



Metabolic fingerprinting for discrimination of DNA-authenticated *Atractylodes* plants using ^1H NMR spectroscopy

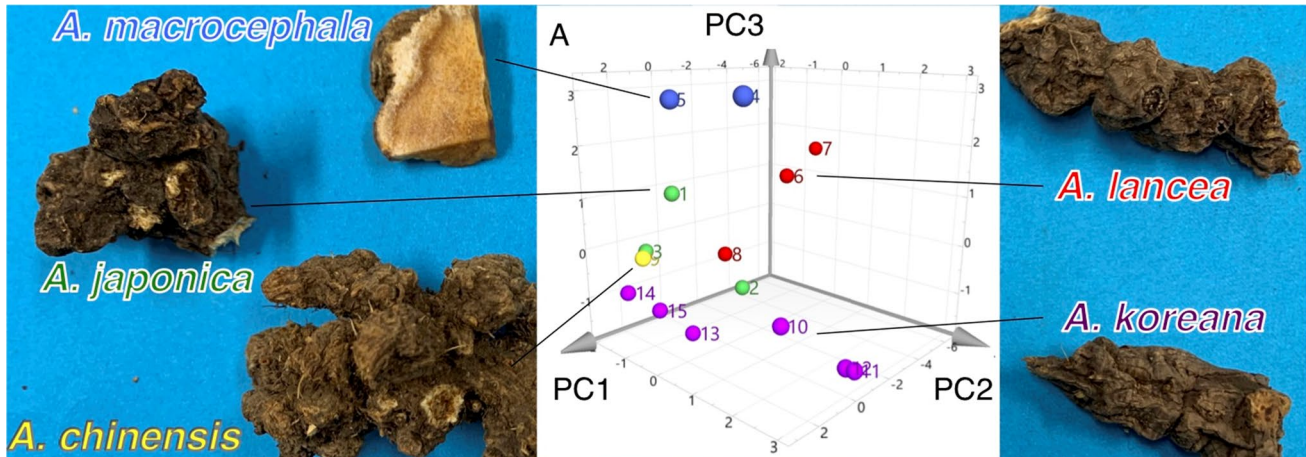
Tatsuya Shirahata^{1,2} · Hiroshi Ishikawa¹ · Teruhisa Kudo¹ · Yumiko Takada¹ · Azusa Hoshino¹ · Yui Taga¹ · Yusaku Minakuchi¹ · Tomoko Hasegawa¹ · Rina Horiguchi¹ · Takehiro Hirayama¹ · Takahiro Konishi¹ · Hiroaki Takemoto¹ · Noriko Sato¹ · Masako Aragane³ · Tetsuro Oikawa² · Hiroshi Odaguchi² · Toshihiko Hanawa² · Eiichi Kodaira¹ · Tatsuo Fukuda¹ · Yoshinori Kobayashi^{1,2}

Received: 24 August 2020 / Accepted: 18 November 2020 / Published online: 11 February 2021
© The Author(s) 2021, corrected publication 2022

Abstract

Identifying different species of the genus *Atractylodes* which are commonly used in Chinese and Japanese traditional medicine, using chromatographic approaches can be difficult. ^1H NMR metabolic profiling of DNA-authenticated, archived rhizomes of the genus *Atractylodes* was performed for genetic and chemical evaluation. The ITS region of the nuclear rDNA was sequenced for five species, *A. japonica*, *A. macrocephala*, *A. lancea*, *A. chinensis*, and *A. koreana*. Our samples had nucleotide sequences as previously reported, except that part of the *A. lancea* cultivated in Japan had a type 5, hybrid DNA sequence. Principal component analysis (PCA) using ^1H NMR spectra of extracts with two solvent systems (CD_3OD , CDCl_3) was performed. When CDCl_3 extracts were utilized, the chemometric analysis enabled the identification and classification of *Atractylodes* species according to their composition of major sesquiterpene compounds. The ^1H NMR spectra using CD_3OD contained confounding sugar peaks. PCA removal of these peaks gave the same result as that obtained using CDCl_3 and allowed species distinction. Such chemometric methods with multivariate analysis of NMR spectra will be useful for the discrimination of plant species, without specifying the index components and quantitative analysis on multi-components.

Graphic abstract



Keywords *Atractylodes* plants · ITS sequence · ^1H NMR spectroscopy · Metabolic profiling

Tatsuya Shirahata, Hiroshi Ishikawa contributed equally to this work.

Extended author information available on the last page of the article

Introduction

Atractylodes rhizome (*Atractylodis rhizoma*, “Byakujutsu” in Japanese) and *Atractylodes lancea* rhizome (*Atractylodis lanceae rhizoma*, “Sojutsu” in Japanese) are often used as crude drugs in traditional Chinese medicine (TCM) and traditional Japanese medicine (Kampo). These plants, belonging to the genus *Atractylodes* and the family Compositae, are perennial herbs distributed in East Asia [1]. These two crude drugs have been used for the opposite clinical purposes in Kampo medicine; Byakujutsu for diaphoretic activity and Sojutsu has antisudorific activity [2, 3].

In the Japanese Pharmacopoeia, *Atractylodes* rhizome is defined as the dried rhizomes of *A. japonica* Koidzumi ex Kitamura and *A. macrocephala* Koidzumi (syn.: *A. ovata* De Candolle), whereas *Atractylodes lancea* rhizome is defined as the dried rhizomes of *A. lancea* De Candolle, *A. chinensis* Koidzumi, and the hybrid between those two species [4]. *A. lancea*, *A. chinensis*, *A. macrocephala*, and *A. koreana* grow wild in Northern China from the midstream reaches of the Yangtze River, northeast toward Inner Mongolia, while *A. japonica* is native to Japan. In the Korean Peninsula, *A. japonica* and *A. koreana* (Nakai) Kitamura grow wild [1]. The rhizome of *A. koreana* is sometimes used as a medicine in limited areas of China [5].

In the Japanese Pharmacopoeia, the two crude drugs are discriminated based on their morphological features [5] as well as by detection of the sesquiterpenoid atractylon [6], which is found only in *Atractylodes* rhizome (*A. macrocephala* and *A. japonica*). However, some varieties of *A. chinensis* have also been found to contain atractylon [7]. In addition, it is difficult to discriminate *A. lancea* and *A. chinensis* because of their overlap in morphological and chemical features [8, 9]. The differentiation of phenotypic trait between *Atractylodes* plants, is affected by environmental factors, such as the developmental stage [10], season of harvest [11], geographic origin of the plant [12], as well as by intra-specific variation [13–15], that makes the discrimination of *Atractylodes*-derived crude drugs difficult. By the use of genetic markers, four species of *Atractylodes* (*A. japonica*, *A. macrocephala*, *A. lancea*, and *A. chinensis*) were distinguished by their internal transcribed spacer (ITS) sequences. However, *A. koreana* could not be distinguished from *A. chinensis* by the ITS sequences alone [16].

Chromatographic fingerprinting (e.g., with HPLC or GC), can be used to clarify the required chemical characteristics of the species [17, 18]. However, the pre-purification step for chromatographic fingerprinting cannot detect all the metabolites present in the crude drug.

NMR spectroscopy can solve the common problems of chromatographic approaches. In recent years, NMR-based

metabolomics technology has been successfully applied in many areas, including the analysis of herbal medicines, due to its unique advantages: (1) simple sample preparation, (2) high robustness, (3) simultaneous detection of primary and secondary metabolites, and (4) specific structural characterization [19–23]. The principal objective of this study was to establish a compound-based discrimination method for the rhizomes of five *Atractylodes* species, which are used in traditional Chinese and Japanese medicine. Thus, we performed ¹H NMR metabolic profiling of these *Atractylodes* species, comparing the authentication with DNA barcoding. We could differentiate the five *Atractylodes* species by comparing the nucleotide sequences of the ITS region. ¹H NMR metabolic profiling of the rhizome extracts with principal component analysis (PCA) also clearly discriminated these five *Atractylodes* species. We believe that our study will improve quality control of herbal medicine because our chemometric model (particularly the multivariate analysis using NMR measurements) will be useful for the assessment of these plant species and potentially others.

