· Research Paper ·

Relationship between DNA Barcoding and Chemical Classification of Salvia Medicinal Herbs

HAN Jian-ping[1*], LIU Chang[2*], LI Min-hui[1,3], SHI Lin-chun[1], SONG Jing-yuan[1], YAO Hui[1], PANG Xiao-hui[1], CHEN Shi-lin[1,4**]

1. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China
2. Molecular Chinese Medicine Laboratory, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China
3. Baotou Medical College, Baotou 014040, China
4. Hubei University of Chinese Medicine, Wuhan 430065, China

Abstract: **Objective** To make the identification of medicinal herbs in *Salvia* L. quickly and accurately. **Methods** In this work, DNA barcoding and chemical fingerprint were compared for the identification of herbs in *Salvia* L. First, the nucleotide sequences of the internal transcribed spacer region two amplified from 48 medicinal plants in *Salvia* L., and three other groups of medicinal plants in Lamiaceae were sequenced. A molecular phylogeny was constructed using the minimum evolution and maximum parsimony methods according to their sequence diversity. Second, the water-solution bioactive components and lipid soluble components were tested by HPLC. Then a chemical phylogeny was built using HPLC fingerprint data. Comparing the molecular and chemical phylogenetic trees revealed many similarities. **Results** DNA barcoding was sequencing based and could therefore provide more accurate results within a shorter time especially in large-scale studies. **Conclusion** The results show that ITS2 region is a novel DNA barcode for the authentication of the species in *Salvia* L. This is the first work to show the relationship between DNA barcoding and chemical components.

Key words: authentication; DNA barcoding; HPLC fingerprint; internal transcribed spacer region 2; quality control; *Salvia* L.
DOI: 10.3969/j.issn.1674-6384.2010.01.002

Introduction

The genus *Salvia* L. (tribe Mentheae, Lamiaceae) represents an enormous and cosmopolitan assemblage of nearly 1000 species and it has undergone marked species propagation in three regions of the world: Central and South America (500 spp.), Central Asia/ Mediterranea (250 spp.), and Eastern Asia (90 spp.) (Walker *et al*, 2004). Approximately 84 species in *Salvia* L. are native to China. Three groups (high-mountain Danshen, low-mountain Danshen, and non-Danshen) (Xiao, Feng, and Xia, 1997) were divided in China using a morphologic character-based numerical taxonomy. Many species of *Salvia* L. have been used as medicinal herbs with active components of traditional Chinese medicines (TCM) for a long time. For example, Danshen, the root and rhizome of *Salvia miltiorrhiza* Bunge, has been used as a herbal drug in the practice of TCM for thousands of years. In China, over 20 species of *Salvia* L. have been used as Danshen in TCM for the treatment of coronary heart disease and stroke (Li *et al*, 2008a). These species differ in their pharmacological activities as well as toxicities in various formulations, and the usage of correct species of *Salvia* L. in the specific formulations are critical to ensure the effectiveness and safety of these drugs.

Morphological characteristics have been used as markers for the identification of the species in *Salvia* L. (Cao and Xie, 2007). However, limitations of phenotypic traits, which will be discussed later, make the unambiguous identification of crude plant materials very difficult,

* contributed equally to the article ** Address for correspondence Prof. Chen SL Address: Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China Tel: +86-10-62899700 Fax: +86-10-62899776 E-mail: slchen@implad.ac.cn
Received: December 1, 2009; Revised: December 15, 2009; Accepted: December 28, 2009

and reports on the misuse species in *Salvia* L. are frequent. For example, non-Danshen, the root of *S. bowleyana* Dunn, was mistakenly used as Danshen in the remote mountain areas of Zhejiang, Jiangxi, and Anhui Provinces, China. Considering the economic and medical importance materials of *Salvia* L., a method that unambiguously and rapidly distinguishes the species of *Salvia* L. and differentiates the species in *Salvia* L. from other similar species would be a powerful tool enabling the production of safe and effective herbal drugs from the species of *Salvia* L.

Two different levels of species identification are relevant to medicinal plants. First, the correct plant species must be used (genetic authentication). Second, the chemical component responsible for the corresponding pharmacological activity must be maximally produced (chemical authentication). DNA barcoding is more suitable for solving the genetic authentication problem, while the second problem is best solved by studying the chemical profiles of the plant materials to ensure that the active components are well defined. Species identification by DNA barcode is based on the sequencing of a short standardized genomic region of the target specimen and comparing this information to that in a reference sequence library from known species (Hebert *et al*, 2003). Although chemical profiling has also been used for plant species identification, DNA barcoding has several advantages. Because the DNA sequence of an individual is definite and remains identical in different plant tissue types, at different development stages, and under various environmental conditions, there is little noise in the DNA sequence for a plant species. Furthermore, the technologies for isolating and computationally analyzing DNA barcodes, such as DNA extraction, PCR, DNA sequencing, and blast searches, have become laboratory routines, making DNA barcoding technology more robust and practical than chemical profiling methods such as HPLC fingerprinting.

Several coding and non-coding regions have been proposed for use as DNA barcodes in plants. These include rbcL, matK, psbA-trnH, and ribosomal intergenic spacer regions (ITS) (Chase *et a*l, 2005; Kress and Erickson, 2007; Kress *et al*, 2005; Lahaye *et al*, 2008). The ITS regions are interspersed among the rRNA genes and can excise themselves during the maturation of the precursor ribosomal RNA (rRNA) transcripts (Miao *et al*, 2008). The ITS regions can be further subdivided into ITS1, which is located between small subunit (SSU) and 5.8S rRNA genes, and ITS2, which separates the 5.8S and large subunit (LSU) rRNA genes. ITS2 is a variable region that is relatively short (200–300 bp long) and easily sequenced. It has been shown to be useful as a possible source of polymorphisms for plant identification (Baldwin *et al*, 1995). A more recent study has shown that it is a double-edged tool for eukaryotic evolutionary comparisons (Coleman, 2003). The study of CHIOU Shu-Jiau showed that ITS2 could be amplified well with specific primers and could be used to authenticate medicinal herbs (Chiou *et al*, 2007; Chen *et al*, 2010). Furthermore, the ITS2 region is one of the more frequently utilized regions for phylogenetic analyses at the genus and species levels (Coleman, 2003). In this study, we used ITS2 to identify medicinal plant of the *Salvia* L.

DNA barcoding could not embody the quality of herbs. At the same time, we highlighted useful analytical techniques that could be employed to analyze DNA for quality assurance, control, and authentication of medicinal plant species. DNA barcoding and chemical fingerprint are two approaches that have recently garnered much attention. However, to date, there has been no report on the relationship between these two. In this study, we compared these two methods for the identification of the genus of *Salvia* L.

Materials and methods

Acquisition of plant materials

Plant specimens were collected from a wide range of geographical areas (Table 1) and were authenticated by Prof. LIN Yu-lin (Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences) and Prof. LI Xi-wen (Kunming Institute of Botany, Chinese Academy of Sciences). The voucher specimens were deposited in the herbarium of the Institute of Medicinal Plant Development, Beijing, China.

