yang, Aiqin Du, Jiurong Yuan, Zhengran Qina Yingying Xia, Hassan Y. Aboul-Enein. Combinational Numerical Fingerprint Spectra of Glycyrrhiza and Analysis of Common Peak Ratio Invariableness In HPLC. *Biomed. Chromatogr.* 2006, 20: 642–655.
- [12] Zou Huabin, Yuan hao, Wang aiwu, Yuan Jiurong, Yue Chunhua. The Common and Variation Peak Ratio Dual Index Sequence Analysis on UV Fingerprint Spectra of Paeonia Lactiflora Pall. *Spectroscopy and Spectral Analysis*, 2007, 28(9): 1815.
- [13] Hua-Bin Zou, Guo-Sheng Yang, Bing Liu, Hong Zhai, Jiurong Yuan, Yingying Xia, Hassan Y. Aboul-Enein. Recognition of Radix Paeoniae Alba herbs (Fam. Ranunculaceae) with HPLC Fingerprint Spectra. *Analytical letters*, 2008, 41: 3309-3323, 2008.
- [14] Xu Guofeng, Zhang Lijia, Meng Xiansheng, Cao Aimin. The analysis on ultra violet fingerprint spectra of trichosanthes peel. *Journal of Liaoning college of TCM*, 2006, 8(1): 87-88.
- [15] Ji Xianling, Gai Yingping, Mu Zhimei, Wang Gengxian, Ji Shenglai. Identification of white Muscardin Silkworms by infrared spectroscopy, 2007, 27(1): 66-69.

- [16] Liu Yuan, Meng Qingyan, Peng Lianxin, An Jing, Wu Yuanbo. Dual—index sequence analysis of common and variant peak ratio in UV specific(fingerprint)chromatogram of *Acanthopanax giraldii* Harms. Chin J Pharm Anal, 2007,27(8):1182-1185.
- [17] Hu Aihua, Xing Shiyan, Kong Qiliang. Study on the Identification of *Ginkgo biloba* L. Woods with FTIR. Chinese Agricultural Science Bulletin, 2009,25(04):88-92.
- [18] Cheng Yunqing, Liu Jianfeng, Liu Qiang, Wang Zhanwu, Wang Shufan. Dual-index sequential analysis of common and variant peak ratio in infrared fingerprint of *Rhodiola sachalinensis*. Journal of Nanjing Agricultural University, 2011, 34(5) : 155-158.
- [19] Huang Liping, Wu Jing. Dual-index sequence analysis of common and variant peak ratio in far-infrared fingerprint of *Pyritum*. China Journal of Chinese Materia Medica, 2011,36 (11) : 1441-1444.
- [20] Shi Yun, Tan Xianhe, Chi Yumei, Lai Huadu, Xi Ming. Study on IR Fingerprint Spectra of *Abelmoschus Corolla*. Chinese wild plant Resources, 2012, 31(2):29-32.
- [21] Wang Yan, Wang Bin, Xu Yinfeng, Xu Huanzhi, Su Li, Zheng Bin, Yang Huicheng. Infrared Fingerprint Analysis of Mussel Coupled with Cluster Analysis and Sequential Analysis of Dual-indexes. Journal of Chinese Institute of Food Science and Technology, 2013, 13(1): 178-182.
- [22] Zhou Ye, Zhang Qingwei, Li Peifu, Han Lifeng, He Lilin, Zhang boli. Identification of Chinese traditional medicine *Angelicae Pubescentis Radix* from different places by fourier transform infrared spectrometry. China Journal of Chinese Materia Medica, 2013,38(19): 3309-3312.
- [23] Wan Zhongmin, An kun, Wang Jiufeng, Jiao Weisun, Shao Huawu, Yang Xiaocheng. Dual—Index Sequence Analysis on FTIR Fingerprint of *Pseudostellaria heterophylla*. Nat Prod Res Dev, 2013, 25(5):594-599.
- [24] Zhuang Lin, Huang Qun, Xu Yanhong. Four Kinds of Phoebe and Machilus Wood Infrared Absorption Spectrum Identification. Jour of Fujian Forestry Sci and Tech, 2014,41(4): 21-25.