Results and discussion

Sequencing and phylogenetic analysis

The ITS sequences of specimens from the five *Atractylodes* species were determined and compared with previous data from the International Nucleotide Sequence Database (INSD) Table 1. Table 2 shows the nucleotide differences among various ITS sequences derived from *Atractylodes* plants. All sequences were of the same length, in which the ITS region, containing ITS1, the 5.8S rRNA gene, and ITS2, was 643 bp. The ITS sequences of No 1, 2, 3 plants were almost concordant with the Type 1 sequence (AB219405) in the previous study [12], which represents the *A. japonica* genotype; there were some differences among individuals, but they did not affect species definition. The ITS sequences of No 4, 5, No 6, 7, 8, and No 9 were concordant with the Type 2 (AB219406: *A. macrocephala*), Type 3 (AB219407: *A. lancea*), and Type 4 (AB219408: *A. chinensis* and AB219409: *A. koreana*) sequences, respectively. The ITS sequences of No 10, 11, 12 and No 13, 14, 15, morphologically identified as *A. koreana*, were almost identical to the Type 4 sequence and were indistinguishable from *A. chinensis* by the ITS genotype, as reported previously [12].

The ITS sequences of the isolated haplotypes from the No 1, 2, 3 were classified into two types, Type 1A and Type 1B, based on sequence similarity. Each individual possessed both types of ITS sequences. The Type 1 sequence can be interpreted as a combination of the two types. This suggests that *A. japonica* is of hybrid origin between unknown

Table 1 Plant specimens used in this study. The *Atractylodes* spp. used in this study were shown in different colors for each species: *A. japonica* in shown in green, *A. macrocephala* in blue, *A. lancea* in

red, *A. chinensis* in yellow, and *A. koreana* in violet. These colors are consistent with other figures and tables in this paper

No	Crude Drug	Japanese name	Species	Wild / Cultivated	Collected year	Locality of collection	ITS type
1	Atractylodes rhizome	Okera	<i>A. japonica</i>	C	1997	Kodaira city, Tokyo, Japan	1
2	Atractylodes rhizome	Okera	<i>A. japonica</i>	W	Unknown	Myougi, Gumma, Japan	1
3	Atractylodes rhizome	Okera	<i>A. japonica</i>	C	Unknown	Oume city, Tokyo, Japan	1
4	Atractylodes rhizome	Oobanaokera	<i>A. macrocephala</i>	C	1990	Kodaira city, Tokyo, Japan	2
5	Atractylodes rhizome	Oobanaokera	<i>A. macrocephala</i>	C	1994	Kodaira city, Tokyo, Japan	2
6	Atractylodes lancea rhizome	Hosobaokera	<i>A. lancea</i>	C	1990	Nanjing, China	3
7	Atractylodes lancea rhizome	Hosobaokera	<i>A. lancea</i>	C	1990	Nanjing, China	3
8	Atractylodes lancea rhizome	Hosobaokera	<i>A. lancea</i>	C	1990	Nanjing, China	3
9	Atractylodes lancea rhizome	Shiraokera	<i>A. chinensis</i>	C	1990	Beijing, China	4
10	-	Nanmanokera	<i>A. koreana</i>	W	Unknown	North Korea	4
11	-	Nanmanokera	<i>A. koreana</i>	W	Unknown	North Korea	4
12	-	Nanmanokera	<i>A. koreana</i>	W	Unknown	North Korea	4
13	-	Nanmanokera	<i>A. koreana</i>	C	2015	North Korea	4
14	-	Nanmanokera	<i>A. koreana</i>	C	2015	North Korea	4
15	-	Nanmanokera	<i>A. koreana</i>	C	2015	North Korea	4

Table 2 Comparison of ITS sequences among various samples derived from *Atractylodes* plants

Site number				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19									
Nucleotide position				39	41	51	74	87	90	91	98	105	111	112	113	119	127	129	140	147	192	196									
No	Japanese name	Species	Type																												
Accession Number AB219405				T	A	G	G	G	Y	G	G	Y	R	T	C	Y	T	C	T	R	A	A									
1,2	Okera	<i>A. japonica</i>	1A	T	A	A	G	G	C	C	G	C	A	T	C	C	T	C	T	A	A	A									
1,2	Okera	<i>A. japonica</i>	1B	T	A	G	C	G	C	C	G	T	G	T	C	T	T	C	T	G	A	A									
3	Okera	<i>A. japonica</i>	1A	T	A	R ^b	C	G	C	C	G	C	A	T	C	C	T	C	T	A	A	A									
3	Okera	<i>A. japonica</i>	1B	T	A	G	C	G	C	C	G	T	G	T	C	T	T	Y	T	G	A	A									
4,5	Oobanaokera	<i>A. macrocephala</i>	2	C ^a	A	A	C	G	C	A	G	C	G	C	A	C	C	C	C	G	A	G									
6,7,8	Hosobaokera	<i>A. lancea</i>	3	T	G	G	C	G	C	C	G	C	G	T	C	C	T	C	T	A	A	A									
9	Shiraokera	<i>A. chinensis</i>	4	T	A	G	T	A	C	C	R	C	G	T	C	T	T	C	T	G	C	A									
10,11,12	Nanmanokera	<i>A. koreana</i>	4	T	A	G	T	A	C	C	R	C	G	T	C	T	T	C	T	G	C	A									
13,14,15	Nanmanokera	<i>A. koreana</i>	4	T	A	G	T	A	C	C	R	C	G	T	C	T	T	C	T	G	C	A									
Site number				20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
Nucleotide position				198	229	230	249	386	389	426	435	446	449	451	459	465	472	494	499	500	504	514	515	542	569	584	620	628	629	630	636
No	G	C	G	T	A	C	G	W	C	C	Y	C	G	R	C	Y	R	Y	A	G	W	C	G	C	G	C	G	C	C		
1,2	G	C	G	T	A	T	G	A	C	C	T	C	G	A	T	T	A	T	A	G	A	C	G	C	C	G	C	C	C		
1,2	K	C	G	T	A	C	A	T	C	C	C	C	A	G	C	T	G	C	A	G	T	C	G	T	C	G	T	C	C		
3	G	C	G	T	A	C	G	A	C	C	T	C	G	A	C	T	A	T	A	G	A	C	G	C	C	G	C	C	C		
3	T	C	G	T	A	C	G	T	C	C	C	C	A	G	C	T	A	C	A	G	T	C	G	T	C	G	C	C			
4,5	G	C	R	T	A	C	G	C	C	C	C	G	C	C	C	G	C	C	C	S	T	T	G	C	C	T	G	C	Y		
6,7,8	G	G	G	T	A	C	G	A	C	C	C	C	G	G	C	T	G	C	A	G	A	C	G	C	C	G	C	C			
9	G	C	G	C	R	C	G	T	Y	Y	C	C	G	G	Y	C	G	C	A	G	T	C	R	C	C	A	C	C			
10,11,12	G	C	G	C	A	C	G	T	C	C	Y	C	G	G	C	C	G	C	A	G	T	C	G	C	C	A	C	C			
13,14,15	G	C	G	C	A	C	G	T	C	C	Y	Y	G	G	C	C	G	C	A	G	T	C	G	C	C	A	C	C			

^aGray background indicates nucleotide sites different from accession number AB219405

^bCharacters other than A, C, G, and T indicate nucleotide additives. R = A + G, W = A + T, S = C + G, Y = C + T, K = G + T

species with Type 1A and Type 1B sequences. Further investigations are needed to verify this hypothesis regarding the origin of *A. japonica*.

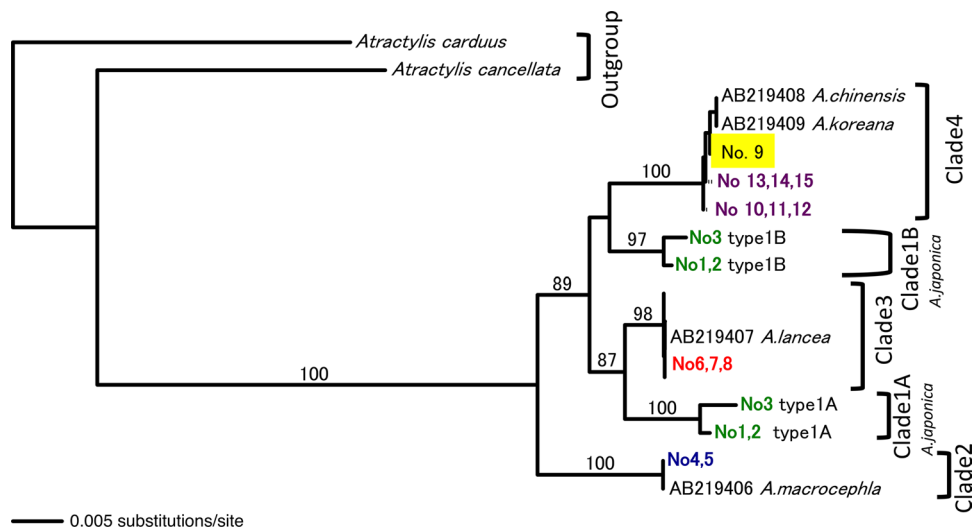
Phylogenetic trees based on the ITS sequences of *Atractylodes* species were constructed with three different methods [neighbor joining (NJ), maximum parsimony (MP), and maximum likelihood (ML)]. Figure 1 shows the NJ tree of the ITS sequences. Five clades (clade 1A, 1B, 2, 3, and 4) were resolved in the ingroup. These five clades correspond to the Type 1A haplotype of *A. japonica*, Type 1B haplotype of *A. japonica*, Type 2 sequence: *A. macrocephala*, Type 3 sequence: *A. lancea*, and Type 4 sequence: *A. chinensis* and *A. koreana*, respectively. All five clades are well supported with more than 97% bootstrap values. The MP and the ML trees showed the same topology as the NJ tree except for differences in the relationships among operational taxonomic units in clade 4, which is beyond the scope of the current study. Thus, the ITS sequence was sufficiently informative to discriminate four *Atractylodes* species (*A. japonica*, *A. macrocephala*, *A. lancea*, and *A. chinensis*). However, *A. chinensis* and *A. koreana*, could not be distinguished solely by the ITS sequence.