DNA extraction and PCR amplification

The leaves used for DNA extraction were silica-dried if collected fresh, or were obtained directly from dried herbarium specimens (Table 1). Total genomic DNA was extracted using the Plant Genomic DNA Kit (Tiangen Biotech Co., China) following the recommended protocols. For PCR amplification, the primers (F-1 and R-2) were designed using Primer Select software as part of the Lasergene package (Burland, 2000). The sequence of F-1

Table 1 Plant specimens used in this study

Voucher No.	Genus	Subgenus	Section	Species + sample name	Origin	Place collected	Genebank Acc. No.
PS0121MT02	*Salvia* L.	*Allagospa-donopsis*	*NA*	*Salvia chinensis* Benth. *S3*	domestic	Nanning, Guangxi, China	FJ883504
PS0121MT01	*Salvia* L.	*Allagospa-donopsis*	*NA*	*Salvia chinensis S4*	domestic	Tianmu mountain, Zhejiang, China	FJ883503
PS1728MT01	*Salvia* L.	*Allagospa-donopsis*	*NA*	*Salvia kiangsiensis* C. Y. Wu.	domestic	Xinning, Hunan, China	FJ883514
PS1702MT01	*Salvia* L.	*Allagospa-donopsis*	*NA*	*Salvia liguliloba* Sun.	domestic	Tianmu mountain, Zhejiang, China	FJ883515
PS1704MT01	*Salvia* L.	*Jungia*	*NA*	*Salvia dugesii* Fernald	foreign	Kunming, Yunnan, China	FJ883508
PS1731MT01	*Salvia* L.	*Jungia*	*NA*	*Salvia farinacea* Benth.	foreign	Yaozhisuo, Beijing, China	FJ883510
PS0153MT01	*Salvia* L.	*Jungia*	*NA*	*Salvia splendens* Ker-Gawl. *S1*	foreign	Nanning, Guangxi, China	FJ883530
PS0153MT02	*Salvia* L.	*Jungia*	*NA*	*Salvia splendens S2*	foreign	Nanning, Guangxi, China	FJ883530
PS0153MT03	*Salvia* L.	*Jungia*	*NA*	*Salvia splendens S3*	foreign	Nanning, Guangxi, China	FJ883530
PS0153MT04	*Salvia* L.	*Jungia*	*NA*	*Salvia splendens S4*	foreign	Lijiang, Yunnan, China	FJ883530
PS1727MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia campanulata* Wall.	domestic	Zhongdian, Yunnan, China	FJ883500
PS1724MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia castanea* Diels	domestic	Wolong, Sichuan, China	FJ883501
PS1706MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia digitaloides* Diels	domestic	Lijiang, Yunnan, China	FJ883507
PS1720MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia evansiana* Hand.-Mazz.	domestic	Lijiang, Yunnan, China	FJ883509
PS176MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia flava* Forrest ex Diels	domestic	Muli, Sichuan, China	FJ883511
PS1730MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia maximowicziana* Hemsl.	domestic	Tianshui, Ganshu, China	FJ883516
PS1712MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia pauciflora* Stib.	domestic	Zhongdian, Yunnan, China	FJ883523
PS1709MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia prezewalskii* Maxim. *S1*	domestic	Lijiang, Yunnan, China	FJ883525
PS1709MT02	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia prezewalskii S2*	domestic	Zhongdian, Yunnan, China	FJ883526
PS1710MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia roborowskii* Maxim.	domestic	Zhongdian, Yunnan, China	FJ883528
PS0155MT01	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia umbratica* Hance *S1*	domestic	Yanqing, Beijing, China	FJ883532
PS0155MT02	*Salvia* L.	*salvia*	*Eurysphace*	*Salvia umbratica S2*	domestic	Nannin, Guangxi, China	FJ883532
PS0151MT01	*Salvia* L.	*salvia*	*Eusphace*	*Salvia fruticosa* Mill	foreign	Athens, Greece	FJ883512
PS1700MT01	*Salvia* L.	*salvia*	*Eusphace*	*Salvia officinalis* Linn. *S1*	foreign	Boston, America	FJ883521
PS1700MT02	*Salvia* L.	*salvia*	*Eusphace*	*Salvia officinalis S2*	foreign	Boston, America	FJ883522
PS0134MT01	*Salvia* L.	*salvia*	*Eusphace*	*Salvia superba*（Silva Tar. & C.K.Schneid.)	foreign	Yaozhisuo, Beijing, China	FJ883531
PS1701MT01	*Salvia* L.	*sclarea*	*Aethiopis*	*Salvia sclarea* L.	foreign	Yaozhisuo, Beijing, China	FJ883529
PS1718MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia bowleyana* Dunn	domestic	Jiuhua mountain, Anhui, China	FJ883499
PS1729MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia cavaleriei var. simplicifolia* Stib.	domestic	Xinning, Hunan, China	FJ883502
PS1723MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia dabieshanensis* J. Q. He	domestic	Yu mountain, Anhui, China	FJ883505
PS1722MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia honania* L. H. Bailey	domestic	Nanyang, Henan, China	FJ883513
PS1719MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia meiliensis* S. W. Su	domestic	Huoshan, Anhui, China	FJ883517
PS1699MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia miltiorrhiza Bunge var. miltiorrhiza f. Alba* C.Y.Wu et H.W.Li *V1*	domestic	Taian, Shandong, China	FJ883520
PS0110MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia miltiorrhiza* Bunge *S1*	domestic	Shangxia, Shanxi, China	FJ883518
PS0110MT02	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia miltiorrhiza S2*	domestic	Yixian, Hebei, China	FJ883519
PS0110MT05	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia miltiorrhiza S3*	domestic	Zhongjiang, Sichuan, China	FJ883519
PS1711MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia prionitis* Hance	domestic	Guilin, Guangxi, China	FJ883527
PS1714MT01	*Salvia* L.	*sclarea*	*Drymosphace*	*Salvia vasta* H. W. Li, Bull	domestic	Luotian, Hubei, China	FJ883533

To be continued

Continued Table 1

Voucher No.	Genus	Subgenus	Section	Species + sample name	Origin	Place collected	Genebank Acc. No.
PS1689MT01	*Salvia* L.	*sclarea*	*Notiosphace*	*Salvia plebeian* R. Br. *S1*	domestic	Zhongdian, Yunnan, China	FJ883524
PS1689MT02	*Salvia* L.	*sclarea*	*Notiosphace*	*Salvia plebeian S2*	domestic	Baiwangshan, Beijing, China	FJ883524
PS1689MT03	*Salvia* L.	*sclarea*	*Notiosphace*	*Salvia plebeian S3*	domestic	Yaozhisuo, Beijing, China	FJ883524
PS1705MT01	*Salvia* L.	*sclarea*	*Plethiosphace*	*Salvia deserta* Schang	foreign	Urumqi, Xinjiang, China	FJ883506
PS0125MT04	*Ajuga* L.	*NA*	*NA*	*Ajuga ciliate* Bunge		Nannin, Guangxi, China	FJ*883495*
PS1738MT01	*Ajuha* L.	*NA*	*NA*	*Ajuga decumbens* Thunb.		Huangshan, Anhui, China	FJ883496
PS0104MT02	*Ajuga* L.	*NA*	*NA*	*Ajuga lupulina* Maxim.		Wolong, Sichuan, China	FJ883497
PS1733MT01	*Phlomis* L.	*NA*	*NA*	*Phlomis melanantha* Diels		Lijiang, Yunnan, China	FJ883498
PS0122MT01	*Scutellaria* L.	*NA*	*NA*	*Scutellaria baicalensis* Georgi		Daqingshan, Shandong, China	FJ883534
PS0120MT01	*Scutellaria* L.	*NA*	*NA*	*Scutellaria indica* Linn.		Daqingshan, Shandong, China	FJ883535

is 5'-ATGCGATACTTGGTGTGAAT-3'. The sequence of R-2 is 5'-GACGCTTCTCCAGACTAACAAT-3'. The PCR reaction mixture consisted of 2 μL DNA (about 15 ng), 2.5 μL of 10 × PCR buffer, 1.5 μL of 25 mmol/L $MgCl_2$, 1.5 μL of 2.5 mmol/L dNTPs, 1.5 U of Taq DNA polymerase (SBS Genetech Co., China), 2.0 μL each of 2.5 μmol/L IT1F and IT2R primers (synthesized by SBS Genetech Co., China) in a final volume of 25 μL. Cycling conditions consisted of an initial 5 min at 94 ℃, followed by 30 s denaturing at 94 ℃, 30 s annealing at 53 ℃ and 45 s elongation at 72 ℃ repeated for 39 cycles, and a final extension of 72 ℃ for 7 min. The PCR products were examined with 1.5% agarose gel electrophoresis and were visualized by ethidium bromide staining under UV.

DNA sequencing

The PCR products were purified by the PCR purification kit (Tiangen Biotech Co., China). All purified PCR products were directly sequenced by the sequencing center at The Chinese Academy of Agricultural Sciences using an ABI 3730 DNA sequencer (Applied Biosystems Industries, USA). The primers used for PCR amplification were also used as the sequencing primers. Multiple reads of the same PCR fragments were subjected to Contig assembly using CodonCode Aligner (CodonCode Co., Germany).

