# Two biological constants for accurate classification and evolution pattern analysis of Subgen.*strobus* and subgen. *Pinus*

Huabin zou

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

Correspondence:huabinzou@126.com

**Abstract** Currently, biological classification and determination of different categories are all based on empirical knowledge,which is obtained relying on morphological and molecular characters. For these methods they lacks of absolutely quantitative criteria ground on intrinsically scientific principles. In fact, accurate science classification must depend on the correct description of biology evolution rules.

In this article a new theoretical approach was proposed, in which two characteristic constants were gained from biological common heredity and variation information theory equation, when it is at the maximum information states, corresponding to symmetric and asymmetric variation states. They are common composition ratios, $P_g$ =0.61, and $P_g$ =0.70. By analyzing the common composition ratios of compounds among oleoresins, two pine subgenus:Subgen.*Strobus* (Sweet) Held and Subgen. *Pinus* could be integrated into one class, Genus pinus, excellently, when $P_g$ = 0.61. These two pine subgenus could be classified into two groups clearly,when $P_g$ = 0.70. The results is somewhat different from that achieved by means of classical classification relying on morphological characters. On the other hand, the evolution relationship of two subgenus was analyzed based on characteristic sequences of samples, it indicated that white pine origin from pinus tabuliformis. The two constants should be used as the classification constants of some biological categories of plants.

**Key words** classification, systematics, biological constant, heredity and variation information, evolution pattern, fingerprint spectra, dual index, pinus.

## Introduction

Biological taxonomy,or systemics is a very classical research field, and it appears that all taxonomic problems have been perfectly solved so far. However, until recently, there exist no any strictly scientific definition and examination about biological species, genus, family, order, class, phylum and kingdom. Thus, this makes biological science building upon the beach. Currently, there are some species definitions, such as biological species concept (BSC), phylogenetic species concept (PSC), evolutionary species concept (ESC), genetic species concept (GSC)， recognition species concept (RSC)， cohesion species concept (CSC) [1,2,3,4,5,6]. All these species concepts are proposed based on empirical knowledge and subjective inference. In fact the definitions of species and other categories are very fundamental for the research on biological diversity, and are the ultimate goal of systemics. Further more, the scientific definition of species is the crown and the holy grail[7]. In reality, very

complex biology phenomenons and behaviors make it hard to put forward a perfect and unified species definition for people. This also leads to the difficulty in biological conservation in practice[8] nowadays.

So far, for the definition of biological species, it lacks support of elemental principle theory based on experimental science and mathematical principles, and lacks theoretically quantitative criteria deduced from these principles. In practice the identification and classification of plants totally depend on a lot of morphological characters[6]. Even in modern numerical taxonomy, the identification standards are also from empirical knowledge, learning/training samples. So, the development of biology is deeply blocked by empirical classification. For these reasons biology is not a rigorous science. In recent classification mathematics, the major methods are some similarity and difference coefficients [9,10], which are short of deterministic criteria for judging species too. There are many pattern recognition methods constructed grounded on statistics, such as systematical cluster, Principal component analysis (PCA)[9,10]. these approaches are able to group organisms into different classes without any prior knowledge related to themselves. But there is no objective or any unsuspected standard for different classes. Some supervised methods, such as Bayesian [6,11] and Fisher determination [9],are certainly able to classify samples, but they must gain their standards relying on learning samples, which only fit to limited sample sets. The maximum likelihood(ML) [4,6,11] classify samples depending on some unproved hypothesis established on experience. It can not make the conclusions undoubtedly. Biological classification science still lacks of precise science principle which likes that existed in physics and chemistry.

It is necessary to search for absolutely quantitative standards, rather than relative criteria and subjective inference standards established by means of empirical knowledge, for classifying different categories and determining accurate evolution relationship among different organisms. Otherwise, one can not discover the true evolution rules of them. The uncertainty in biological classification brings about difficulty to build up a systematically biological science, and influences deep research on biology seriously. The present state of biology is similar to that of chemistry in the early of 19 century, when thermodynamics was not built up, and periodic table of elements was not be discovered, that is, chemistry theory was an empirical system in that time.

The key feature of principle theory is that it can predict matter's intrinsic properties accurately and quantitatively. However, the strait of theory based on statistical description can only offer probability. Author Zou found that the two characteristic constants derived from biological inheritance and variation information theory equation, that is dual index information theory equation could accurately identify combination Chinese medicines[12,13,14,15], which are very complex biology systems. That is, the two constants are qualified to be the absolutely quantitative criteria of traditional Chinese medicines (TCM), which consist of extracts of several kinds of herbs. So far there is no report whether the two constants are suited to be as the classification criteria of different organisms, and fit to distinguish biological categories, such as species, genus, subgenus, family and so on.

In generally , a certain and undoubted classification system, accurate definition of species, genus and family and so forth are one of the ultimate aims of biological science[7].

In the respect of classifying plants of pinus, the major methods are based on morphological or macro characters[6]. In modern taxonomic field, the classification of pinus' plants is also grounded on macro characters. That is ,the number of vascular bundle within needles. They are divided into two subgenus, haploxylon with unidimensional bundle, and diploxylon with double vascular bundles[16,17]. In 1966, *P. krempfii* Lecomte was discovered, and it was one new subgenus[18], thus pinus includes three subgenus.

Plants of Pinus originated from China are classified into two subgenus, haploxylon and diploxylon[19]also, according to the number of bundles in a needle. Song Zhenqian divided pinus, growing in China, into three subgens too, that is, subgen. *Strobus*(Sweet)Rehd, subgen.*Pinus* and subgen, Parrya (Mayr) Chu et X.P.Li [20].For the third subgenus, it merely contains two kinds of pines.

On the other hand, a series of researches have revealed that the compositions of rosins of different pines can fully reflect their genetic traits,hereditary characters.

According to literature[20], Herty C.H. (1908) once studied *P.elliottii* Englem. and *P. palustris* Mill.. Dupont G and Barrand M (1925) investigated *P.nigra*, Krestinsky V et al (1932) researched *P. sylvestris*. They also found that for the same kinds of pines, the compositions in rosins varied little. Oudin A(1938) studied *P. pinaster* and obtained the same conclusions. For the same pine trees their compositions of rosins varied lightly in different seasons in an year. The study indicated the collecting rosin methods and ages of pine trees did not influence the compositions[21].