- [25] Shi Chaojing, Liu Biwang, Li Yani, Xue Huiqing. Comparative study on different genuine producing areas of *Astragalus membranaceus* soil by FTIR. Journal of Shanxi university (Nat.Sci.Ed.), 2015,38 (2) : 335-242.
- [26] Peng zhou, Zhao Yuqin, Wang Bin, Qu Youle. Journal of Anhui Agri.Sci. 2015,43(12) : 32-34.
- [27] Zou Huabin, Han Zhifeng, Zhai Hong, Wei Yingqin, Wang Ying. The First and Second Cluster Analysis of Dual Index Grade Sequence of IR Fingerprint Spectra of Guifudihuang Pill and Jinkuishenqi Pills Samples and Their Quality Evaluation. Spectroscopy and Spectral Analysis, 2007,27(12): 2432-2436.
- [28] Zou Huabin, Zhang Xinling, Zhai Hong, et al. Dual index grade sequence pattern recognition of extracts with ethanol of Mingmu Dihuang pills and Zhibai Dihuang pills. China Journal of Chinese Materia Medica 2008,33(13):1543-1549.
- [29] Dong Fengjuan, Zou Huabin, LU Weijie, Du Aiqin. The dual index grade sequence pattern recognition method for analyzing infrared fingerprint spectra of extracts with the reform of Shiquan Dabu Pills. Journal of Shandong university (Engineering science), 2010, 40(6): 113-119.
- [30] Wang Lishuang, Zou Huabin, Wang Sishun, Wu Can. The high resolution identification and

- quality evaluation of Uyghur compound formula Zukamu. Chinese Traditional Patent Medicine, 2011,33(7):1198-1201.
- [31] Zou Huabin, Wang Lishuang, Zhai Hong. *N*-bio-system dual index grade sequence individualized pattern recognition of H-NMR fingerprint spectra of Zuogui Pills and Yougui Pills and Jinkui Shenqi Pills. Journal of Shandong university (Engineering science), 2012, 42(4): 114-119.
- [32] Zou Huabin, Wang Lingling, Du aiqin, Shi Peimen. The accurate classification of four kinds of compound Danshen medicines and theoretical evaluation of their medicalefficiency. World Chinese medicine , 2015,10(6):904-909.
- [33] Zou Huabin, Dong Fengjuan, Zhang Xinling, Du Aiqin. HPLC fingerprint spectra analysis of Buzhong Yiqi Pills and Shiquan Dabu Pills by the relative peak area difference ΔS grade sequence pattern recognition method. Journal of Shandong university (Engineering science), 2011,41(6): 97-102.
- [34] Zou Huabin, Ayiguzaili. Abulimiti, Zhai Hong. Analysis of EDS Fingerprint Spectra of Mineral Drug Montmorillonite Powder Relying on Dual Index Grade Sequence Individualized Pattern Recognition Method and their Quickly Quality Evaluation. Spectroscopy and Spectral Analysis, 2013,33(12):3321-3325.
- [35] Zou Huabin. Dual index information markedly similar sequence clustering analysis on IR fingerprint spectra of extracts of Guifu Dihuang and Jingui Shenqi pills with ethanol. China Journal of Chinese Materia Medica, 2009,34(18): 2325-2330.
- [36] Zou Huabin. A systematically theoretical distinguish approach for traditional Chinese medicine with identical quality. World Chinese medicine, 2015,10(7):1078-1082.
- [37] Zou Huabin. Fingerprint spectra-based mathematical theory in determining the intrinsic quality grade of biological system of Chinese medicine. World Chinese medicine, 2016,11(9):1876-1881.
- [38] Xu kexue. Biological mathematics. Beijing: Science press, 2002.
- [39] Song Xianmin, Guo Weijia. New edited national Chinese Traditional Patent Medicine. Beijing: people's publish house, 2002.
- [40] Zou Huabin, Yuan Jiurong, Wang Wei. Theoretical identification of common peaks in fingerprint of Chinese medicine -a W- testing and discriminatory method. World science and technology/ modernization of traditional Chinese medicine and materia medica, 2004, 6(2):50-56.