Data analyses

The data describing the abundance of the major chemical components in the specimens were obtained from our previous studies (Li *et al*, 2008a; Li *et al*, 2008b). The missing data points from those studies were transformed, so that "—" was replaced with 0.01 and "+" was replaced with 0.05 (Table 2). Although four additional compounds (F, G, H, and I) were described in the previous studies, their abundances were below the detection limits in all specimens tested. As a result, they were not included in the current study. Multiple sequence alignment was carried out using Clustalw (Larkin *et al*, 2007) either as a standalone application or as part of the MEGA4 software package (Tamura *et al*, 2007). The divergence between pairs of sequences was calculated using the Kimura 2-parameter model (Kimura, 1980), and all positions containing gaps and missing data were eliminated from the data set (complete deletion option). Phylogenetic analysis was conducted using MEGA4. Hierarchical clustering analysis of the chemical profiling data was performed using JMP software (Version 7.0, SAS, Cary, NC, USA). Multivariate data analyses, such as principle component analysis (PCA) and partial least squares (PLS), were carried out using the SIMCA software (Version 10, Umetrics, Sweden). Haplotype tagging SNPs were identified using the BEST software (Sebastiani *et al*, 2003).

Results

Sequence analysis

The forty-eight specimens (Table 1) used in this study belong to four genera, including *Salvia* L. (42 specimens), *Ajuga* L. (3 specimens), *Phlomis* L. (1 specimen), and *Scutellaria* L. (2 specimens). The forty-two specimens of *Salvia* L. belong to 30 unique species and multiple specimens were obtained from the following species: *S. chinensis* Benth. (2 specimens), *S. officinalis* L. (2), *S. prezewalskii* Maxim. (2), *S. umbratica* Hance (2), *S. plebeian* (3), *S. miltiorrhiza* Bge. (4), and *S. splendens* Ker-Gawl (4). Genomic DNA was extracted from individual dried specimens. ITS2-containing fragments

Table 2 Distribution of major chemical components in the specimens used for the construction of chemical phylogeny

Species	Compounds								
	A	B	C	D	E	I	J	K	L
S. bowleyana	0.05	0.05	0.05	0.05	0.05	0.05	8.03	6.34	76.45
S. campanulata	0.05	0.05	0.05	0.05	0.05	0.7	17.55	0.01	11.44
S. castanea	0.05	0.05	0.05	0.36	0.05	0.18	11.31	0.01	0.05
S. cavaleriei	0.05	0.05	0.05	0.05	0.05	0.05	14.49	1.94	54.07
S. chinensis S3	0.01	0.01	0.01	0.01	0.05	0.01	2.78	1.11	25.95
S. chinensis S4	0.01	0.01	0.01	0.01	0.05	0.01	1.95	0.92	9.5
S. dabieshanensis	0.14	0.36	0.14	0.62	0.05	0.05	5.38	4.54	55.36
S. deserta	0.01	0.01	0.01	0.01	0.05	0.05	4.77	6.36	0.01
S. digitaloides	0.04	0.05	0.05	0.4	0.05	0.05	11.35	0.01	0.05
S. dugesii	0.01	0.01	0.01	0.01	0.01	0.05	0.56	0.01	0.01
S. evansiana	0.05	0.05	0.05	0.05	0.24	0.05	25.32	0.05	15.15
S. farinacea	0.01	0.01	0.01	0.01	0.05	0.01	7.15	0.01	0.01
S. flava	0.05	0.05	0.05	0.05	0.05	0.05	29.41	0.01	1.37
S. fruticosa	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. honania	0.1	0.18	0.3	0.51	0.05	0.05	3.66	1.66	18.37
S. kiangsiensis	0.01	0.01	0.01	0.01	0.05	0.05	17.73	0.05	11.01
S. liguliloba	0.01	0.01	0.01	0.01	0.53	0.01	5.46	0.01	0.01
S. maximowicziana	0.05	0.05	0.05	0.15	0.05	0.29	26.31	0.01	1.66
S. meiliensis	0.11	0.37	0.18	0.8	0.05	0.05	2.81	2.2	33.37
S. miltiorrhiza S1	0.24	0.44	0.81	1.39	0.33	0.05	4.57	3.6	45.38
S. miltiorrhiza S2	0.21	0.29	0.64	1.46	0.29	0.21	4.42	1.08	55.77
S. miltiorrhiza S3	0.51	1.84	0.85	2.82	0.3	0.05	1.47	0.97	30.95
S. miltiorrhiza V1	0.16	0.25	0.11	2.6	0.48	0.05	3.12	1.71	50.69
S. officinalis S1	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. officinalis S2	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. pauciflora	0.05	0.12	0.13	1.06	0.05	0.35	10.99	0.01	0.05
S. plebeia S1	0.05	0.05	0.01	0.01	0.05	0.05	0.05	0.01	0.05
S. plebeia S2	0.05	0.05	0.01	0.01	0.87	0.42	4.58	0.01	0.05
S. plebeia S3	0.05	0.05	0.01	0.01	0.05	0.05	6.36	0.01	0.05
S. prezewalskii S1	0.53	1.41	1.07	4.94	0.05	0.48	3.53	0.01	1.44
S. prezewalskii S2	0.19	0.56	0.28	1.78	0.05	0.05	2.49	0.01	2.55
S. prionitis	0.05	0.13	0.05	0.05	0.01	0.05	1.62	0.52	1.79
S. roborowskii	0.05	0.05	0.06	0.1	0.05	0.05	7.8	0.01	0.01
S. sclarea	0.01	0.01	0.05	0.01	0.01	0.05	1.01	0.01	0.01
S. splendens S1	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. splendens S2	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. splendens S3	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. splendens S4	0.01	0.01	0.01	0.01	0.05	0.05	0.05	0.01	0.01
S. superba	0.01	0.01	0.01	0.01	0.01	0.01	3.35	0.01	0.01
S. umbratica S1	0.05	0.05	0.05	0.05	0.01	0.01	0.05	0.01	0.05
S. umbratica S2	0.05	0.05	0.05	0.05	0.01	0.01	0.05	0.01	0.01
S. vasta	0.51	0.18	0.79	0.59	0.05	0.05	7.56	2.82	63.26