Mirov N.T. analyzed the compositions of 77 kinds of rosins corresponding to 77 kinds of pines, and found that they could be used for identifying pines[22]. He also researched the extracts from different kinds of pines.Their extracts differed from each other obviously, and could be used in chemical classification (chemotaxonomy)[23]. In the early of 1960s, chemotaxonomy was put forward[24,25,26], and this method was applied to classifying pines.

Song Zhenqian studied the oleoresins of *P.koraiensis* Sieb.et Zucc (red pines) grew in Dunhua district of Jilin province in China and in the area of Russa in the European, He found that the compositions of oleoresins from the two areas were highly identical, even though these oleoresin samples produced apart from thousands miles away. The compositions of the oleoresins of Zhan pine originated from Fuyuan district of Zhejiang province in China and from the kamaka mountain in Peru were highly similar to each other[20].Thus ,the compositions in oleoresins of different pines can provide us with plentifully molecular characters for classifying pines. Generally speaking, compared to genes, substances in metabolites enable to reflect the live physiology states of organisms more perfectly, since genes representation are regulated by many factors. Thus, gene information can not reflect the physiological states lively.

In august of 1988 and 1989, for each kind of pine, 5 to10 oleoresin samples from the 5 to 10 pine trees were collected, randomly from its distribution center by means

of the same manner. Totally hundreds of oleoresins samples from 50 trees belong to 24 kinds of pines were studied systematically. He designed a method for Gas chromatography (GC) analysis without advanced separation, all compositions in oleoresins were analyzed successfully.

In terms of classical classification results of pines,Song analyzed oleoresins of 24 kind of pine samples Subgen.*Strobus* and Subgen.*Pinus* originated in China, and performed the chemical classification of these pines. However, there is no Hierachical Cluster Analysis and pattern recognition being carried out by means of mathematical method so far.

For matter, its constructs determine its functions. For this reason the reliable information for classifying samples should be the material structure information, which pose more rigorous independent characters. Compared to morphological characters, molecular construct characters ought to well represent species and evolution traits of biology.

In mathematical classification of biology, whether there exist some strictly certainty principle theory, which can give us certain classification results. In this article grounded on the biological heredity and variation information theory equation proposed by author Zou[12], two characteristic constants, corresponding to symmetric and asymmetric variation states were obtained, when the information values are at the maximum. These constants could be used as the absolute criteria for classifying complex biology systems TCM based on chemical fingerprint spectra[12-15], and the two constants were defined as the species constants of TCM too[15].In fact, the compositions of pine oleoresins can be viewed as chemical fingerprint spectra of compounds. So, we can utilize the some theory treating fingerprint spectra to analyze compound information in them.

In this paper, 24kind of pine samples belonging to Subgen.*Strobus* and Subgen.*Pinus* were quantitatively classified perfectly found on the two constants and the chemical compositions of their oleoresins. The results indicated that the new results both comply with that of classical classification, again exist obvious difference compared with that of classical classification. When using $P_g$ =61% as the theoretical standard, these 24 samples were clustered into one class/group,that is , Pinus. While when using $P_g$ = 70% as the theoretical standard, these 24 samples were divided into two Subgenus, that is, Subgen.*Strobus* and Subgen.*Pinus*. However, the results also showed that 3 samples out of 9 Subgen.*Strobus* (in terms of classically morphological classification), belong to Subgen.*Pinus* relying on the new method. Thus among these 24 samples, 18 of them belong to Subgen.*Pinus*, 6 of them were Subgen.*Strobus*. Moreover, by means of the characteristic sequences of these samples, the evolution relationship between the two subgenus could be revealed. The conclusion is Subgen.*Strobus* originated from Subgen.*Pinus*. On the other hand, the extremely large differences in geography and weather, can cause greater variation in chemical compositions of oleoresins of pines in the same Subgenus. This research uncovered that the two constants are qualified to be as the biological constants for

Subgen.*Strobus* and Subgen.*Pinus*. Moreover, whether these two constants are suited for classifying other plants,and even used as category constants, such as species, genus, subgenus, family and so on.This is worth conducting much more researches.

## 2 Material

### 2.1 Oleoresin samples of Subgen.*Strobus* and Sugen.*Pinus*

The chemical composition characteristics come from oleoresin samples of 24 kinds of pines. These 24 oleoresin samples were listed in table1.

**Table 1 Oleoresin resources of 24 samples of Subgen.*Strobus* and Sugen.*Pinus*[a]**

| Sample | Latin name | Sampling location in China |
|---|---|---|
| | **Subgen.*Strobus* (Sweet)Rehd.** | |
| S1 | *P.koraiensis* Sieb.etZucc | Dunhua Luotuo mountain in Jilin Province |
| S2 | *P.sibirica*(Loud.)Mayr | Altay of Xinjiang |
| S3 | *P.armandi* Franch. | Shanxi Zhongtiao mountains |
| S4 | *P.dabeshanensis* Cheng et Law | Anhui Dabie mountain |
| S5 | *P.fenzeliana* Hand.-Mzt. | JianfengLing of Hainan Island |
| S6 | *P.griffithii* McClelland | Liuku of Yunnan Province |
| S7 | *P.kwangtungensis* Chun ex Tsiang | Mang mountain of Hunan Province |
| S8 | *P.wangii* Hu et Cheng | Xichou of Yunnan Province |
| S9 | *P.Strobus* L. | Fuyang of Zhejiang Province |
| | **Subgen.*Pinus*** | |
| S10 | *P.densata* Mast. | Lijiang of Yunnan Province |
| S11 | *P.densiflora* Sieb. et Zucc. | Dalian of Liaoning Province |
| S12 | *P. kesiya* Royle ex Gord.var. *Langbianensis* (A.Chev.) Gaussen | Simao of Yunnan Province |
| S13 | *P. latteri* Mason | Bawangling in Hainan Island |
| S14 | *P. massoniana* Lamb. | Zhanjiang of Guangdong Province |
| S15 | *P.massoniana* Lamb.var.*hainanensis* Cheng et L.K.Fu | Bawangling in Hainan Island |
| S16 | *P.sylvestris* L. var. *liangnanesis* Hort. | Tahe of Heilongjiang Province |
| S17 | *P.sylvestriformis* (Takenouchi) T.Wang ex Cheng | Baihe of Jiling Province |
| S18 | *P. tabulaeformis* Carr. | Shanxi Zhongtiao mountains |
| S19 | *P. taiwanensis* Hayata | Huang mountain of Anhui Province |
| S20 | *P. taiwanensis* Hayata var. *damingshanensis* Cheng et L.K.Fu | Daming mountain of Guangxi Province |
| S21 | *P. takahasii* Nakai | Mi mountain of Heilongjiang province |
| S22 | *P. wulinensis* C.J.Qi et Q.Z.Ling | Zhangjiajie of Hunan Province |
| S23 | *P. yunnanensis* Franch. | Yongren of Yunnan Province |
| S24 | *P. roxbourghii* Sarg. | Jilong of Xizang |

*a*. this table was after from literature [20],p18-21.