Identification of metabolites from *Atractylodes* plants

The ^1H NMR spectra of the CD_3OD extracts of the five species are shown in Fig. 2a. We assigned signals to seven compounds the chemical structures of which are shown in Fig. 3. The color cubic signs beside each chemical name relate to the corresponding buckets in the loading plot in PCA analysis (vide infra) in Fig. 3. Detailed peaks were assigned to the metabolites by comparisons to pure, isolated compounds as shown in Table 3, except sucrose

in CDCl_3 . These assignments included five sesquiterpenoids, atractylon (ATN), atractylenolide II (ATOII), atractylenolide III (ATOIII), hinesol (HIN), and β -eudesmol (EUD), a polyacetylenic compound, atractylodin (ATD), and a sugar, sucrose (SUC). In the ^1H NMR spectra using the CD_3OD extracts, the main peaks were assigned to sucrose (δ 5.35, δ 4.00–3.30). In addition, H-12 and H-13 of hinesol and β -eudesmol were identified as a singlet at δ 1.16. Expanded spectra in the lower field were required to clarify each characteristic peak (Fig. 2b). The H-12 peak of atractylon characteristically appeared as a singlet at δ 7.40. A polyacetylated compound, atractylodin, was also detected in the olefinic region at δ 7.49 (d, $J=1.5$ Hz), 6.84 (d, $J=16.0$ Hz), 6.48 (d, $J=3.3$ Hz), 6.47 (dd, $J=3.3, 1.5$ Hz), 6.32 (dd, $J=15.5, 6.9$ Hz), 6.10 (d, $J=16.0$ Hz), and 5.64 (ddd, $J=15.5, 1.8, 1.0$ Hz). In the CD_3OD spectra, only the above peaks were recognized in the standard compounds. In contrast, many characteristic peaks of each compound were identified in the CDCl_3 extract spectra (Fig. 2c). The H-15 peaks of the characteristic sesquiterpenoids in *A. japonica* and *A. macrocephala* were observed as doublets or doublets of doublets at δ 4.87–4.56, such as atractylon (δ 4.86, 4.70), atractylenolide II (δ 4.86, 4.56), and atractylenolide III (δ 4.87, 4.60). These types of buckets were called “sesquiterpenoid-1” and indicated by green squares. To identify *A. lancea* and *A. chinensis*, major sesquiterpenoids were observed, such as the H-1 of hinesol (δ 5.31), and H-15 of β -eudesmol (δ 4.72, 4.45), according to their characteristic signals (Table 3). These types of buckets were called “sesquiterpenoid-2” and shown by orange squares. The H-15 of atractylon (δ 7.05), H-12 and H-13 of hinesol and β -eudesmol (δ 1.21), and the olefinic proton of atractylodin (δ 7.49, 6.84, 6.48, 6.47, 6.32, 6.10, and 5.64) showed the same pattern as with the CD_3OD extract.

Fig. 1 The NJ tree based on the ITS sequences of *Atractylodes* species. The bootstrap values above 80% are shown along the branches



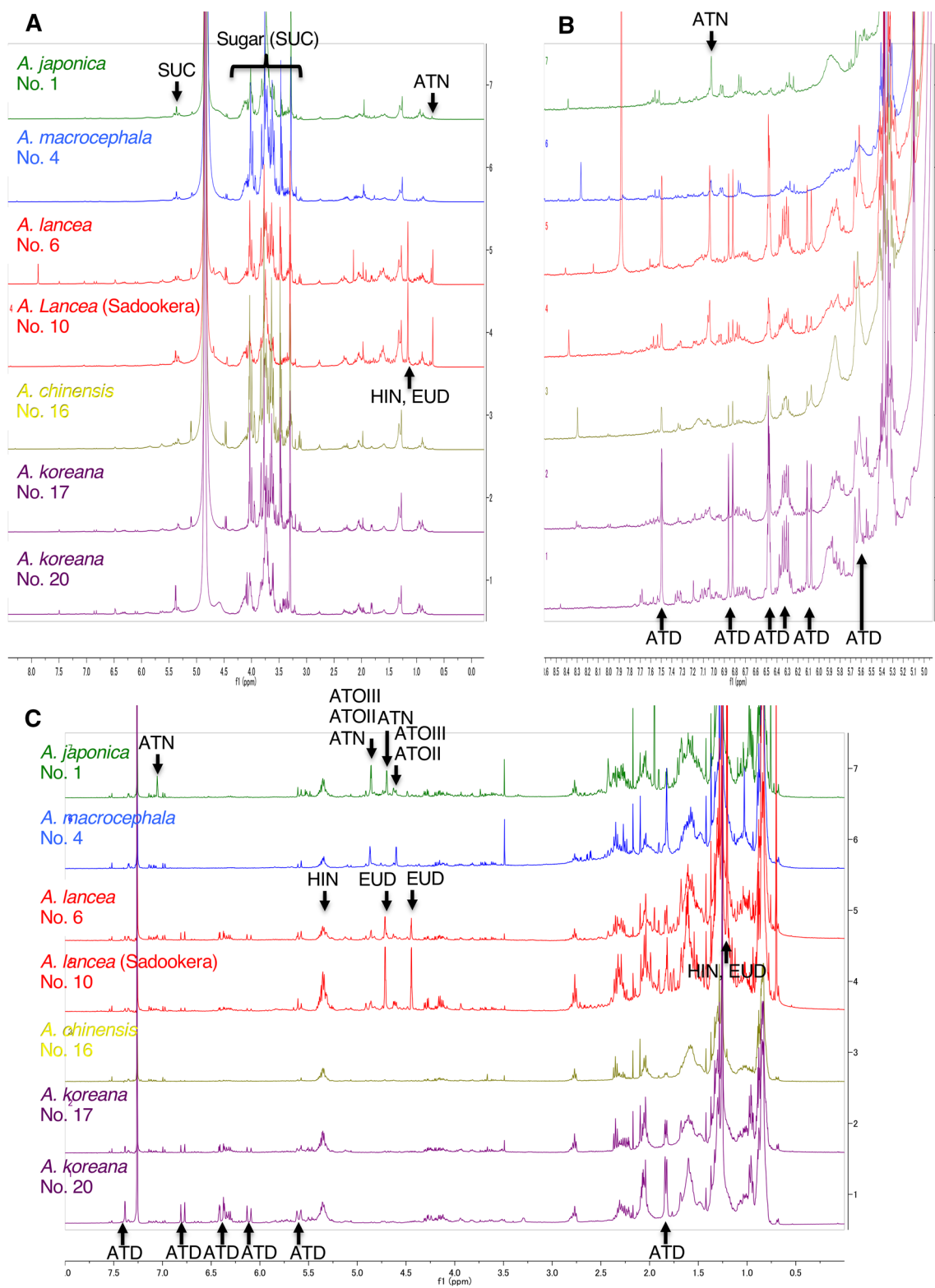


Fig. 2 Representative ¹H-NMR spectra of **a** CD₃OD and **c** CDCl₃ extracts of *Atractylodes* plants. **b** Expanded spectral region from 0.60 to 1.8 ppm of CD₃OD extract. Signal numbers follow those listed in Table 3 for metabolite identification using ¹H NMR