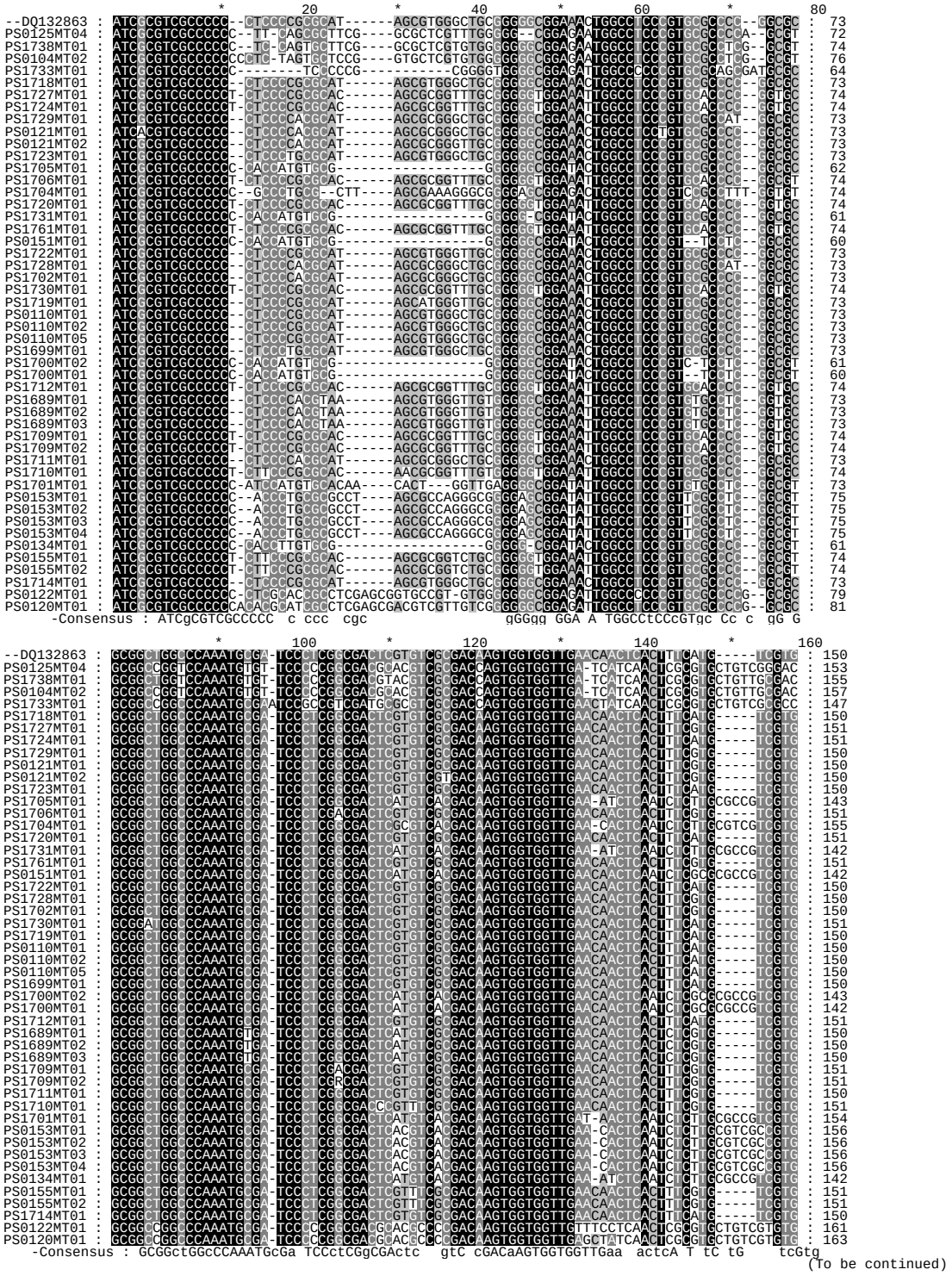
were amplified by PCR and sequenced. The resulting sequences were used to search the public sequence database, and a multiple sequence alignment was generated (Fig. 1). One of the previously known ITS2 sequences (Genebank Acc No: DQ132863) from *S. miltiorrhiza* was used as a template, and this determined the starting and ending positions of the ITS2 sequences. The alignment of all the ITS2 sequences is shown in Fig. 1.

Phylogenetic analysis of species in *Salvia* L.

As shown in Table 1, the 48 specimens belong to four genera, and the 42 specimens from the *Salvia* L. belong to four sub-genera: *Salvia* L., *Sclarea*, *Jungia* L., and *Allagospadonopsis*. The sub-genus *Salvia* L. contains two sections: *Eurysphace* and *Eusphace*. By contrast, the

the sub-genus *Sclarea* contains four sections: *Drymosphace*, *Notiosphace*, *Plethiosphace*, and *Aethiopis*. In this study, all species belonging to section *Plethiosphace*, section *Eusphace*, sub-genus *Jungia* L., and section *Aethiopis* were introduced originally or were sampled directly from territories outside of China.

The ITS2 regions range from 219 to 230 base pairs (bps) long. The aligned length of the data was 245 bps. With regions of ambiguous alignment or ambiguous sequences excluded, the total length of included characters was 243 bps. Of these 243 characters, 92 were conserved, 151 were variable, and 85 (56.3%) were potentially parsimony-informative. These data for a total of 48 taxa were used to construct the phylogenetic tree using MEGA4. ITS2 sequences from three non-*Salvia* L. genera were included as the outgroup. The evolutionary history was inferred using the minimum evolution (ME) and maximum parsimony (MP) methods, respectively (Fig. 2).



(To be continued)

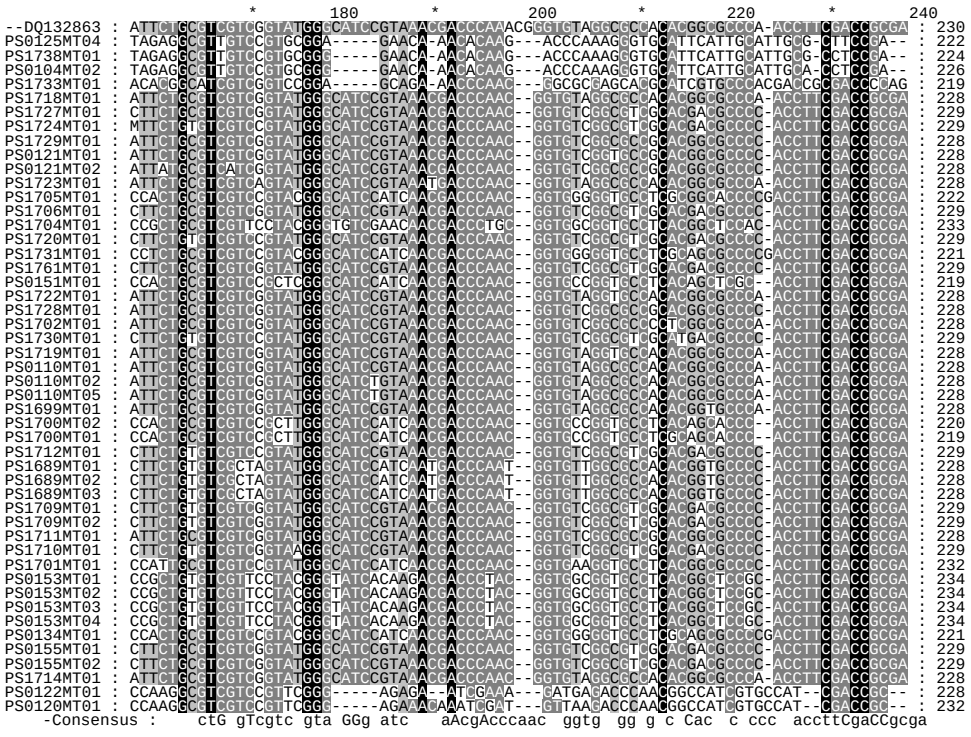


Fig. 1 Multiple sequence alignment of the forty-eight ITS2 sequences used in this study

The sequences were aligned using Clustalw (Larkin *et al*, 2007). The sequence DQ132863 is a previously known ITS2 sequence and was used here as a template to define the structure of the newly isolated ITS2 sequences. The nucleotides in each column are shaded based on their level of conservation. Nucleotides conserved in at least 100%, 80%, and 60% of the specimens are shown as white characters in a dark background, white characters in a gray background, and black characters in a grey background, respectively. The base colors are black characters in a white background. The positions of the alignment are shown above the alignment. The number of the last nucleotide in each sequence is shown to the right of the sequence. The consensus sequences are shown below each block, with identical nucleotides in uppercase letters, the most conserved nucleotides in lowercase letters, and blanks indicate that a gap is most abundant at this position

Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). For constructing the ME tree, the evolutionary distances were computed using the Maximum Composite Likelihood method (Sneath and Sokal, 1973) and are expressed as the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Inter-change algorithm (Nei and Kumar, 2000) at search level 1. The neighbor-joining algorithm was used to generate the initial tree. The MP tree was also obtained using the close-neighbor interchange algorithm (Nei and Kumar, 2000) with search level 3, in which the initial trees were obtained with the random addition of sequences (10 replicates).