### 2.2 Compositions of oleoresins

From[20],the oleoresin compositions of 24 kinds of samples were obtained by means of methods in section 6, seen table 2-1,2-2,2-3 below.

**Table 2-1 Compositions of oleoresins[a]**

| Sample | The codes of compounds in oleoresins 1 to 19[b] | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** |
| **S1** | 1 | 2 | | 4 | 5 | | 7 | 8 | 9 | 10 | 11 | | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| **S2** | 1 | 2 | | 4 | 5 | | 7 | 8 | 9 | 10 | 11 | | | | | | | 18 | 19 |
| **S3** | 1 | 2 | | 4 | 5 | 6 | | 8 | | 10 | | 12 | 13 | | 15 | 16 | 17 | 18 | 19 |
| **S4** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | 11 | | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| **S5** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | | | 13 | 14 | | | 17 | 18 | 19 |
| **S6** | 1 | 2 | | 4 | 5 | 6 | 7 | 8 | | 10 | 11 | | | 14 | | | | | 19 |
| **S7** | 1 | | | 4 | 5 | | 7 | 8 | | | 11 | | | | | | | | 19 |
| **S8** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | 11 | | | | | | 17 | 18 | 19 |
| **S9** | 1 | 2 | 3 | 4 | 5 | 6 | | 8 | | 10 | | 12 | | | | | | | 19 |
| **S10** | 1 | 2 | 3 | 4 | 5 | | 7 | 8 | 9 | 10 | | | 13 | | | 16 | 17 | 18 | 19 |
| **S11** | 1 | 2 | | 4 | 5 | | | 8 | 9 | | | | | | | 16 | 17 | 18 | 19 |
| **S12** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S13** | 1 | 2 | | 4 | 5 | | 7 | 8 | | 10 | | | | | | | 17 | | |
| **S14** | 1 | 2 | | 4 | 5 | | 7 | 8 | | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S15** | 1 | 2 | 3 | 4 | 5 | | 7 | 8 | | 10 | | | 13 | | | 16 | 17 | 18 | 19 |
| **S16** | 1 | 2 | 3 | 4 | 5 | | 7 | 8 | 9 | 10 | | | 13 | | | | | 18 | |
| **S17** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S18** | 1 | 2 | 3 | 4 | 5 | 6 | | 8 | | 10 | | | 13 | | | 16 | | 18 | 19 |
| **S19** | 1 | 2 | | 4 | 5 | 6 | | 8 | | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S20** | 1 | 2 | | 4 | 5 | 6 | | 8 | | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S21** | 1 | 2 | | 4 | 5 | 6 | | 8 | 9 | 10 | | | | | | 16 | 17 | 18 | 19 |
| **S22** | 1 | 2 | | 4 | 5 | 6 | | 8 | 9 | 10 | | | | | | 16 | 17 | | |
| **S23** | 1 | 2 | | 4 | 5 | | | 8 | | 10 | | | 13 | | | 16 | 17 | 18 | 19 |
| **S24** | 1 | 2 | | 4 | 5 | | | 8 | | | | 12 | 13 | 14 | | | | 18 | |

**Table 2-2 Compositions of oleoresins**

| Sample | The codes of compounds in oleoresins 20 to 38 | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **36** | **37** | **38** |
| **S1** | 20 | 21 | | 23 | 24 | 25 | | 27 | | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| **S2** | 20 | 21 | | 23 | 24 | 25 | 26 | 27 | 28 | | 30 | 31 | 32 | 33 | | | 36 | 37 | 38 |
| **S3** | 20 | 21 | | | | | 26 | | 28 | 29 | 30 | 31 | 32 | | 34 | 35 | 36 | 37 | 38 |
| **S4** | 20 | 21 | 22 | 23 | 24 | 25 | | 27 | | | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| **S5** | 20 | 21 | | 23 | 24 | | 26 | 27 | | | 30 | | | | 34 | 35 | 36 | 37 | 38 |
| **S6** | 20 | 21 | | | | | | | | 29 | | 31 | 32 | | 34 | 35 | 36 | 37 | 38 |
| **S7** | 20 | | | | | | | | | 29 | 30 | | | 33 | 34 | 35 | 36 | 37 | 38 |
| **S8** | | 21 | | 23 | 24 | | 26 | 27 | | | | | 32 | | 34 | 35 | 36 | 37 | 38 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S9** | 20 | | | 23 | 24 | | 26 | | | 29 | 30 | | 32 | | 34 | 35 | 36 | 37 | 38 |
| **S10** | | 21 | | | | | | | | 29 | 30 | 31 | 32 | | 34 | 35 | 36 | 37 | 38 |
| **S11** | | 21 | | | | 25 | | | 28 | | | 31 | 32 | | | 35 | 36 | 37 | 38 |
| **S12** | | 21 | 22 | 23 | 24 | | | 27 | 28 | 29 | | 31 | 32 | | | 35 | 36 | 37 | 38 |
| **S13** | | 21 | | | | | | | | | 30 | | | | | | 36 | 37 | 38 |
| **S14** | 20 | 21 | 22 | 23 | 24 | 25 | 26 | | 28 | 29 | 30 | 31 | | | | 35 | 36 | 37 | 38 |
| **S15** | 20 | 21 | 22 | 23 | 24 | | | 27 | 28 | 29 | 30 | 31 | | | | 35 | 36 | 37 | 38 |
| **S16** | | | | 23 | 24 | | | | | | | | | | 34 | 35 | 36 | 37 | 38 |
| **S17** | 20 | 21 | 22 | 23 | 24 | | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| **S18** | 20 | 21 | | | | | | | | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
| **S19** | 20 | 21 | 22 | 23 | 24 | | | 27 | 28 | 29 | | 31 | | | | 35 | 36 | 37 | 38 |
| **S20** | 20 | 21 | | 23 | | | | | | 29 | 30 | | | | 34 | 35 | 36 | 37 | 38 |
| **S21** | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | | 29 | 30 | 31 | | 33 | 34 | 35 | 36 | 37 | 38 |
| **S22** | | | | | 24 | | | | 28 | | 30 | 31 | | 33 | 34 | 35 | 36 | 37 | 38 |
| **S23** | | 21 | | 23 | | | 26 | 27 | | 29 | | 31 | 32 | | 34 | 35 | 36 | 37 | 38 |
| **S24** | | | | | | | | 27 | 28 | | | 31 | | 33 | | 35 | 36 | 37 | 38 |