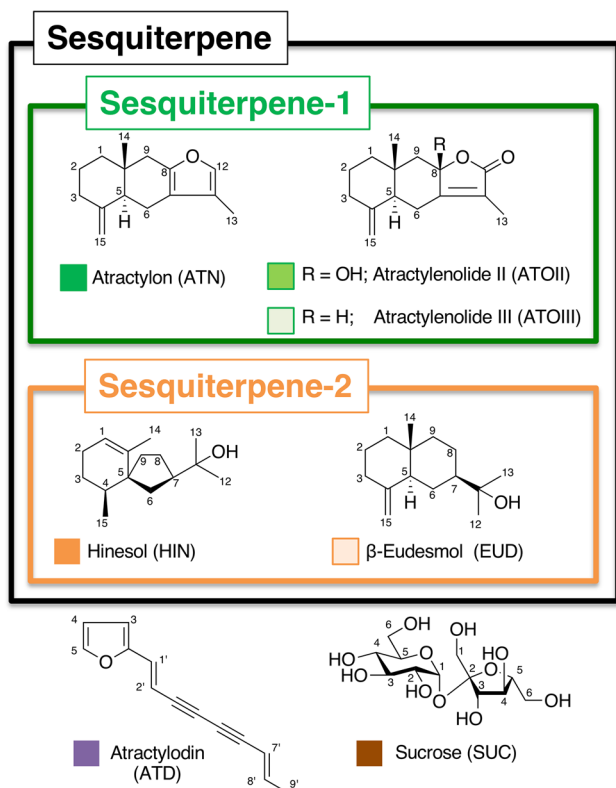


Fig. 3 Chemical structures of the major components identified from *Atractylodes* plants

Principal component analysis using ^1H NMR spectra of CD_3OD extracts (CD_3OD model)

Principal component analysis is a widely applied unsupervised method in multivariate data analysis that aims to reduce the dimensionality of a multivariate dataset [24]. In this study, PCA was used for species classification and identification of marker metabolites. The bucketed data derived from the ^1H NMR spectra of these samples were subjected to PCA. The data were scaled using the Pareto scaling method, which lies in-between non-scaling and the unit variance method [24, 25]. PCA extracted three significant principal components cumulatively accounting for 83% of the total variance (PC1 = 45%, PC2 = 26%, PC3 = 12%), which is shown in a 3D score plot (Fig. 4a), where each point represents an individual sample. A partial separation was observed between *A. lancea* (red) and the samples from the other four species. To understand the differentiation of *A. lancea* in more detail, the score plot of the first two principal components generated from the 15 samples is shown in Fig. 4b. Two samples of *A. macrocephala* (blue) were adjacently located. The *A. koreana* samples (purple) were not clustered as one group, but other groups were clustered based on their species. The loading plot was analyzed to clarify the buckets that contributed to form these groupings

(Fig. 4c). The buckets of sucrose (3–5 ppm), including the anomeric proton at 5.18 ppm, were spread on the far right side (PC1 positive value). The buckets defined as the sesquiterpene-2 group were found in the third quadrant, where the characteristic bucket from the protons of the dimethyl group at δ 1.16 ppm (s) was also found. Secondary metabolite components from the essential oil of *Atractylodes lancea* rhizome contributed to the grouping of samples from *A. lancea*. For the further consideration of this model, the score and loading plot of PC1 and PC3 were studied, which showed that sesquiterpene-1 and -2 were not clustered on the loading plot. The PC1 and PC3 data were showed in the Electronic supplementary material (ESM)(Figures S1, S2 and S3, vide infra). Unfortunately, these PCA models could not differentiate the plant species due to the confounding resonances of sucrose (3–5 ppm). As a second step, PCA was performed using the ^1H NMR spectra of compounds extracted with a low polar solvent to prevent detection of sucrose and allow the delineation of the plant species.

Principal component analysis using ^1H NMR spectra of CDCl_3 extracts (CDCl_3 model)

PCA using ^1H NMR spectra of *Atractylodes* samples extracted with CDCl_3 , which is unaffected by primary metabolites, was performed utilizing bucket integral values every 0.04 ppm in the range of 0.00–10.00 ppm. The results showed three significant principal components cumulatively accounting for 82% of the total variance (PC1 = 50%, PC2 = 21%, PC3 = 11%), which are presented in the 3D score plot in Fig. 6a. The 3D score plot in Fig. 5b, shows three major, distinct clusters corresponding to the different species studied. The samples of *A. koreana* (purple) were clustered in one group, unlike the CD_3OD extracts. The score plot of PC1/PC2 was analyzed to understand the details of the clustering (Fig. 5b). Plots of *A. koreana* were found in the second quadrant. The PC1 and PC2 loading plots were studied further to clarify potential metabolic markers contributing to the discrimination of the different species (Fig. 5c). In the fourth quadrant, buckets of 0.74 and 0.78 were confirmed corresponding to the proton at C14 (0.76 ppm, s) in atractylon, which is a specific compound found in *Atractylodes* rhizome. In contrast, in the third quadrant, there are buckets derived from protons at 6.11, 6.33, 6.37, 6.42, 6.79, and 7.38 ppm, which are considered to be inherent to atractylodin from *A. koreana* (shown in purple). In the fourth quadrant, the buckets of the sesquiterpene-2 group derived from *Atractylodes lancea* rhizome were involved. In addition, the protons of the dimethyl group were found to be 1.22 ppm (s) deviated from the center. From the above, it was possible to distinguish the species on the basis of PCA plots by detecting known, base-like compounds by NMR and performing multivariate analysis (PCA).

Table 3 $^1\text{H-NMR}$ chemical shifts (δ) and coupling constant (J , Hz) of *Atractylodes* plants

Number	Compound	CD_3OD		CDCl_3						
		Position	$^1\text{H-NMR}$	Position	$^1\text{H-NMR}$					
Atractylon	ATN	12	7.04	s		12	7.05	D	$J = 1.0$ Hz	
		15a	4.80	d	$J = 1.6$ Hz	15a	4.86	Dd	$J = 3.5, 1.6$ Hz	
		15b	4.71	d	$J = 1.6$ Hz	15b	4.70	Dd	$J = 3.2, 1.6$ Hz	
		3,6,9	2.42–2.02	m		3,6,9	2.44–2.00	M		
		13	1.92	s		13	1.95	D	$J = 1.0$ Hz	
		1,2,5	1.70–1.49	m		1,2,5	1.72–1.45	M		
		14	0.74	s		14	0.76	S		
	Atractylenolide II	ATOII	8	4.97	dd	$J = 11.5, 6.4$ Hz	8	4.79	Dd	$J = 11.7, 6.5$ Hz
			15a	4.87	d	$J = 1.4$ Hz	15a	4.86	D	$J = 1.4$ Hz
			15b	4.65	d	$J = 1.4$ Hz	15b	4.56	D	$J = 1.4$ Hz
			6a	2.77	dd	$J = 13.9, 3.9$ Hz	6a	2.68	Dd	$J = 14.0, 3.9$ Hz
			6b	2.43	m		3a	2.36–2.29	M	
			3a	2.37	m		6b	2.28	M	
			9a	2.28	dd	$J = 12.0, 6.4$ Hz	9a	2.25	Dd	$J = 12.4, 6.5$ Hz
Atractylenolide III	ATOIII	3b	2.06–1.98	m		3b	1.98–1.89	M		
		5	1.90	m		5	1.82–1.79	M		
		13	1.78	t	$J = 1.8$ Hz	13	1.78	T	$J = 1.6$ Hz	
		1a, 2	1.66–1.55	m		1a, 2	1.62–1.53	M		
		1b	1.40–1.33	m		1b	1.27	dt	$J = 12.9, 5.0$ Hz	
		9b	1.10	dd	$J = 12.0, 11.5$ Hz	9b	1.10	dd	$J = 12.4, 11.7$ Hz	
		14	0.92	s		14	0.86	s		
		15a	4.85	d	$J = 1.4$ Hz	15a	4.87	d	$J = 1.3$ Hz	
		15b	4.63	d	$J = 1.4$ Hz	15b	4.6	d	$J = 1.3$ Hz	
		6a	2.66	dd	$J = 13.1, 3.1$ Hz	6a	2.66	dd	$J = 13.2, 3.2$ Hz	
		6b	2.42	m	$J = 13.1, 1.5$ Hz	6b	2.45	dd	$J = 13.2, 1.5$ Hz	
		3a	2.38–2.33	m		3a	2.01–1.92	m	$J = 13.6$ Hz	
		9a	2.21	d	$J = 13.3$ Hz	9a	2.26	d		
		3b	2.06–1.96	m		3b	2.01–1.92	m		
5	1.91–1.87	m		5	1.86–1.82	m	$J = 1.3$ Hz			
13	1.78	d	$J = 1.4$ Hz	13	1.82	d				
2	1.69–1.60	m		1b, 2a, 2b	1.72–1.55	m				
1b	1.56–1.51	m		9b	1.56	d	$J = 13.3$ Hz			
9b	1.47	d	$J = 13.3$ Hz	1a	1.24	dt	$J = 11.8, 6.4$ Hz			
1a	1.29	dt	$J = 13.1, 5.1$ Hz	14	1.03	s				