The ME consensus tree and the MP consensus tree are shown on the left and right sides of Fig. 2, respectively. Species from *Salvia* L. were separated from those genera of *Scutellaria*, *Ajuga*, and *Phlomis*. The overall topologies of the two trees are similar. The taxa were arranged in the same order as much as possible to facilitate comparisons of the topologies of the two trees. To the right of the ME tree, the sections (or subgenera when no sections were available) are shown. As shown, both the ME and MP trees support the notion that there are two clades: Clade Ⅰ and Clade Ⅱ. Clade Ⅰ exclusively contains specimens originating from inside of China, while Clade Ⅱ exclusively contains specimens originating from outside of China. Clade I can be further divided into sub-clades ⅠA, ⅠB, and ⅠC. Sub-clade ⅠA contains species from sect. Drymosphace and sub-genus *Allagospadonopsis*, while sub-clade ⅠB and ⅠC contain species from sect. *Eurysphace* and sect. *Notiosphace*, respectively. Clade Ⅱ can also be divided into two sub-clades. Sub-clade ⅡA contains species from sect. *Plethiosphace*, sect. *Eusphace*, subgenus *Jungia* L., and sect. *Aethiopis*, while sub-clade

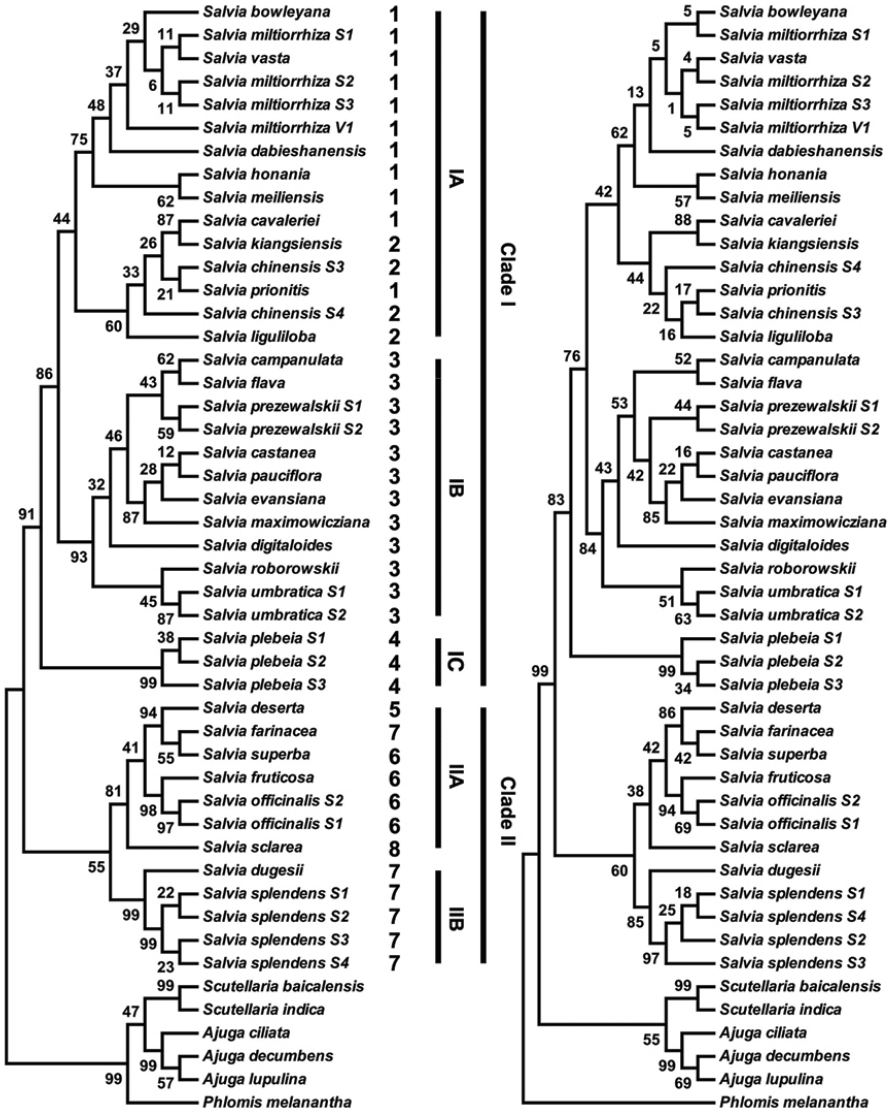


Fig. 2 Strict consensus phylogenetic trees constructed using the ME (left) and MP (right) methods based on the ITS2 sequences of 48 taxa of Lamiaceae

1. Sect. *Drymosphace* 2. Subg. *Allagospadonopsis* 3. Sect. *Eurysphace* 4. Sect. *Notiosphace*
5. Sect. *Plethiosphace* 6. Sect. *Eusphace* 7. Subg. *Jungia* 8. Sect. *Aethiopis*

Bootstrap values are shown above the branches. The section or the subgenus to which these taxa belong are shown between the two trees. Based on the trees, these taxa were divided into two clades (Ⅰand Ⅱ), which have three (ⅠA, ⅠB, ⅠC) and two subclades (ⅡA, ⅡB), respectively

ⅡB only contains species from subgenus *Jungia* L. The separation of various species at the clade level is reasonably well-supported, with bootstrap scores greater than 90, but the support for deeper level separations is poor. Furthermore, we found that *S. chinensis* and *S. miltiorrhiza* are not monophyletic. Although further investigation on the specimens to ensure their correct authentication will be required, these results are not surprising. A previous study has shown that *Salvia* L. is not monophyletic (Walker *et al*, 2004). In summary, the molecular phylogeny reconstructed using the ITS2 sequences is consistent with the taxonomical classification of the species in *Salvia* L. based on their morphologies, supporting the notion that ITS2 is suitable for phylogenetic comparisons. In fact, the ITS2 molecular phylogeny helped us to correct the misidentification of certain species in *Salvia* L. For example, we suspected that one *S. officinal* species used in our institute was *S. superba* based on the analysis of its ITS2 sequence. Further examination of the morphologies and chemical compositions confirmed that this species was misidentified and indeed should be classified as *S. superba*. This provided a practical example of using ITS2 sequence as a DNA barcode for

both species identification and species discovery.

Genetic distance analysis

Next we determined the genetic distances between specimens with regard to three different categories: those belonging to the same species (intra-species), those belonging to different species of the same genus (inter-species intra-genus), and those belonging to a different genus (inter-genus) (Table 3). Column 2 shows the pairs of specimens that were compared. Column 3 shows the number of pairs of specimens included in the comparison. The means and standard deviations of the distances are shown in the next columns. We used either percent dissimilarity (*p*) or the Kimura-2-parameter DNA substitution model (Kimura 2) to calculate the distances. As shown in columns 4 and 5, the average intra-species *p*-distance ranged from 0 to 3.1. The average inter-species intra-genus *p*-distance ranged from 4.53 to 11.06, while the average inter-genus *p*-distance ranged from 24 to 36.09. As expected, the average *p* distances followed the order: intra-species < inter-species intra-genus < inter species inter-genus. Similar patterns were observed for the corresponding average distances calculated using the Kimura-2-parameter model (Columns 6 and 7). One interesting observation is that the average inter-species genetic distance (11.06 +/— 6.26) is significantly higher than that of the intra-species distance of specimens in *Salvia* L., with the maximum being 3.10 for *S. chinensis.* This suggests that the inter-species variations of the ITS2 sequence are significantly smaller than its intra-species variations. As a result, ITS2 sequences can be used to assign a specimen in *Salvia* L. to the correct species with less likelihood of mis-assigning it to a different species.

Construction of a chemical phylogeny

Similarly to our use of ITS2 sequences to construct a molecular phylogeny, these chemical profiles were used to construct a chemical phylogeny. The data were pre-processed in order to replace the missing data with regard to noise levels and detection limits (see materials and methods). The resulting data matrix (supplementary Table 1) was then subjected to two-way clustering using the JMP software (Fig. 3). The chemical phylogeny showed several patterns that are similar to the molecular phylogeny (Fig. 2). For example, clusters 1 and 2 contain mostly specimens from the *Drymosphace* and *Eurysphace* sections, respectively, and these correspond to the subclades ⅠA and ⅠB. Similarly, cluster 3 contains specimens from sect. *Eusphace* and subgenus *Jungia* L., corresponding to clade ⅡB in Fig. 2. The relationships among other specimens were not as obvious as those observed in the molecular phylogeny. In addition to studying the relationships of the specimens based on the chemical profiles, we also investigated the relationships of the compounds based on their distribution profiles

Table 3 Intra- and inter-species variations observed in the ITS2 sequences under study