## Table 2-3 Compositions of oleoresins

| **Sample** | **The codes of compounds in oleoresins 39 to 56** | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 |
| **S1** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | | 49 | 50 | | | | | | | |
| **S2** | 39 | | 41 | 42 | 43 | 44 | 45 | | 47 | | 49 | 50 | 51 | | | | | | |
| **S3** | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | | |
| **S4** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | | |
| **S5** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | | | | | | | |
| **S6** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | | 49 | 50 | 51 | | | | | | |
| **S7** | 39 | | 41 | 42 | 43 | 44 | 45 | | 47 | 48 | 49 | 50 | 51 | | | | | | |
| **S8** | 39 | | 41 | 42 | 43 | 44 | 45 | | 47 | | 49 | 50 | 51 | | | | | | |
| **S9** | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | | | | | | | |
| **S10** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | 53 | | | 56 | |
| **S11** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | 56 | |
| **S12** | 39 | | 41 | 42 | 43 | | 45 | | 47 | 48 | 49 | 50 | 51 | | | | | 56 | |
| **S13** | 39 | | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | 54 | 55 | | |
| **S14** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | 54 | | 56 | |
| **S15** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | 56 | |
| **S16** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | 56 | |
| **S17** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | | 56 | |
| **S18** | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | | 54 | | 56 | |
| **S19** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | 54 | | 56 | |
| **S20** | 39 | | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | 54 | | 56 | |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S21** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | | | 56 |
| **S22** | 39 | | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | | 54 | 56 |
| **S23** | 39 | 40 | 41 | 42 | 43 | | 45 | 46 | 47 | 48 | 49 | 50 | 51 | | 53 | | |
| **S24** | 39 | 40 | 41 | 42 | 43 | 44 | 45 | | 47 | 48 | 49 | 50 | 51 | 52 | | 54 | |

*a*. the compositions of oleoresins of Subgen.*Strobus* and Sugen.*Pinus* samples were after from [20], p123-127, 131-136. *b*. the codes of these compounds in oleoresins were listed as follows.

1,α-pinene；2，Camphene；3，β-phellandrene；4，β-pinene；5，α-Myrcene；6，β-Myrcene；7, 3-Carene；8, Limonene；9，γ-Terpinene；10,α-Terpinene；11，Undecane；12,1-Borneol；13，α-Terpineol；14, Tridecane；15,α-Cubebene；16,β-copaene；17, Isolongifolene；18,longifolene；19,Caryophyllene；20,α-Humulene；21，trans-β-Farnesene；22,β-Caryophyllene；23,β-Guaien；24,γ-Elemene;25, β-Bisabolene；26, Aromadendrene；27,β-Elemene28,Cedrol；29, 8,13-Pimarene；30，Isokaurene；31，8,13-Abietadiene；32, Artis-15-ene；33,Cembrene；34,Pimaral；35，Sandaracopimaral；36，8,15-Isopimaric acid；37，pimaric acid；38，Communic acid；39，Sandaracopimaric acid；40，Pimarol；41，Isopimaric acid；42，Palustric acid；43，Levopimaric acid；44，Lambertianic acid；45，Dehydroabietic acid；46，8,12-Abietadienoic acid,47，Abietic acid；48, 6,8,11,13-Abietatetraenoic acid；49，Neoabietic acid；50, 7,13,15-Abietatrienoic acid；51,7α-Hydroxydehydroabietic acid;52, *p*-Allylanisol ； 53, Isopimaric acid ； 54, 8,13(15)-Abietadienoic acid；55, Mercusic acid，56,α-Cedrene。

## 3 Dual index sequence analysis of oleoresin samples

### 3.1 Constructing dual index sequence of different oleoresins

To apply dual index sequence analytical method [27-33], and to carry out the analysis based on oleoresin compositions. This method is briefly elucidated as follows.To calculate common composition ratios and variation ratios of the 24 oleoresins. Firstly, to select one sample being as the standard or reference, secondly, to calculate common composition ratios and variation ratios of all others to this reference. Then to arrange these samples in terms of the order that common composition ratio is from high to low value. For each sample by this way, one binary sequence with sample symbols, common composition ratios and variation ratios can be achieved. That is dual index sequences. The dual index sequences of 24 oleoresin samples were showed in **supplementary 1**.

### 3.2 Analysis on dual index sequences

#### 3.2.1 two constants from biological heredity and variation information theory

This biological common heredity and variation information theoretical equation[12], so-called dual index information theoretical equation, is displayed below.

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

By means of this equation,we are able to calculate the mutual-action information value between any two organisms or any two evolution stages of a biological system.

The definitions of all variables and coefficients are showed there,according to literature[27-33].

$$P_{\mathrm{g}} = \frac{N_g}{N_g + n_a + n_b} \times 100\% = \frac{N_g}{N_d} \times 100\% \qquad (2)$$

Common composition ratio $P_{\mathrm{g}}$:the ratio of common compositions $N_g$ existed in any two oleoresin samples $a,b$ to the independent compositions $N_d$ in the $a,b$ .The number of $N_d$ is equal to the kinds of different compounds in both sample $a,b$ . $P_{\mathrm{g}}$ can be briefly expressed as $P$. This index $P_{\mathrm{g}}$ is the same as the Jaccard and Sneath, Sokal coefficients [9] intrinsically. $n_a$ , $n_b$ are the variation compositions in oleoresin sample $a,b$, respectively.

Other parameters are presented as follows.