Table 3 (continued)

Number	Compound	CD ₃ OD		CDCl ₃		
		Position	¹ HNMR	Position	¹ HNMR	
Hinesol	HIN	14	1.04	s		
		1	5.28	m	1	5.31
		2	2.00-1.90	m	2	1.96-1.90
		3,4,6,7,8,9	1.79-1.28	m	3,4,6,7,8,9	1.78-1.25
		14	1.68	m	14	1.68
		12	1.16	s	12	1.21
		13	1.16	s	13	1.21
		15	0.92	d	15	0.92
		15a	4.69	dd	15a	4.72
		15b	4.45	dd	15b	4.45
		3a	2.30	ddt	3a	2.31
		3b	2.01	m	3b	2.00
		1a	1.77	m	1a	1.77
		1b,2,5,6,7,8,9	1.68-1.14	m	1b,2,5,6,7,8,9	1.66-1.18
		12	1.16	s	12	1.21
Atractylodin	ATD	13	1.16	s	13	1.21
		14	0.71	s	14	0.70
		5	7.49	d	5	7.38
		1'	6.84	d	1'	6.79
		3	6.48	d	4	6.42
		4	6.47	dd	3	6.37
		8'	6.32	dq	8'	6.33
		2'	6.10	d	2'	6.11
		7'	5.64	ddd	7'	5.6
		9'	1.82	dd	9'	1.83
		H-1	5.38	d		
		H-3'	4.09	d		
		H-4'	4.01	dd		
		H-3,5,6,5',6'	3.84-3.67	m		
		sucrose	SUC	H-1'	3.63	d
H-1'	3.59			d		
H-2	3.41			dd		
H-4	3.37-3.34			m		

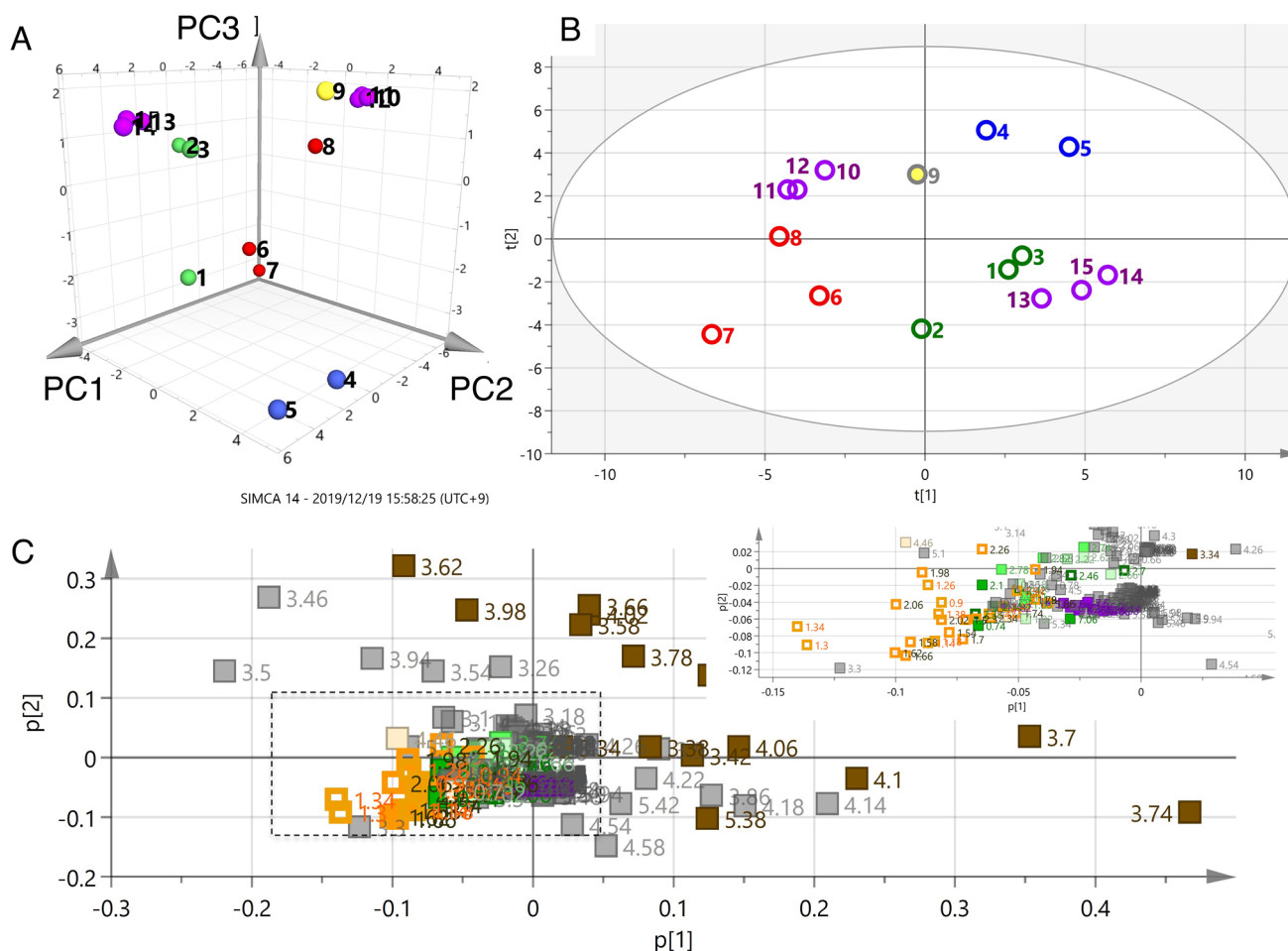


Fig. 4 Principal component analyses (PCA) using ^1H -NMR spectrum extracted with CD_3OD of different *Atractylodes* samples. **a** 3D score plot of PC1, PC2 and PC3 scores, **b** score plot of PC1 and PC2 scores, **c** loading plot for PC1 and PC2 components

Principal component analysis excluding sugar-based buckets in ^1H NMR spectra of CD_3OD extracts (CD_3OD sugar KO model)

Using deuterated chloroform made it possible to distinguish basic species without the influence of highly water-soluble primary metabolites, especially sugars (sucrose in this instance). Therefore, PCA was performed excluding the sugar region (3.30 to 5.40 ppm) of the ^1H NMR spectra of the CD_3OD extracts, because the PCA derived from the reduced NMR dataset was in general agreement with the PCA using CDCl_3 extracts. The results showed three significant principal components cumulatively accounting for 83% of the total variance. The 3D score plot showed major distinct clusters corresponding to the five different species (Fig. 6a, the contributions of PC1, PC2, and PC3 were 46%, 19%, and 18%, respectively). Especially, samples of *A. koreana* (purple) formed a cluster similar to that from the CDCl_3 extracts. The PC1/2 loading plot was used to verify group formation on the score plot (Fig. 6b). There

were many sesquiterpene-2-group buckets such as the chemical shifts of the dimethyl group resonances in hinesol and β -eudesmol, appearing at 1.16 ppm in the fourth quadrant, and whose clustering supported the grouping of *A. lancea* on the score plot. In contrast, the sesquiterpene-1 group (green) bucket was positioned in the second quadrant in the loading plot, which was in agreement with the distribution of the *A. japonica* group on the score plot. The *A. koreana* samples were distinguished as one group by the presence of the bucket corresponding to the resonance of atractylodin.