Groups	Samples compared	No.	*p*		Kimura 2	
			Mean	Std	Mean	Std
intra-species	*S. chinensis* samples *vs S. chinensis* samples	1	3.10	*NA*	3.16	*NA*
intra-species	*S. miltiorrhiza* samples *vs S. miltiorrhiza* samples	6	0.74	0.54	0.29	0.25
intra-species	*S. officinalis* samples *vs S. officinalis* samples	1	0.46	*NA*	0.46	*NA*
intra-species	*S. plebeia* samples *vs S. plebeia* samples	3	0.00	0.00	0.00	0.00
intra-species	*S. prezewalskii* samples *vs S. prezewalskii* samples	1	0.00	*NA*	0.44	*NA*
intra-species	*S. splendens* samples *vs S. splendens* samples	6	0.00	0.00	0.00	0.00
inter-species	*S. umbratica* samples *vs S. umbratica* samples	1	0.00	*NA*	0.00	*NA*
inter-species intra-genus	*Ajuga* species *vs Ajuga* species	3	4.53	1.64	4.70	1.76
inter-species intra-genus	*Salvia* species *vs Salvia* species	842	11.06	6.26	12.65	7.47
inter-species intra-genus	*Scutellaria* species *vs Scutellaria* species	1	8.97	*NA*	9.36	*NA*
inter-species intra-genus	*Ajuga* species *vs Phlomis* species	3	26.62	0.81	29.82	0.54
inter-species intra-genus	*Ajuga* species *vs Salvia* species	126	36.09	1.13	44.97	2.45
inter-species intra-genus	*Phlomis* species *vs Salvia* species	42	30.38	1.51	39.30	3.06
inter-species intra-genus	*Scutellaria* species *vs Ajuga* species	6	27.15	0.57	34.82	1.18
inter-species intra-genus	*Scutellaria* species *vs Phlomis* species	2	24.00	0.52	32.14	1.06
inter-species intra-genus	*Scutellaria* species *vs Salvia* species	84	31.92	1.31	45.61	2.78

Two distance metrics, *p* and Kimura 2, were used to calculate the distances among individuals of the same species (intra-species), among specimens that belong to different species of the same genus (inter-species intra-genus), and among specimens that belong to different genera (inter-genus). The mean and standard deviation of the distances of all pair-wise comparisons in the corresponding group are shown

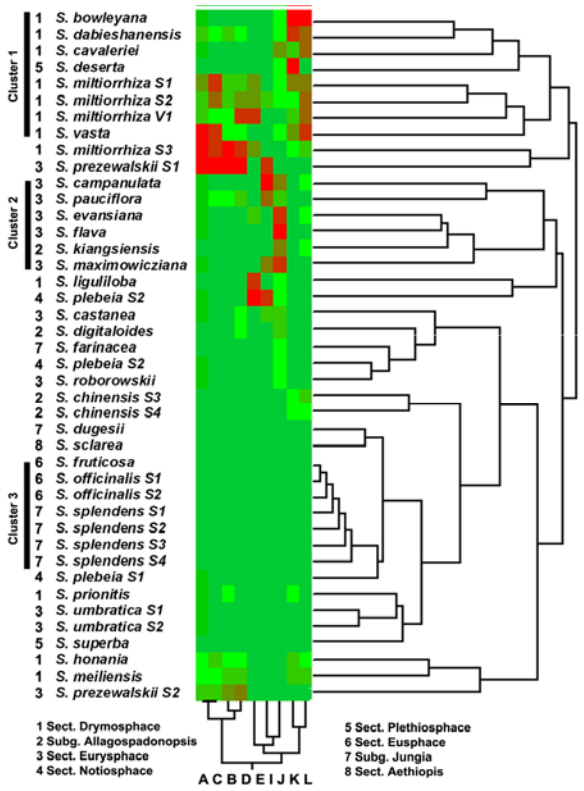


Fig. 3 The chemical phylogeny constructed using compounds identified in HPLC experiments

Nine different compounds (A–M) and their relative abundances in the corresponding specimens were determined using HPLC. Two-way hierarchical clustering analyses, which cluster the specimens and the compounds simultaneously, were carried out using the JMP software. The corresponding principle components are presented by "c1" to "c5". The compounds are A (dihydrotanshinone I), B (cryptotanshinone), C (tanshinone I), D (tanshinone IIA), E (danshensu), F (procatechuic acid), G (procatechuic aldehyde), H (chlorogenic acid), I (caffeic acid), J (rosmarimic acid), K (lithospermic acid), L (salvianolic acid B), and M (salvianolic acid A)

across the specimens. We calculated the pair-wise Pearson correlation coefficients of the distribution profiles of the nine compounds. Four compounds (A, B, C, and D) had very high correlation coefficient scores, suggesting that they have very similar distribution profiles. It is possible that these compounds are produced through some common metabolic pathway that is shared among all these specimens in *Salvia* L. (data not shown).

Comparison of the genetic and chemical phylogenies

One of the key questions is how well the molecular and chemical phylogenies correlate with each other. Because multiple ITS2 sequence sites and multiple chemical compounds are involved, multivariate analysis methods are most appropriate for answering this question. The ITS2 sequence data and the HPLC fingerprinting data for each specimen were joined. The resulting data matrix consists of two blocks of variables (*X* and *Y*). The *X* block is composed of 152 *X* variables, each corresponding to an ITS2 sequence site, and the *Y* block is composed of 9 *Y* variables, each corresponding to a chemical compound. Our hypothesis is that some *X* variables (ITS2 sequence sites) correlate with some *Y* variables (chemical compounds) as mediated by principle components for the *X* and *Y* blocks, respectively. To test this hypothesis, we first determined whether or not there are any principle components for the *X* and *Y* variables using PCA. Fig. 4A is the score plot for the *X* variables, which shows the specimen distribution in the *X* principle component space. This shows that specimens of *Salvia* L. form five clusters (arrows) in the score plot, suggesting the existence of a few principle components that can explain a significant portion of the variation seen at all ITS 2 variable sites. Fig. 4B shows the overview plot for the PCA-*X* model, that is, the PCA model built using the *X* variables. The goodness-of-fit of a PCA model is measured using two parameters: R2 and Q2. R2 is the percentage of variation in the data set that is explained by the PCA model. Thus, R2 is a measure of the degree of fit between the model and the data. A large R2 (close to 1) is a necessary but not sufficient condition for a good model. Q2 is the percentage of variation in the data set that can be predicted by the model according to cross validation. Q2 indicates how well the PCA model predicts new data. A large Q2 (Q2 > 0.5) indicates good predictivity. As shown in Fig. 4B, a good PCA model was obtained for the *X* block. This suggests that a few principle components (up to five as seen in Fig. 4B) can explain > 80% of the total variation observed among the ITS2 sequences with > 60% predictivity.

PCA analysis was then performed for the *Y* variables, and the score plot and model overview plot are shown in Fig. 4C and 4D, respectively. As shown in Fig. 4C, the specimens are scattered in the principle component space. In addition, Fig. 4D shows that Q2 decreases when R2 increases, which means that the PCA-*Y* model has very low predictivity. These observations suggest that the *Y* variables are very noisy.

After having evaluated the principle components of the *X* and *Y* variables, PLS analysis was performed to determine if the *X* variables correlate with the *Y* variables. A PLS model can be thought of as identifying the

principle components of *X* and *Y* variables simultaneously (different from the principle components described above, which were calculated using *X* and *Y* variables independently), so that variations in the *Y* variables can be best explained by those observed in the *X* variables. Fig. 4E and 4F are the score plot and model overview plot for the PLS model. Fig. 4E shows that the *X* and *Y* variables are very poorly correlated. Fig. 4F shows that the