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

$$P_{\mathrm{vb}} = \frac{n_b}{N_g} \times 100\% \qquad (4)$$

$$P_a = \frac{n_a}{N_d} \times 100\% \qquad (5)$$

$$P_b = \frac{n_b}{N_d} \times 100\% \qquad (6)$$

$P_{va}$ and $P_{vb}$ are the variation composition ratios of sample $a,b$, respectively. $P_a$ and $P_b$ are the ratios of $n_a$ and $n_b$ to $N_d$, respectively. They means the existed probabilities of $n_a$ and $n_b$. More importantly, these parameters $P_{\mathrm{g}}$, $P_{va}$ , $P_{vb}$ , $P_a$ and $P_b$ are all established based on variables measured in experiments, without any coefficient determined by empirical knowledge. In fact, this information can be represented as a function $I_b = f(n_a, n_b, N_g)$.

The relationship among these variables and parameters are elucidated below.

$$N_{\mathrm{d}} = N_{\mathrm{g}} + n_{\mathrm{a}} + n_{\mathrm{b}}, \quad N_A = N_{\mathrm{g}} + n_{\mathrm{a}}, \quad N_B = N_{\mathrm{g}} + n_{\mathrm{b}} \qquad (7)$$

$N_A$ and $N_B$ are the number of compounds in sample $a,b$, respectively. The relationships of these variables may also be represented simply as follows.

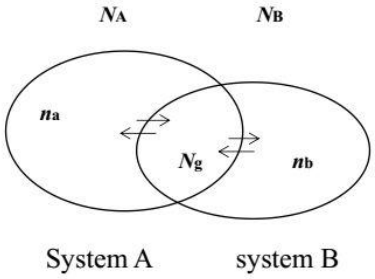


**Fig.1 the relationships among different variables**

### 3.2.2 To definite degree of symmetry

In order to research the symmetry and asymmetry of variation among any two biological systems,a new parameter was defined as $\alpha$.

$\alpha = \frac{P_b}{P_a}$, $0 \leqq \alpha \leqq 1$. When $\alpha = 1$, it shows two samples $a,b$ are in symmetric variation state. When $\alpha = 0$, it expresses $a,b$ are in absolute asymmetric variation state. When $0 \leqq \alpha \leqq 1$, it represents $a,b$ are in different asymmetric variation states.

Depending on the two states , which are symmetric variation $n_a = n_b$, $\alpha = 1$ ,and asymmetric variation $n_a \neq 0$, $n_b = 0$, $\alpha = 0$, with the maximum information values, two common composition ratios $P_g$=0.610 and $P_g$=0.695 can be achieved, respectively. Most interestingly, $P_g$= 0.610 is very closed to gold ratio 0.618.

These two characteristic constants, corresponding to the two states $\alpha = 1$ ,and $\alpha = 0$ , combing with the maximum number of effective sample method[13], are employed to analyze the 24 oleoresin samples based on their dual index sequences as in[12-15], the results were showed in section 3.2.2 and 3.2.3.

### 3.2.3 Classification results relying on characteristic constant $P_g$ = 0.61 (61%).

The construction of characteristic sequences of samples is represented as follows.

**Characteristic sequence = core sequence + <u>related sequence</u>**

**S1**:S1[a] S2 S3 S4 S5 S6 S8 S10 S12 S14 S15 S17 S18 S19 S20 S21S23
**S2**:S1 S2 S4 S5 S6 S8 S12 S14 S15 S17 S19 S21
**S3**:S1 S3 S4 S5 S6 S8 S9 S10 S11 S12 S14 S15 S17 S18 S19 S20 S21 S22 S23
**S4**:S1 S2 S3 S4 S5 S6 S8 S10 S11 S12 S14 S15 S17 S18 S19 S20 S21 S22 S23

**S5**:S1 S2 S4 S5 S8 S9 S3 S12 S14 S15 S16 S17 S19 S20 S21 S23
**S6**:S1 S2 S3 S4 S6 S7 S8 S9 S10 S18 S19 S20 S23
**S7**:S6 S7 S9 S18 S20
**S8**:S1 S2 S3 S4 S5 S6 S8 S9 S12 S17 S19 S20 S21 S23
**S9**:S3 S5 S7 S6 S8 S9 S10 S14 S15 S16 S17 S18 S19 S20 S21S23
**S10**:S1 S3 S4 S6 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S11**:S3 S4 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S12**:S1 S2 S3 S4 S5 S8 S10 S11 S12 S14 S15 S17 S18 S19 S20 S21 S22 S23
**S13**:S10 S13 S14 S20 S22
**S14**:S1 S2 S3 S4 S5 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S15**:S1 S2 S3 S4 S5 S9 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S16**:S5 S9 S10 S11 S14 S15 S16 S17 S18 S20 S21 S22 S23
**S17**:S1 S2 S3 S4 S5 S8 S9 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S18**:S1 S3 S4 S6 S7 S9 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24
**S19**:S1 S2 S3 S4 S5 S6 S8 S9 S10 S11 S12 S14 S15 S17 S18 S19 S20 S21 S22 S23
**S20**:S1 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S21**:S1 S2 S3 S4 S5 S8 S9 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23
**S22**:S4 S3 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S24
**S23**:S1 S3 S4 S5 S6 S8 S9 S10 S11 S12 S14 S15 S16 S17 S18 S19 S20 S21 S23
**S24**:S18 S22 S24

*a*. S1:S1$^{a}$S2---,S1 belongs to its most similar sample, each sample belongs to its own most similar sample too.

Among these 24 samples, according to classically morphological classification results, S1~S9 are Subgen. *Strobus* pines, S10~S24 are Subgen. *Pinus* pines, which are all originated from China. The results based on $P_g$ = 0.61= 61% were optimized. Samples in their characteristic sequences all belong to one group/set, there was no sample in related sequences. All samples in their core sequences were limited to one class, that is *Pinus*.

In accordance with characteristic sequence of each sample, S13 and S24 were of very short sequence, and differ from others greatly, although the 24 pines are all in one class. This may be because origin districts of S13 and S24 are unique. S13 is *P.latteri* Mason, it distributes in the district Hainan province of China, south of Guangdong province and south of Guangxi province in China, and indo-China penisula, the Malay penisula and the Philippines. These districts belong to tropical and subtropical areas, which are different from that of other pines obviously.

S24 belongs to Pinus roxburghii pine,one can know that it belongs to Subgen.*Pinus* by samples in its characteristic sequence. However, there were distinct differences compared to the characteristic sequences of other samples. This may result from that its distribution area is of great differences related to that of other samples. S24 originated from Jilong district in south of Tibet Plateau, where is at an altitude of 2 100 to 2 200 meters. The distribution center is at the southern slop of Himalaya mountain. On the other hand, it also originated from the kingdom of Bhutan, the Sikkim district, India, Nipper, even from Afghanistan. S24 come from Jilong, which is

of significant differences compared to origin districts of other samples. This may leads to vast difference in compositions compared with other samples. So does the S13.