Thus, by modifying the calculation method, it was possible to fully utilize the performance of ^1H NMR metabolic profiling without changing the extraction solvent. In this method, highly water-soluble primary metabolites, such as saccharides, are excluded from the calculation to obtain characteristic metabolic fingerprints from the *Atractylodes* plants. On the other hand, this knocking out method would also exclude signals corresponding to glycosides at the same time. However, in this time, the KO model showed good agreement with the CDCl_3 model for the discrimination of

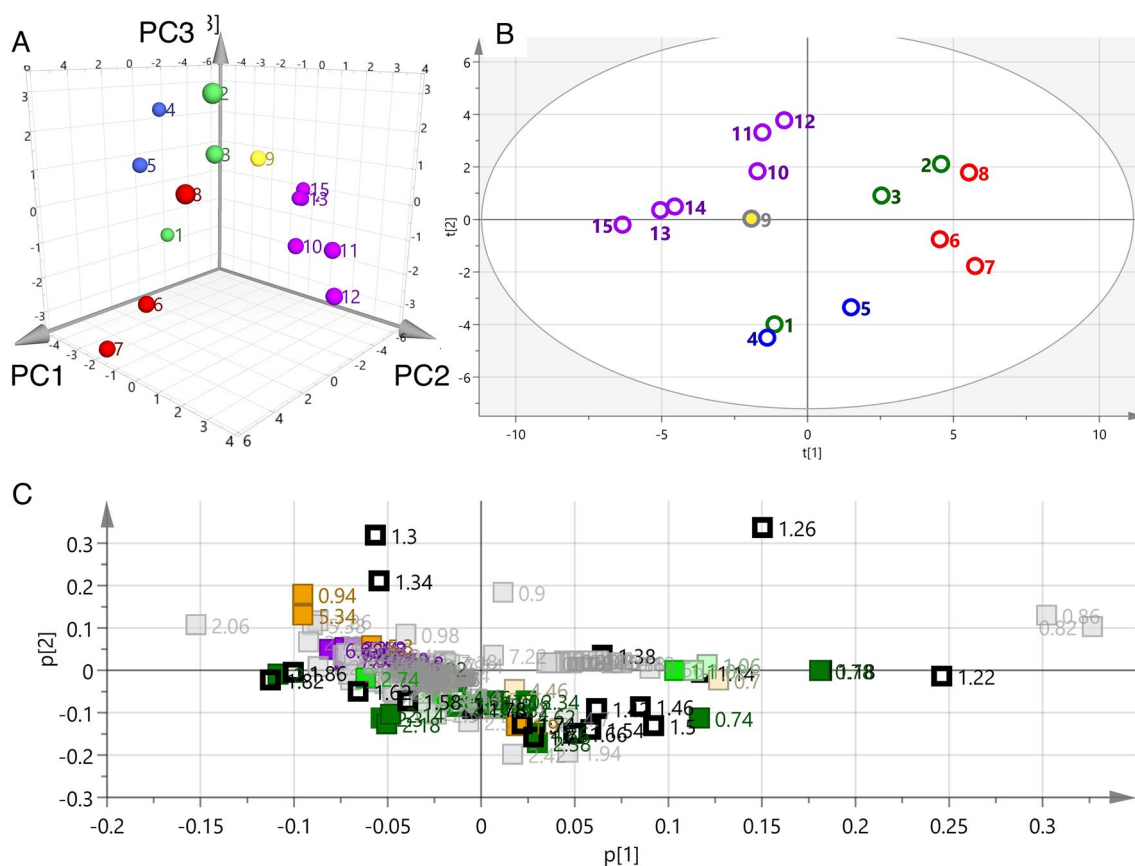


Fig. 5 Principal component analyses (PCA) using ¹H-NMR spectrum extracted with CDCl₃ of different *Atractylodes* samples. **a** 3D score plot of PC1, PC2 and PC3 scores **b** score plot of PC1 and PC2 scores **c** loading plot for PC1 and PC2 components

Atractylodes species. The reason for the discrimination is supposed to be that sucrose was the main component during the knockout region.

Comparison between three models using calculation parameters

The merit of NMR metabolic profiling is that the classification of plant species can be performed without specifying the index components. We successfully established a suitable method for ¹H NMR metabolic profiling of *Atractylodes* plants by selecting a bucket according to the desired parameters and performing statistical calculations with the same solvent.

The multiple correlation coefficient (R^2) and predictive ability parameter (Q^2) in each of the three PCA models, extraction of CD₃OD, CDCl₃ and CD₃OD knockout of sugar region (CD₃OD sugar KO), are shown in Table 4. Two-way orthogonal partial least squares (O2PLS-DA) was performed on the three models to add various classes of *Atractylodes* species as objective variables, checked the multiple correlation coefficient (R^2) and predictive ability parameter (Q^2)

of the resulting model, and performed a permutation test. R^2 and Q^2 values indicate the model's linearity and predictive ability, respectively. A discrimination model with an R^2 value of > 0.65 and a Q^2 value of > 0.5 is regarded as adequate for discrimination [31].

A permutation test is used to validate the incidence of overfitting in a predictive model [32, 33]. In this test, provisional discrimination models were constructed on the basis of various data matrices in which objective and explanatory variables were randomly combined many times, and R^2 and Q^2 were calculated for each provisional model. The original and permuted data matrices and the correlation coefficients of R^2 and Q^2 were plotted on the x and y axes, respectively. The y intercept of the regression line in the plot is used as the estimated index for overfitting: generally, $R^2 < 0.4$ and $Q^2 < 0.05$ [32, 33]. Table 4 shows the results of each estimated index for all the discrimination models.

The value of R^2Y did not change in any of the models, but the value of Q^2Y showed a good fit in the CD₃OD sugar KO model. This suggested that the CD₃OD sugar KO model could be used as the discrimination model. From the results of the permutation test, it was found that using the CD₃OD

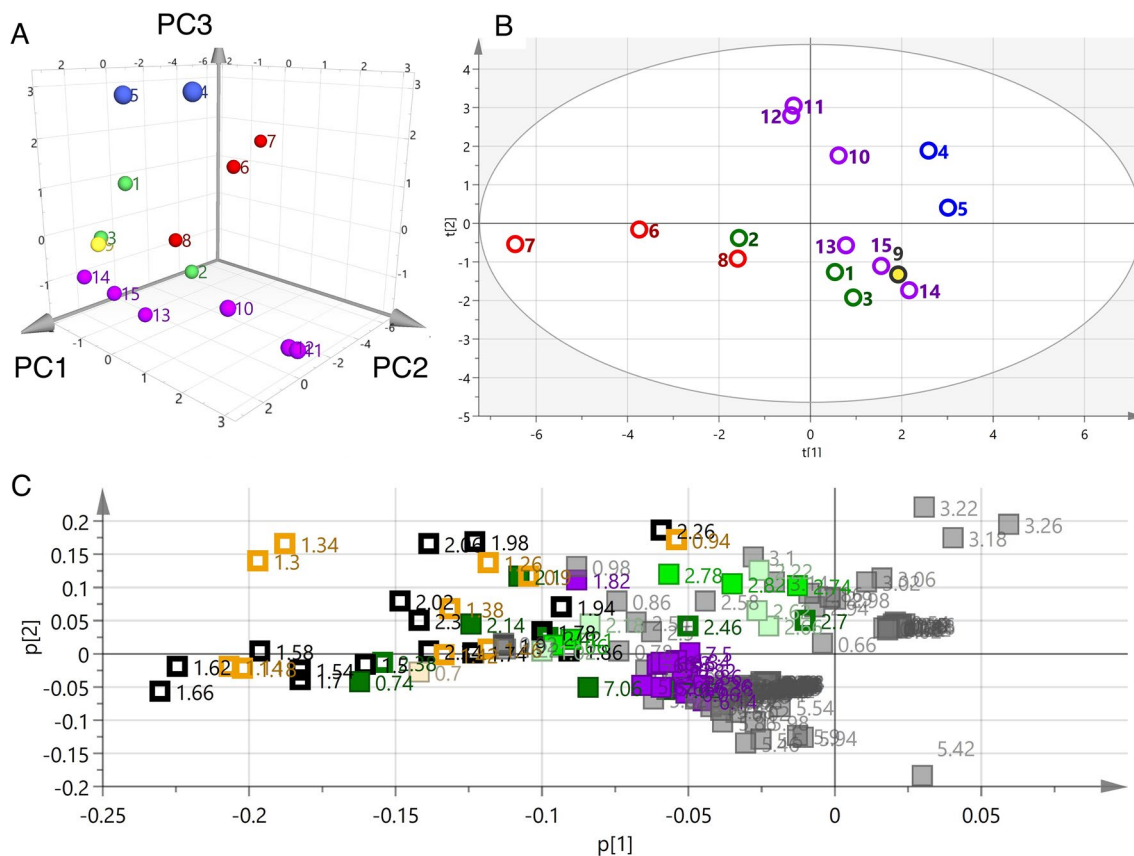


Fig. 6 Principal component analyses (PCA) using ¹H-NMR spectrum removed a sugar-based region (δ 3.30–5.40) extracted with CD₃OD of different *Atractylodes* samples. **a** 3D score plot of PC1, PC2 and

PC3 scores, **b** score plot of PC1 and PC2 scores, **c** loading plot for PC1 and PC2 components

Table 4 Estimation index results for all the discrimination models

Model	Permutation test											
			<i>A. japonica</i>		<i>A. macrocephala</i>		<i>A. lancea</i>		<i>A. chinensis</i>		<i>A. koreana</i>	
	<i>R</i> ²	<i>Q</i> ²	<i>R</i> ²	<i>Q</i> ²	<i>R</i> ²	<i>Q</i> ²	<i>R</i> ²	<i>Q</i> ²	<i>R</i> ²	<i>Q</i> ²	<i>R</i> ²	<i>Q</i> ²
CD ₃ OD	0.850	0.390	0.518	− 1.21	0.423	− 1.30	0.353	− 1.37	0.579	− 0.617	0.514	− 1.29
CDCl ₃	0.886	0.310	0.445	− 1.06	0.477	− 1.40	0.441	− 1.52	0.535	− 0.911	0.430	− 1.20
CD ₃ OD sugar KO	0.871	0.480	0.386	− 1.21	0.398	− 0.858	0.402	− 1.02	0.495	− 0.521	0.423	− 1.02

sugar KO model was suitable for predicting *A. japonica* and *A. macrocephala*.