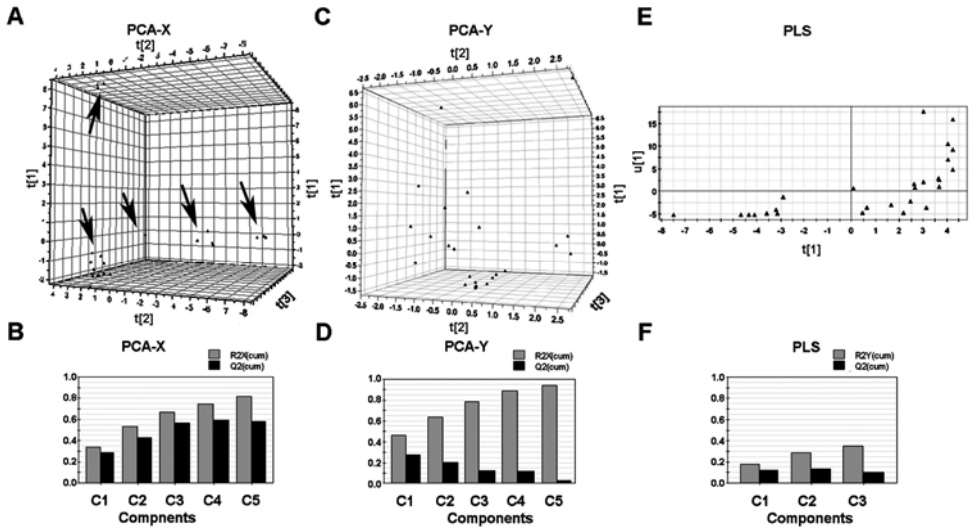


Fig. 4 Multivariate analyses of the ITS2 sequences and the chemical profiles

Ninety-two variable sites in the ITS2 sequences were considered as the *X* variables, and the abundances of the nine chemical compounds were considered as the *Y* variables. PCA was carried out for the *X* and *Y* variables, and a PLS analysis was carried out for the *X* and *Y* variables together. (A) PCA score plot for the *X* variables; (B) overview plot for the PCA-*X* model; (C) PCA score plot for the *Y* variables; (D) overview plot for the PCA-*Y* model; (E) score plot for the PLS model; (F) overview plot for the PLS model

PLS model neither fits the data well (R2 < 0.4 given three principle components) nor predicts the data well (Q2 decreases with the addition of principle components and becomes less then 0.1 given three principle components) in cross-validation. A poor Q2 is obtained when the data have much noise (with PCA), when the relationship *X* > *Y* is poor (with PLS), or when the model is dominated by a few scattered outliers. We think that the most likely reasons for the failure to obtain a good PLS model include the following: 1) the chemical compound data are very noisy (are highly variable); and/or 2) the ITS2 sequences (*X* variables) do not correlate well with the chemical distribution profiles (*Y* variables).

As an alternative approach, a tree comparison method was used to compare the molecular phylogeny and the chemical phylogeny. The chemical phylogeny tree shown in Fig. 3 was converted to the phylip (Felsenstein, 2005) tree format and TopD software (Puigbo, Garcia-Vallve, and McInerney, 2007) was used to compare the genetic phylogeny and the chemical phylogeny. Our results show some statistically significant similarities between the two phylogenies (data not shown). In conclusion, although the molecular and chemical phylogenies share obvious similarities, their similarity is not statistically significant based on the results obtained from PCA, PLS, and tree comparison analyses.

Identification of haplotype tagging SNPs

We next asked what minimal set of variable sites is both necessary and sufficient to discriminate the set of the specimens in *Salvia* L. To answer this question, the 42 ITS2 sequences from the specimens of *Salvia* L. (Table 1) were retrieved from the 48 ITS2 sequences and subjected to multiple sequence alignment. This generated an alignment 238 bp in length, among which 92 sites are non-ambiguous variable sites. These 92 sites were then analyzed using the BEST software, which was designed to identify haplotype tagging SNPs (htSNP). Please note that each variable site is equivalent to a SNP in the context of this study, and the two terms will be used interchangeably hereafter. The alignment of the 92 variable sites is shown in Fig. 5. On top of the alignment, the number of SNPs and the SNP types are shown. The SNPs were numbered from 1 to 92 based on their order in the sequence alignment. In terms of type, three types of SNPs were identified. The first type, indicated with a number in the SNP type field, is the binary equivalent SNP, which are redundant SNPs that have equivalent SNPs. For example, SNP 11 shows “4” in the SNP type. This means

that SNP 11 is equivalent to SNP 4, thus SNP 11 can be determined if SNP 4 has been determined. The second type of SNP is the derived SNP, indicated with an "x". These SNPs have no equivalent SNPs, but the information contained in these SNPs can be derived from the information contained in other SNPs. The dependency chart of these SNPs, that is, how these SNPs can be derived from other SNPs will be provided upon request. The third type of SNP is the haplotype tagging SNP, indicated with a blank in the SNP type field. The htSNPs have no equivalent SNPs and can not be deduced from information contained in the other SNPS. The minimal number of htSNPs was found to be 14, which includes the following SNPs: 17, 22, 34, 35, 38, 39, 59, 63, 70, 80, 81, 84, 88, and 89 (Fig. 5). This set of htSNPs is necessary and sufficient for the identification of the specimens in *Salvia* L. under this study.

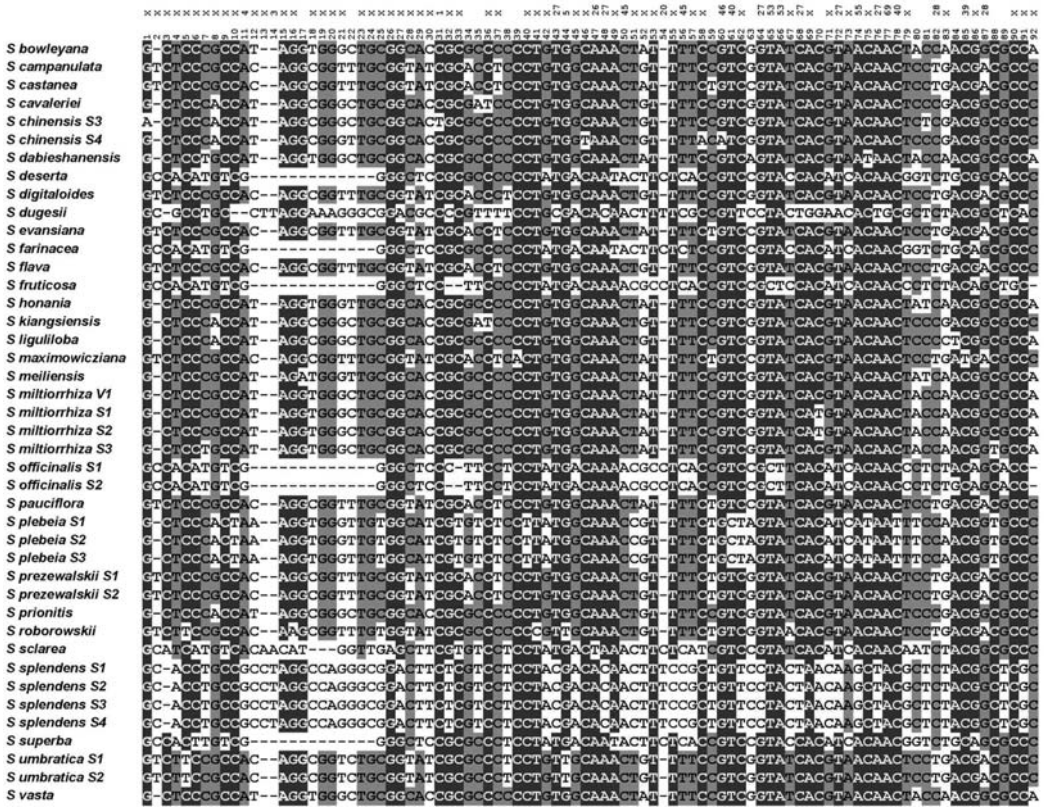


Fig. 5 Identification of a minimal set of SNPs

Ninety-two variable sites in the 243 bp sequence alignment of the ITS barcode were extracted and presented here. The nucleotides in each column are shaded based on their level of conservation as described in Fig. 1. The specimen names are shown to the left of the alignment. Each row represents a haplotype and each column represents a SNP. At the top of the alignment, the number of SNPs and the SNP type are shown. The SNPs are of three different types: binary equivalent SNPs (indicated with the number of the SNP to which the SNP is equivalent), derived SNP (indicated with "x"), and haplotype tagging SNP (for which the field was left blank)

Discussion

Various species of *Salvia* L. have been used extensively as components of traditional medicines, including *S. digitaloides* Diels, *S. przewalskii*, *S. yunnanensis* C.H.Wright, *S. miltiorrhiza* Bunge var. *miltiorrhiza f. alba* C. Y. Wu et H. W. Li and *S. miltiorrhizae*. These species have distinct pharmacological activities and their authentication is critical for the effective and safe application of Chinese materia medica products that contain components derived from these plant materials. One of the well-known problems is the difficulty of discriminating closely related species solely based on their morphological characteristics. This is due to the following facts: 1) closely related species usually have subtle morphological differences that are discernible only by experts, and 2) environmental factors may affect the morphologies of the plants. As a result, individual plants might develop phenotypes that deviate from their commonly observed ones. Because of these problems, morphological analysis alone cannot guarantee

the correct authentication of plant materials. Needless to say, processed plant materials cannot be authenticated by morphology at all because of the change of physical appearance and properties during processing. In the current study, we isolated and characterized ITS2 sequences for use as a DNA barcode for the identification of the species in *Salvia* L.