The results showed ahead expressed that $P_g \geqq (61\pm3)\%$ could be employed as the absolute theory standard [14,15]to determine pines of Pinus in practice. This verified that common heredity and variation information theory equation fit to describe biological evolution rules.

### 3.2.4 Classification results based on characteristic constant $P_g$= 0.70

In order to distinguish Subgen. *Strobus* from Subgen. *Pinus*, $P_g$ =70% was accepted as the absolute theory standard, and practical criterion $P_g \geqq (70\pm3)\%$ [14,15] was used to be classification standard, combing the maximum number of effective sample method[13]. By practical analysis, the optimized results were obtained at common composition ratio $P_g$ = 69%,and these results were listed in table 3.

**Table 3 The number of effective samples based on $P_g \geqq (70 \pm x$ [a]$)\%$**

| $P_g$% | Optimized results | | |
|---|---|---|---|
| | The maximum number of effective samples[b] | Results of classification | |
| | | number of white pines[c] | number of Chines pines[d] |
| ≧68 | 145 | 6 | 15+3[e] |
| ≧69 | 162 | 6 | 15+3 |
| ≧70 | 169 | 2 | 15+7 |

*a*. $x$ =-2,-1,0. *b*.the optimum results, when the number of effective sample was the maximum. *c*. White pines, that is spurce, belong to Subgen. *Strobus*, *d*. Chinese pines, that is Pinus tabuliformis, belong to Subgen. *Pinus*. *e*. In 15+3, 15 means the number of Chinese pines and 3 means the number of spurce according to the classical classification results.

According to table 3, although the number of effective samples was the maximun, when $P_g \geqq 70\%$ , the results were not correct, since there were only 2 samples belonging to Subgen. *Strobus*, 22 samples were Subgen. *Pinus*. This case is overclassification.

Combining with the maximum number of effective sample method[13],

$$Y = \sum_{i=1}^{M} (N_{ci} - N_{ri}) \qquad (8)$$

$Y$ : the maximum number of 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 sequences, the classification is more ideal.The number of the most ideally effective samples is $Y_{idea} = \sum_{i=1}^{n} N_i^2$ , in which $N_i$ is the number of samples in $i$ th class, $n$ is the number of classes. In this case, the $N_{ci}$ is equal to that of samples in a class,and $N_{ri}$ =0, tending to ideal classification. For instance, in this study, the number of samples belonging to Subgen.*Strobus* is 6, and that of samples belonging to Subgen.*Pinus* is 18, then the $Y_{idea} = 6^2 + 18^2 = 360$. This method is an excellent one for quantitatively judging whether the result is pros and cons. At the same time, a big advantage for this approach is that it can avoid over classification, compared to other pattern recognition methods,and assure the result is objective classification compared with other methods. On the other hand, in this research the $Y$ is obviously lower than $Y_{idea}$ , it indicated there are great variation among samples. Depending on the compounds in oleoresins, classification results of these 24 samples of Subgen.*Strobus* and Subgen.*Pinus* were showed as follows.

The construction of characteristic sequences of the 24 samples is :

**characteristic sequence = core sequence + related sequence**

**Class A(Subgen.*Strobus*)**

S2:S1 S2[a] S8[b] S4 [c]
S5:S1 S5 S8 S4
S1:S1 S2 S5 S4 S17 S21
S8:S2 S5 S8 S4 S12 S23
S6:S6 S7
S7:S6 S7

**Class B(Subgen.*Pinus*)**

S3:S3 S4 S9 S10 S14 S15 S17 S18 S19 S20 S21S23
S4:S4 S3 S12 S15 S17 S19 S21S23 S1 S2 S5 S8
S9:S3 S9
S10:S3 S10 S11 S14 S15 S16 S17 S18 S20 S21 S23
S11:S10 S11 S12 S14 S17 S23
S12:S4 S11 S12 S14 S15 S17 S19 S20 S21 S23 S8
S13:S13 S20 [0.667][d]
S14:S3 S10 S11 S12 S14 S15 S17 S19 S20 S21 S23
S15:S3 S4 S10 S12 S14 S15 S16 S17 S18 S19 S20 S21 S23
S16:S16S10 S15
S17:S3 S4 S10 S11 S12 S14 S15 S17 S18 S19 S20 S21 S23 S1
S18:S3 S10 S15 S17 S18 S19 S20 S21 S22 S23
S19:S3 S4 S12 S14 S15 S17 S18 S19 S20 S21 S22

S20:S3 S10 S12 S14 S15 S17 S18 S19 S20 S21 S23
S21:S3 S4 S10 S12 S14 S15 S17 S18 S19 S20 S21 S23 S1
S22:S18 S19 S22
S23:S3 S4 S10 S11 S12 S14 S15 S17 S18 S20 S21S23 S8
S24:S24S18[0.614 ]S22[0.610 ][d]

S2:S1S2[a]S8[b]S4[c] , this sequence was the characteristic sequence of S2. *a*. S2, each sample is the most similar sample in its own characteristic sequence. *b*. S1S2S8 was the core sequence of S2. *c*. S4 was the sample in related sequence of S2. *d*. S13:S13S20[0.667], 0.667 was the common composition ratio between S13 and S20. S24:S24S18[0.614]S22[0.610], the part sequence with double underline was the close sequence of S24, in which the composition ratios of sample S18,S22 related to S24 were lower than 69%. This part sequence shows that S24 belongs to Subgen.*Pinus*.

**4 Evolution pattern analysis of the two subgenus**

In above characteristic sequences, the common composition ratios of samples to reference samples were equal to or larger than 69%. These results represented that Subgen.*Strobus* and Subgen.*Pinus* could be distinguished clearly with the standard $P_g \geqq 69\%$. In class A Subgen.*Strobus*, the characteristic sequences of S1,S2, S5, S8 were with obviously related sequence made up of samples of Subgen.*Pinus*. This uncovered that Subgen.*Strobus* is greatly of the properties of Subgen.*Pinus*. The characteristic sequences of samples of Subgen.*Strobus* included samples of Subgen.*Pinus*, and the ratios of Subgen.*Strobus*' samples to Subgen.*Pinus*' samples were very high, and variations among characteristic sequences of Subgen.*Strobus* were distinct. Compared with Subgen.*Pinus*' samples, their characteristic sequences almost contain no Subgen.*Strobus*' samples. This asymmetry revealed that Subgen.*Strobus* should evolved from Subgen.*Pinus*. These relationships were displayed qualitatively in figure 2.