Conclusion

In conclusion, we performed ¹H NMR metabolic profiling of DNA-authenticated, archived rhizomes of the *Atractylodes* genus for genetic and chemical quality evaluation. The nucleotide sequence of the ITS region of the nuclear rDNA was confirmed for five species, *A. japonica*, *A.*

macrocephala, *A. lancea*, *A. chinensis*, and *A. koreana*. An unbiased approach using multivariate statistical analysis of ¹H NMR spectra of CD₃OD extracts was adopted to reveal compositional differences in the primary and secondary metabolites among *Atractylodes* species, however, we failed to discriminate these plant species by this condition. Therefore, we prepared analytical samples by CDCl₃ extraction, in which clustering of each plant species was achieved to detect species-specific compounds on the score plot of the PCA. Removing the sugar peaks from the ¹H NMR spectra of the CD₃OD extracts with PCA gave the

same results as the PCA using CDCl_3 extracts. This biased chemometric model was able to successfully discriminate these plant species. The present study revealed that ^1H NMR-based metabolic profiling and genetic assessment are useful to identify members of the *Atractylodes* genus, which are categorized as different drugs in the Japanese Pharmacopoeia.

Experimental section

General experimental procedures

Polymerase chain reactions (PCRs) were performed with Ex Taq DNA polymerase (Takara, Kyoto, Japan). Sequencing reactions were carried out with Big Dye Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems, CA, USA), and the amplicons were electrophoresed on an ABI 3130 Genetic Analyzer (Applied Biosystems, CA, USA). All ground *Atractylodes* samples were extracted with an Accelerated Solvent Extraction system (ASE 350) from Dionex Corporation (Sunnyvale, CA, USA). The extracts were dried using a Thermo-Fisher Savant SC250EXP SpeedVac™ equipped with an RVT4104 refrigerated vapor trap. Freeze-drying was performed on a Labconco Freezone 4.5 (Kansas City, MO, USA). A precise Mettler Toledo XS105 dual range analytical balance was employed to prepare extracts for UHPLC and quantitative ^1H NMR (qHNMR) analyses. Samples for NMR analyses were prepared with a Pressure-Lock gas syringe (VICI Precision Sampling Inc., Baton Rouge, LA, USA) and calibrated glass pipets (cat. no: 2-000-200, Drummond Scientific, Broomall, PA, USA). Standard NMR tubes of 3 mm × 7 in. were purchased from Shigemi Co., Ltd. (no. PS005, Hachioji-city, Tokyo, Japan). ^1H NMR spectra were recorded on an Agilent Technologies 400-MR (400 MHz).

Reagents

Atractylodes standards, β -eudesmol, and atractylenolide III were obtained from Fujifilm Wako Pure Chemical Co. (Osaka, Japan). For NMR acquisition, $\text{CD}_3\text{OD}-d_4$ (99.8% D) and CDCl_3 (99.8% D) were purchased from Merck KGaA (Darmstadt, Germany). The signals in the ^1H NMR spectra of *Atractylodes* extracts were assigned to individual metabolites on the basis of thorough analyses of the 2D NMR spectra and spiking experiments. The PCR-grade tubes, tips, and most biological reagents used for DNA authentication were acquired from Qiagen (Valencia, CA, USA) and/or Thermo-Fisher Scientific and Beckman Coulter (Indianapolis, IN, USA). LO3 agarose for gel electrophoresis was purchased from Takara (Kyoto, Japan).

Plant material

Fifteen specimens of five *Atractylodes* species were analyzed: *A. japonica*, *A. macrocephala*, *A. lancea*, *A. chinensis*, and *A. koreana*. Details of the plant materials are shown in Table 1. Samples were botanically/macrospectically verified prior to inclusion. All the voucher specimens were deposited in Medicinal Plant Garden, School of Pharmacy, Kitasato University, Kanagawa, Japan.

DNA extraction, PCR amplification, and sequencing

Total DNA was extracted from 100–200 mg of leaf tissue using hexadecyltrimethylammonium bromide (CTAB) solution following the method of Doyle [26] with minor modifications. The primer pair ITS5 (5'-GGA AGT AAA AGT CGT AAC AAG G-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') [27] was used to amplify the ITS region of nrDNA. PCR amplification was performed in a 50- μL reaction volume containing 1 × reaction buffer for Ex Taq DNA polymerase, 0.2 mM of each dNTP, 0.2 μM of each primer, 0.5 units of Ex Taq DNA polymerase (Takara, Kyoto, Japan), and approximately 10 ng of template DNA. PCR was performed under the following cycling conditions: (95 °C, 3 min) × 1 cycle, (95 °C, 1 min; 52 °C, 1 min; 72 °C, 1 min 30 s) × 30 cycles, and (72 °C, 8 min) × 1 cycle. Since it is known that there are additive nucleotides in the ITS sequence of *A. japonica* [12], single strand conformation polymorphism (SSCP) analysis with the first PCR products was carried out to isolate individual alleles, following the method of Watano et al. [28]. Segregated bands were excised and purified before use as template DNA for the second PCR. The PCR products were purified with Amicon® Ultra Centrifugal Filter Units (Merck, Germany). Sequencing reactions using purified PCR products were carried out with BigDye Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems, CA, USA) under the following cycling conditions: (96 °C, 1 min) × 1 cycle, (96 °C, 10 s; 50 °C, 5 s; 60 °C, 3 min 30 s) × 40 cycles, and (60 °C, 5 min) × 1 cycle. A specific sequence primer, AJITSF3 (5'-CCG CGA ACA TGT AAT GAC AAC CGG GC-3') or AJITSR3 (5'-AAG CGT CGT CGC GAG GCG AC-3') was used to avoid non-specific annealing. The reaction products were analyzed using the ABI3130 Genetic Analyzer (Applied Biosystems, CA, USA).

Phylogenetic analysis

Sequence data were edited and aligned using BioEdit [29]. Phylogenetic analysis by three different methods (NJ, MP, and ML) was performed using PAUP 4.0b10 [30]. Bootstrap analysis of 1000 replicates was conducted for the NJ and MP trees.

Sample extraction and preparation

Each *Atractylodes* plant sample was extracted using 0.1 mg of material in 1.0 mL of CD₃OD-*d*₄ (99.8% D) or CDCl₃ (99.8% D) by ultrasonication at room temperature for 1 h. The mixture was centrifuged at 3,000 rpm (Kubota 3740, Japan) for 5 min. The supernatants were filtered through an Ekicrodisc® 30 mm syringe membrane filter (0.45-µm pore size) and transferred into 3-mm standard NMR tubes.

¹H NMR acquisition and ¹H NMR multivariate data analysis

The 1D ¹H NMR spectra were acquired at 298 K using a 45° excitation pulse experiment (Bruker pulprog: zg). The probe was frequency-tuned and impedance-matched before each acquisition. For each sample, 64 scans (ns) and 4 dummy scans (ds) were recorded with the following parameters: spectral width of 16 ppm, relaxation delay (D1) of 3.0 s, and receiver gain (RG) set to 256. The total duration of each ¹H NMR acquisition was 15 min. Off-line data processing was performed using MNova Lite (Mestrelab Research, S.L.) and ALICE2 software (JEOL). The ¹H NMR spectra were automatically Fourier-transformed using ALICE2 software (JEOL). The spectra were referenced to CH₃OH at 3.30 ppm and CHCl₃ at 7.26 ppm. Spectral intensities were reduced to integrated regions, referred to as buckets, of equal width (0.04 ppm) within the region of δ 10.0 to 0.00 ppm using ALICE2 for metabolome version 6 software (JEOL). In the case of CD₃OD, the region between δ 5.0 and 4.6, corresponding to residual water signals, was removed. The total integral value on the spectrum was set to 100 to provide bucket tables. In the resulting bucket tables, all rows were scaled to the total intensity, and Pareto scaling was applied to the columns preceding PCA and O2PLS-DA using SIMCA software. For another experiment, the regions between 4.00 and 3.00 ppm, corresponding to the sugar region, and the sucrose anomeric resonance at 5.169 ppm (doublet *J* = 3.75 Hz) were removed from the bucket table.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11418-020-01471-0>.