Until now, most discussions on DNA barcodes have focused on the identification of a known species by comparing its barcode sequences to those in a reference library. In these cases, searching the reference library will simply give a "yes" or "no" answer. However, for a novel species whose standard barcode is not in the reference library, it would be ideal if the DNA barcode could also serve as a phylogenetic marker that places the species in the correct position in the phylogenetic tree. In such cases, it is important to test the usefulness of a DNA barcode for phylogenetic analysis. Using ITS2 sequences, we constructed a molecular phylogeny of these species, and this was highly consistent with the morphological phylogeny (Fig. 2). This shows that ITS2 is a good phylogenetic marker that can also be used to identify the novel species of *Salvia* L., whose ITS2 sequences are not available in the reference library.

An extensive chemical profiling study was carried out for this set of specimens in *Salvia* L. (Li *et al*, 2008a). This provided us with an opportunity to compare the molecular and chemical phylogenies (Li *et al*, 2008a). The underlying assumption is that ITS2 sequences might co-evolve with genes encoding the enzymes involved in the metabolic pathways that produce these compounds. As a result, the evolution of ITS2 sequences might correlate with the differentiation of chemical profiles. The goal is to evaluate to what extent variations in DNA barcode sequences might be able to predict variations in the chemical compositions of individual plants. For this purpose, we constructed a chemical phylogeny, and this showed similarity to the molecular and morphological phylogenies. However, PCA, PLS, and tree comparison analyses could not identify any statistically significant correlation between these phylogenies. This could be due to several reasons. First, the ITS2 sequences might not co-evolve with the genes involved in the metabolism of these secondary metabolites used to define the chemical profiles. Second, environmental factors may have significantly affected the chemical compositions. As a result, the noise levels in the chemical profiles would be too high.

While various coding and intergenic DNA regions have been proposed as potential DNA barcodes, but methods for defining the exact regions required for DNA barcoding have not been investigated extensively. In the last part of our analysis, we subjected the full-length ITS2 sequence to htSNP identification. The set of htSNPs we obtained comprise a minimal set of sites that are necessary and sufficient for species identification in a specific taxonomic group. This opens up a venue to further improve DNA barcoding technology. For example, a pilot study could be used to identity the htSNPs, and then a minimal barcode that covers all htSNPs could be specified. One could still use DNA sequencing technology to obtain the sequence of this minimal barcode. Obtaining the minimal barcode sequence might be less challenging and expensive, since it should be shorter than the original DNA barcode. Alternatively, more specific techniques, such as those used to detect genotyping SNPs, could be introduced into this species identification area.

A number of conclusions can be drawn from this study:

1) The ITS2 region is suitable for DNA barcoding study because it has conserved flanking regions for primer design, and it has sufficient genetic variation to differentiate closely related species regardless if they come from the market or are cultivated from seed leaves, flowers, or herbs in their early life stages. Owing to their short length, however, they could be amplified from processed plant materials. The limitation of DNA barcoding is that it cannot be used to assess the quality of crude drug. ITS2 was presented here as a promising phylogenetic tool. This sequence has proven to be important for the identification process, and it turned out to be a useful marker for studying systematics.

2) The chemical profile determined by HPLC could serve as a fingerprint for the quality control of the species in *Salvia* L. Herbs usually contain up to hundreds or even thousands of different phytochemicals. Many factors may affect the ultimate chemical profile of any herb. Routine chemotaxonomic studies provide only a qualitative account of secondary metabolites. The chemical composition of the sect. *Eusphace* Benth was very similar. In addition, the chemical composition of the species in *Salvia* L. could be varied greatly from its habitats.

Thus, chemical composition analysis could only be used to determine the quality of the samples rather than distinguish adulterants. Clearly, variations in ITS2 DNA sequences provide more reliable information for the latter purpose.

3) As a result, DNA barcoding technology and HPLC fingerprinting technology can complement each other in determining the identity and chemical composition of a plant specimen. The ITS2 sequence combined with the HPLC fingerprint could not only be developed to authenticate the species of *Salvia* L. but also could be used to optimize authentic location. Additionally, to explore unknown species in a given taxonomic group, DNA barcoding might be used for species identification on a higher taxonomic level.

Acknowledgements

This work was supported by the International Cooperation Program of Science and Technology (No. 2007DFA30990) and the Special Founding for Healthy Field (No. 200802043) awarded by the Chinese Ministry of Science and Technology. This work was also supported by grants from the Hong Kong Research Grant Council (HKU 7526/06M) to C.L. Thank our colleagues in the Institute of Medicinal Plant Development who have helped in sample collection, identification, laboratory work, and manuscript preparation, including Profs. LIN Yu-lin, LI Xi-wen, ZHANG Zhao, and Drs. MA Xin-ye, LUO Kun, LI Ying and WU Qiong. We are thankful to Dr. LI Jian-ying from University of North Carolina for helpful suggestions regarding the methods of data analyses.

References

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ, 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on Angiosperm ahylogeny. *Ann Mo Bot Gard* 82: 247.

Burland TG, 2000. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol* 132: 71-91.

Cao Z, Xie XL, 2007. The research of various methods to distinguish *Salvia miltiorrhiza* Bulge. *Lishizhen Med Mater Med Res* 5(8): 1861-1863.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V, 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc B Biol Sci* 360: 1889-1895.

Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li Y, Li XW, Jia XC, Lin YL, Christine L, 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5(1): e8613. doi:10.1371/journal.pone.0008613

Chiou S, Yen J, Fang C, Chen H, Lin T, 2007. Authentication of medicinal medicinal herbs using PCR-amplified ITS2 with specific primers. *Planta Med* 73: 1421.

Coleman AW, 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* 19: 370-375.

Felsenstein J, 2005. PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington, Seattle.

Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003. Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270: 313-321.

Kimura M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111-120.

Kress WJ, Erickson DL, 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS one* 2(6): 1-10.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH, 2005. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102: 8369-8374.

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V, 2008. DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA* 105: 2923.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.

Li MH, Chen JM, Peng Y, Wu Q, Xiao PG, 2008a. Investigation of Danshen and related medicinal plants in China. *J Ethnopharmacol* 120: 419-426.

Li MH, Chen JM, Peng Y, Xiao PG, 2008b. Study on the distribution regularity of water-solution bioactive components of *Salvia* L. in China. *World Sci Technol* 10: 46-52.

Miao M, Warren A, Song W, Wang S, Shang H, Chen Z, 2008. Analysis of the internal transcribed spacer 2 (ITS2) region of *Scuticociliates* and related taxa (Ciliophora, Oligohymenophorea) to infer their evolution and phylogeny. *Protist* 159: 519-533.

Nei M, Kumar S, 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, USA.

Puigbo P, Garcia-Vallve S, McInerney JO, 2007. TOPD/FMTS: a new software to compare phylogenetic trees. *Bioinformatics* 23: 1556.

Sebastiani P, Lazarus R, Weiss ST, Kunkel LM, Kohane IS, Ramoni MF, 2003. Minimal haplotype tagging. *Proc Natl Acad Sci USA* 100: 9900-9905.

Sneath P H A, Sokal RR, 1973. *Numerical Taxonomy*. Springer.

Tamura K, Dudley J, Nei M, Kumar S, 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596.

Walker JB, Sytsma KJ, Treutlein J, Wink M, 2004. *Salvia* (Lamiaceae) is not monophyletic: implications for the systematics, radiation, and ecological specializations of Salvia and tribe Mentheae 1. *Am J Bot* 91: 1115-1125.

Xiao XH, Fang QM, Xia WJ, 1997. Numerical taxonomy of medicinal *Salvia* L. and the genuineness of Danshen. *J Plant Res* 6: 17-21.