**Fig.2 the sample set of characteristic sequences of Subgen.*Strobus* and Subgen.*Pinus***
A, Subgen.*Strobus*, B. Subgen.*Pinus*
Green/deep color region express Subgen.*Pinus*, yellow/light color region express Subgen.*Strobus*. The large size of area show the number of samples was large.

On the other hand, in characteristic sequences of Subgen.*Strobus*, the number of samples in core sequence of every Subgen.*Strobus* sample was small, and there were great change among their core sequences. This fact pointed out there were obvious variations among them, and the origins should be from multiple sources.

From geographic space point of view, if Subgen.*Pinus* is in a concentration distribution space, Subgen.S*trobus* should distribute around the space of Subgen.*Pinus*. In the evolution chain,Subgen.*Pinus* was first originated from it ancestor, then Subgen.*Strobus* came from Subgen.*Pinus*. This involving relation existed in characteristic sequences was the equal of the overlapping relation between different sets. By this view, the origins of different organisms could be inferred from the mutual involving relation, and evolution history could be portrayed briefly. This verified that the common heredity and variation information theoretical equation could discover the quantitative evolution principle of biology, and it provides a new theory for investigating biological evolution.

From the related sequences of Subgen.*Strobus*' samples,the samples in the related sequences were mainly S4,S12,S17, S21 and S23, which are Subgen.*Pinus*. This inferred that Subgen.*Strobus* were mainly evolved from S4,S12,S17, S21 and S23 of Subgen.*Pinus*. Grounded on the core sequences of Subgen.*Strobus*' samples, they can be divided into two subgroups, subgruop1:S1,S2,S5,S8 and subgroup 2: S6,S7. Their core sequences construct two independent sets. This showed there are obvious differences in property between the two subsets.

S4, S12, S17, S21 and S23 existed in related sequences of S1,S2,S5,S8 , this manifested there exist the kinship between S1,S2,S5,S8 and Subgen.*Pinus*.

In accordance with characteristic sequences, samples S3,S4,S9, which are Subgen.*Strobus* based on the classical classification results, could be classified into Subgen.*Pinus* perfectly by means of this new approach.

On the other flip side, there were no or few samples in the related sequences of S3,S4,S9~S24.This revealed that there are distinct differences between Subgen.*Strobus* and Subgen.*Pinus*. Depending on the involving relation between characteristic sequences of Subgen.*Strobus* and subgen. *Pinus*. The maximum likelihood is Subgen.*Strobus* evolving from Subgen.*Pinus*. This also means Subgen.*Pinus* is ancestor , Subgen.*Strobus* is the son.

The results offered above represented that $P_g \cong 69\%$ could be used as the absolutely quantitative standard for distinguishing Subgen.*Strobus* and Subgen.*Pinus*. Two subgenus are of significant asymmetry variation in evolution. No mater from the common composition ratio values, or from the properties, the conclusion are all fit to theoretical derivation. This demonstrates that dual index information theory equation is capable of accurately describing biological evolution rules.

In Subgen.*Pinus*, some kinds of pine species were of distinct feature. Although S4 was divided into class B, that is Subgen.*Pinus*, its related sequence contained many samples of Subgen.*Strobus*, S4:S3S4S12S15S17S19S21S23S1S2S5S8, this expressed S4 is in the transition stage of evolution from Subgen.*Strobus* to Subgen.*Strobus*. S4 may be named the transition species from Subgen.*Pinus* to Subgen.*Strobus*.

Although S9 was Subgen.*Pinus*, there were great differences between S9 and other samples of Subgen.*Pinus*, based on its characteristic sequence. S9 is Pinus *strobus* (eastern white pine), the geographic feature of its origin is extremely different from that of other samples. This brings about its variation being greatly compared to other samples. Pinus *strobus* Linn originated from eastern north of America, its most similar

sample was S3, that is Pinus armandii Franch, which originate from Hua mountain in Shanxi province of China. The geographic feature of origin districts is similar with each other. The climate is moist and warm cool.

S13 is Pinus *Latteri* Mason, seen in section 3.2.2. Its geographic areas differ from that of other pine species. This resulted in its properties, chemical compositions being of unique feature. From its characteristic sequence, its most similar sample was S20, whose common composition ratio was merely equal to 67%, less than 69%. Thus S13 could be viewed as an independent class.

S20 is pinus taiwanensis Hayata var, it distributes in Daming mountain of Guangxi province and Fanjing mountain of Guizhou province in China. Its geographic features differ from that of other pinus origins significantly.

S24 is Pinus roxburghii (Chir Pine, or longleaf lndian) pine, and belongs to Subgen.*Pinus*,depending on samples in characteristic sequence. However, there existed obvious differences between it and other pine species of Subgen.*Pinus*. This was because of the extreme difference in its growth geographic environment between it and other samples of Subgen.*Pinus*. Its origin center is in the southern slop of Himalaya mountains. S24 was collected from this center,its unique geographic environment makes it differ from other samples of Subgen.*Pinus* significantly, and determines that there was no sample with high similarity to it. The most similar samples to it were S18, S22, while the common composition ratios are merely 0.614 and 0.610, which are far less than 69%, respectively.

Compared the characteristic sequence of classes A with that of class B, one can find out that most characteristic sequences of samples in Subgen.*Pinus* vary slightly. This indicated that Subgen.*Pinus* is in steady evolution state. In contrast, the characteristic sequences of Subgen.*Strobus* samples change greatly and randomly.These revealed that Subgen.*Strobus* is in unsteady evolution state. The analytical results expressed that the characteristic constant $P_g$=70%, obtained from dual index information theory equation, relying on its asymmetric variations, combing with the maximum number of effective sample method[13], could be employed as absolutely theoretical standard for classifying Subgen.*Strobus* and Subgen.*Pinus*. Another constant $P_g$=0.61 ,based on symmetric variation, could be as the absolute theoretical standard for determining which trees are Pinus.