Funding This work was supported by JST COI Grant number JPMJCE1301.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Fu SM, Fang HJ, Liu GS, Xiao PG (1981) A study on the medicinal plants of the genus *Atractylodes*. *Acta Phytotaxon Sin* 19:195–202
2. Hiraoka N (1993) Medicinal and aromatic plants. In: Bajaj YPS (Ed), Springer-Verlag, Berlin, pp. 79–91
3. Tang W, Eisenbrand G (1992) Chinese drugs of plant origin. Springer-Verlag, Berlin, pp 209–211
4. Ministry of Health, Welfare and Labor of Japan (2016) The Japanese Pharmacopoeia, 17th edn, Ministry of Health, Welfare and Labor of Japan. Tokyo, Japan
5. Liu GS, Fu SM, Feng SX (1979) Manual of Chinese Medicines; Institute of Materia Medica of Chinese Academy of Medicinal Sciences, People's Health Publishing House: Beijing, pp. 152–164
6. Nishikawa Y, Watanabe Y, Seto T (1975) Studies on the evaluation of crude drugs. I. Comparative studies on the components of *Atractylodes Rhizomes*. *Shoyakugaku Zasshi* 29:139–146
7. Takeda O, Miki E, Terabayashi S, Okada M, Lu Y, He H-S, He S-A (1995) Variation of essential oil components of *Atractylodes chinensis* growing in China. *Yakugaku Zasshi* 115:543–552
8. Takeda O, Miki E, Terabayashi S, Okada M, Lu Y, He HS, He SA (1996) A comparative study on essential oil components of wild and cultivated *Atractylodes lancea* and *A. chinensis*. *Planta Med* 62:444–449
9. Terabayashi S, Miki E, Takeda O, Lu Y, He H-S (1997) Anatomical variations in rhizomes of *Atractylodes lancea* (Thunb.) DC. and *A. chinensis* Koidz. (Compositae), original plants of Chinese natural medicine, “Cangzhu.” *J Jpn Bot* 72:238–248
10. Hiraoka N, Matoba K, Ogawa N, Kagoshima K (1997) *Atractylon* is a major sesquiterpene in young *Atractylodes lancea* plants raised from shoot cultures. *Nat Med* 51:417–419
11. Takeda O, Miki E, Terabayashi S, Okada M, He S, Sashida Y (1996) Seasonal variation of essential oil components in *Atractylodes lancea* (Thumb.) DC. propagated by division of their rhizomes. *Chem Pharm Bull* 44:823–828. <https://doi.org/10.1248/cpb.44.823>
12. Kawanishi F, Takahashi T, Omukai T, Zhang B-G, Li Z-LXP-G (1994) Comparison of the outer morphologies, growth and the components in the rhizomes of *Atractylodes* plants cultivated in Kyoto and Beijing. *Nat Med* 48:1–10
13. Takeda O, Miki E, Morita M, Okada M, Lu Y, He H-S (1994) Variation of essential oil components of *Atractylodes lancea* growing in Mt. Maoshan area in Jiangsu Province, China. *Nat Med* 48:11–17
14. Kohda H, Goto K, Anetai M, Yamagishi T (1994) Studies on the botanical origin of *Atractylodes lancea*. *Nat Med* 48:58–62
15. Takeda O, Miki E, Terabayashi S, Okada M, Lu Y, He H-S, He S-A (1995) Variation of essential oil components of *Atractylodes lancea* growing in China. *Nat Med* 49:18–23
16. Shiba M, Kondo K, Miki E, Yamaji H, Morota T, Terabayashi S, Takeda S, Sasaki H, Miyamoto K, Aburada M (2006) Identification of medicinal *Atractylodes* based on ITS sequences of nrDNA. *Biol Pharm Bull* 29:315–320. <https://doi.org/10.1248/bpb.29.315>
17. Xia YG, Yang BY, Wang QH, Liang J, Wang D, Kuang HX (2013) Species classification and quality assessment of Cangzhu (*Atractylodes Rhizoma*) by high-performance liquid chromatography and

- chemometric methods. *J Anal Methods Chem* 2013:1–7. <https://doi.org/10.1155/2013/497532>
18. Cho HD, Kim U, Suh JH, Eom HY, Kim J, Lee SG, Choi YS, Han SB (2016) Classification of the medicinal plants of the genus *Atractylodes* using high-performance liquid chromatography with diode array and tandem mass spectrometry detection combined with multivariate statistical analysis. *J Sep Sci* 39:1286–1294. <https://doi.org/10.1002/jssc.201501279>
 19. Yoshitomi T, Wakana D, Uchiyama N, Tsujimoto T, Kawano N, Yokokura T, Yamamoto Y, Fuchino H, Hakamatsuka T, Komatsu K, Kawahara N, Maruyama T (2020) ¹H NMR-based metabolomic analysis coupled with reversed-phase solid-phase extraction for sample preparation of *Saposhnikovia* roots and related crude drugs. *J Nat Med* 74:65–75. <https://doi.org/10.1007/s11418-019-01343-2>
 20. Tsujimoto T, Yoshitomi T, Maruyama T, Yamamoto Y, Hakamatsuka T, Uchiyama N (2018) ¹³C-NMR-based metabolic fingerprinting of Citrus-type crude drugs. *J Pharm Biomed Anal* 161:305–312. <https://doi.org/10.1016/j.jpba.2018.08.044>
 21. Gilard V, Balayssac S, Malet-Martino M, Martino R (2010) Quality control of herbal medicines assessed by NMR. *Curr Pharm Anal* 6:234–245. <https://doi.org/10.2174/157341210793292365>
 22. van der Kooy F, Maltese F, Hae Choi Y, Kyong Kim H, Verpoorte R (2009) quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. *Planta Med* 75:763–775. <https://doi.org/10.1055/s-0029-1185450>
 23. Kim HK, Choi YH, Verpoorte R (2010) NMR-based metabolomic analysis of plants. *Nat Protoc* 5:536–549. <https://doi.org/10.1038/nprot.2009.237>
 24. Goodacre R, Shann B, Gilbert RJ, Timmins ÉM, McGovern AC, Alsberg BK, Kell DB, Logan NA (2000) Detection of the dipicolinic acid biomarker in *Bacillus* spores using curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Anal Chem* 72:119–127. <https://doi.org/10.1021/ac990661i>
 25. Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AWM, Verpoorte R (2004) Metabolomic differentiation of *Cannabis sativa* cultivars using ¹H NMR spectroscopy and principal component analysis. *J Nat Prod* 67:953–957. <https://doi.org/10.1021/np049919c>
 26. Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
 27. White TJ, Bruns T, Lee S (1990) Taylor JW. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols—a guide to methods and applications. Academic Press Inc., New York, pp 315–322
 28. Watano Y, Kanai A, Tani N (2004) Genetic structure of hybrid zones between *Pinus pumila* and *P. parviflora* var. *pentaphylla* (Pinaceae) revealed by molecular hybrid index analysis. *Am J Bot* 91:65–72. <https://doi.org/10.3732/ajb.91.1.65>
 29. Hall AT (1999) BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
 30. Swofford DL PAUP* (2002) Phylogenetic analysis using parsimony (*and Other Methods). Version 4; Sinauer Associates: Sunderland, Massachusetts
 31. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P (2015) PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol Biosyst* 11:13–19. <https://doi.org/10.1039/C4MB00414K>
 32. Eriksson L (2001) Multi- and megavariate data analysis principles and applications. Umetrics Academy, Umea, Sweden
 33. Marcia MCF (2017) Encyclopedia of physical organic chemistry. Online Library; Wiley, Hoboken, NJ, USA, pp 2041–2058

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Tatsuya Shirahata^{1,2} · Hiroshi Ishikawa¹ · Teruhisa Kudo¹ · Yumiko Takada¹ · Azusa Hoshino¹ · Yui Taga¹ · Yusaku Minakuchi¹ · Tomoko Hasegawa¹ · Rina Horiguchi¹ · Takehiro Hirayama¹ · Takahiro Konishi¹ · Hiroaki Takemoto¹ · Noriko Sato¹ · Masako Aragane³ · Tetsuro Oikawa² · Hiroshi Odaguchi² · Toshihiko Hanawa² · Eiichi Kodaira¹ · Tatsuo Fukuda¹ · Yoshinori Kobayashi^{1,2}

✉ Yoshinori Kobayashi
kobayashiy@pharm.kitasato-u.ac.jp

¹ School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

² Kitasato University Oriental Medicine Research Center, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

³ Tokyo Metropolitan Institute of Public Health, 24-1 Hyakunin-chou, 3-chome, Shinjuku-ku, Tokyo 169-0073, Japan