## 5 Conclusion

The two constants derived from biological common heredity and variation information theory equation, combining with compounds in oleoresins metabolized by pines, could accurately identify Subgen.*Strobus* and Subgen.*Pinus* without any help of empirical knowledge related to learning samples. This showed that common heredity and variation information theory equation can uncover biological evolution rule excellently. The two constants can reveal the features of some biological categories. In other hand, the characteristic sequences of samples in each class were able to express evolution information of organisms. The involving relationships between core

and related sequences, which belong to different classes, respectively, can uncover evolution history and relationship of different classes. One can determine which is ancestor, and which is the son. This research indicated that Subgen.*Strobus* should originate from Subgen.*Pinus* and come from multiple species of Subgen.*Pinus*. Common heredity and variation information theory provides us with a new approach for accurately classifying biology, analyzing evolution relationships of biology by applying molecular structure, molecular kind information. The two constants may be used as absolutely theoretical criteria for classifying different biological categories, and may revise the results obtained based on classical classification methods.For these categories, such as species, subgenus and genus with very close kinship, their inner similarity is high, the two constants can be adopted as the theoretical criteria for determining them. However, for many other categories, such as family,order, class,phylum, and kingdom, being of distant relatives, samples in one category of them are of low similarity. These two constants should not suit to these categories. For family, the two constants may fit to, or not match with the classification of them. Further more, whether biological common heredity and variation information theoretical equation is suitable for analyzing structure information in DNA , protein, all these need to be further investigated for us.

## 6 Instruments and methods[20]

**To see supplementary 2.**

## References

[1] Ernst Mayr. The growth of biological thought-diversity, evolution and inheritance, Chengdu: Sichuan education press,2010.

[2] Zhou Changfa, Yang Guang. The status and definition of species,Beijing: Science press,2011.

[3] Wang Jinwu.Spermatophyte taxonomy.Beijing:High Education Press，2011（second edition).

[4] Douglas J. Futuyma. Evolution。Beijing:High Education Press，2016 (third edition).p443-453.

[5] Nicholas H.Barton，Derek E.G.Briggs，Jonathan A.Eisen，David B.Goldstein, Nipam H.Patel. Evolution. Beijing: Science press, 2010.P648-686.

[6] Michael G.Simpson. Plant systematics. Beijing: Science press, 2011(second edition).

[7] Edward O. Wilson.The diversity of life. Beijing: China citic press, 2016. p59-62.

[8] Stephen T. Garnett, Les Christidis. Taxonomy anarchy hampers conservation. Nature,2017, 546:25–27.

[9] Xu Kexue.Biological mathematics. Beijing: Science press,2002.

[10] Gurcharan Singh. Plant systematics－an integrated approach. Beijing:Chemical industry press, 2010.

[11] Huang yuan. Molecular phylogenetics.Beijing: Science press, 2012.

[12] Zou Huabin.Dual index information markedly similar sequence clustering analysis on IR fingerprint spectra of extracts of Guifu Dihuang and JinguiShenqi pills with ethanol.China Journal of Chinese Materia Medica, 2009,34(18): 2325-2330.

[13] Zou Huabin.A systematically theoretical distinguish approach for traditional Chinese medicine with identical quality.World Chinese medicine，2015,10(7):1078-1082.

[14] 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.

[15] Huabin zou. Two Chinese medicine species constants and the accurate identification of Chinese medicines.Jul. 20, 2017. bioRxiv doi: http://dx.doi.org/10.1101/166140

[16] Shaw G R .The genus Pinus (Arnold Arboretun Pub. No.5)[M], Boston, Houghton Mifflin Co., 1914.

[17] Pilger R. Genus Pinus[M]//Die Natürlichen Pflanzenfamilien, Vol. XIII. Engler A, Prantl K. Gymnospermae. Leipzig:Wilhelm Engelmann, 1926.

[18] Critchfield W B, Little E L. Geographic distribution of the world [M]. Washington, D.C.: Forest Service, 19.

[19] Wu Zhonglun. Classification and distribution of Chinese pine genus. Acta phytotaxonomica Sinica, 1956,5(3):131-163.

[20] Song Zhanqian. Oleoresin characteristics and chemical classification. Hefei: University of Science and Technology of China Press, 2009. p8-12,12-14, 18-21,22-25, 123-127, 131-136.

[21] Oudin, A. Les amendements et engrais dans les reboisements. REV EAUX ET FOReTS 1939, 77(4): 335-341.

[22] Mirov N T. Composition of gum turpentines of pines[R]. USDA Technical Bulletin No. 1239, 1961.

[23] Mirov N. T . The genus Pinus [M].New York:The Ronald Press Company,1967.

[24] Hegnauer R. Chemotaxonomic der Pflanzen[M]. Part 1. Basel :Birkhaüser Verlag,1962.

[25] Swain T.Chemical Plant Taxonomy.Academic Press, London, United Kingdom., 1963

[26] Alston R. E, Turner B. L..Biochemical Systematics. Science, 1963, 141 (3582): 709.

[27] Zou Huabin, Yuan Jiurong, Yuan Hao. Study of HPLC-FPS and 4-Dimensional UV-FPS of Peru's Ginseng healthcareproduct. Chinese Traditional Patent Medicine,2003, 25(4):88-92.

[28] Zou Huabin, Yuan Jiurong, Lv Qingtao,Rong Rong.The DualIndex Sequence Analytical Method of Common Peak Ratioand Variant Ratio for Analysing UV Fingerprint Spectra of Radix Glycyrrhizae.Jorunal of Chinese Medicinal Materials,2003,26(9):625-629.

[29] 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.

[30] Zou Huabin , Yuan Jiurong , Du Aiqin , Sun Linlin , Hassan Y. Aboul Enein .. Dual-Index Sequence Analytical Method for IR Fingerprint Spectra of Ethanolic Extract of Various Gylcyrrhizae's Root Species components. Analytical letters, 2005, 38 (7): 1167 - 1178

[31] Huabin Zou , Guosheng Yang , Aiqin Du , Jiurong Yuan , Zhengran Qina Yingying Xia , Hassan Y. Aboul-Enein. Combinational Numeral Fingerprint Spectra of Glycyrrhiza and Analysis of Common Peak Ratio Invariableness In HPLC. Biomed. Chromatogr. 2006, 20: 642–655.

[32] 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 Pal1. Spectroscopy and Spectral Analysis，2007, 28(9):1815-1819.

[33] Hua-Bin Zou, Guo-Sheng Yang, et al. Recognition of Radix Paeoiae Alba herbs（Fam. Ranunculaceae）with HPLC Fingerprint Spectra. Analytical letters, 2008,41:3309-3323.