CrossMark

Identification of saponin composition and their geographical distribution in Chinese cultivated soybean (*Glycine max*)

Y. Takahashi · X. H. Li · L. J. Qiu · C. Tsukamoto · K. J. Wang

Abstract A total of 2959 Chinese cultivated soybeans were examined for saponin composition by thin-layer chromatography. Thirteen components were distinguished as five groups, group A (Aa, Ab, Ax, Ad, Ae and Af), DDMP (αg and βg), group B (Ba), group α (H-αg, I-αg and K-αg), and group β (KA-αg). The detected composition included four common phenotypes (Aa, Ab, AaBc, AbBc) and five new types (Aaβ_{KA}, Abα_H, Aaα_I, Abα_I, and Abα_K). The group α components and group β were first found in cultivated soybeans. The frequencies of the $Sg\text{-}1^a$ and $Sg\text{-}1^b$ alleles for Xyl and Glc at the C-22 terminal sugar were 39.3 and 60.7%, respectively, and the $Sg\text{-}4$ gene for Ara at the second sugar of C-3 was rare (0.07%). The frequency distribution in botanical characters indicated that the $Sg\text{-}1$ locus was related to soybean evolution and ecotypes; small seed, prostrate stem, black seed-coat, indefinite growth, broad-leaved shape, low-oil content, and spring-sowing types accumulated relatively higher frequencies of $Sg\text{-}1^a$ allele. The frequency of $Sg\text{-}1^a$ allele was regionally higher (>60%) along the Yellow River valley.

Integrating archaeological documents, the frequency distribution of $Sg\text{-}1^a$ allele and the occurrence of group α in soybeans with a relatively higher frequency in the northwest, a single area of origin for soybeans is proposed as the middle and lower reaches of the Wei River valley, i.e., the middle-upper reaches of the Yellow River valley in China. Group α components in soybeans are safe for humans because these saponins have been ingested for several 1000 years in China.

Introduction

Saponins are secondary metabolites that are abundant in the soybean [*Glycine max* (L.) Merrill], where they amount to 0.6–6.5% of dry weight (Shiraiwa et al. 1991b). Soybean saponins have a lot of health functional activities (Tsukamoto and Yoshiki 2006; Fenwick et al. 1991; Murata et al. 2006; Kuzuhara et al. 2000; Rowlands et al. 2002; Ellington et al. 2006; Kang et al. 2005; Lee et al. 2005; Ishii and Tanizawa 2006; Yang et al. 2015). Recent studies have reported that group-A saponins can prevent memory impairment (Hong et al. 2014). However, group-A saponins have a bitter and astringent taste owing to the acetylation in hydroxyl groups of the terminal sugar of the sugar chain attached at the C-22 position

Y. Takahashi · X. H. Li · L. J. Qiu · K. J. Wang (✉)
Institute of Crop Sciences, Chinese Academy of
Agricultural Sciences, Beijing 100081, China
e-mail: wangkj@caas.net.cn

C. Tsukamoto (✉)
Faculty of Agriculture, Iwate University,
Morioka 020-8550, Japan
e-mail: chigen@iwate-u.ac.jp

(Okubo et al. 1992), and there is a demand for group-A saponins to be genetically decreased or deleted from seeds in soybean breeding for soy milk (Kato et al. 2007).

Saponin composition has been investigated in wild soybean (*Glycine soja* Sieb. & Zucc.) and cultivated soybeans (Kudou et al. 1992, 1993; Shiraiwa et al. 1991a, c; Tsukamoto et al. 1993; Krishnamurthy et al. 2014b), and soybean saponin composition are classified into several groups, including group A, DDMP (and their decomposition), and group α saponins (formerly *Sg-6* saponin, here designated as the α saponins). Disintegrated DDMP saponins would create group B or E (Kudou et al. 1993; Tsukamoto et al. 2009). The group α saponins are a newly identified group in wild soybean accessions (Honda et al. 2009; Takahashi et al. 2016c) and are reported to be genetically controlled by the *Sg-6* locus (Krishnamurthy et al. 2014b), but group α saponins have not yet been found in cultivated soybeans. The group α has been found to have at least four components H, I, J, and K, with only a single sugar chain at the C-3 position of their aglycones (Takahashi et al. 2016b). Recently, a new group β saponin (KA-αg component) was found in a semi-wild soybean accession lacking group A saponins in China (Takahashi et al. 2016a, 2017).

Thus far, six loci *Sg-1* (Shiraiwa et al. 1990; Tsukamoto et al. 1993), *Sg-3*, *Sg-4* (Tsukamoto et al. 1993; Takada et al. 2012), *Sg-5* (Tsukamoto et al. 1998; Takada et al. 2013), *Sg-6* (Krishnamurthy et al. 2014b), and *Sg-7* (Takahashi et al. 2016a) have been found to be involved in saponin biosynthesis in soybeans, and thirteen alleles (*Sg-1^a^/Sg-1^b^/sg-1^0^*, *Sg-3/sg-3*, *Sg-4/sg-4*, *Sg-5/sg-5*, *Sg-6/sg-6*, *Sg-7/sg-7*) have been identified. Saponin composition phenotypes for soybean varieties or accessions can be described by the combination of saponin components, which depends on the combination of dominant and recessive alleles. Tsukamoto et al. (1993) analyzed the saponin composition of 800 Japanese soybean varieties and 329 wild soybeans and identified eight phenotypes (Aa, Ab, AcAf, AaAb, AaBc, AbBc, AaAbBc and A0Bc). Based on the components of group A and α saponins and the absence of Aa or Ab of group A, or loss of the whole group A, eleven types of saponin phenotypes (Aa, Ab, AaBc, AbBc, Aa + α, AaBc + α, AbBc + α, AuAeBc, A0-S, A0Bc-S, and A0Bc) were distinguished in Korean wild soybeans (Krishnamurthy et al. 2014b).

China is thought to be the area of origin of cultivated soybeans, and investigation of saponin polymorphisms in Chinese cultivated soybeans has not been carried out yet. The group α saponins have been newly identified in wild soybeans. Little is known about the functional activities of α saponins at present so the public are suspicious of their safety as α saponins do not exist in cultivated soybeans. If cultivated soybeans contain the group α saponins, there is no need to worry at all about this group of saponins. Our objective in this work was to clarify the general situation of the polymorphisms of saponins in Chinese soybeans and their geographical distribution, and to examine whether α saponins exist in cultivated soybeans in the area of origin, China. We also would like to probe into the area of the origin of soybeans considering taking another otherwise indicator of chemical composition for exploration of this issue such as saponins.

Materials and methods

Plant materials and chemicals

A total of 2959 cultivated soybean accessions, stored in the national crop gene bank at the Chinese Academy of Agricultural Sciences (CAAS), were used in this study. These accessions all were landraces, randomly taken from all the administrative regions of China (Table 1).

To assess the distribution among geographical regions for saponin composition and phenotypes, the accessions were zoned into eight geographical regions (Northeast, Northwest, North, Centre, Eastern North, Southeast, Southwest, South) according their provenances (Table 1; Fig. 3). For insight into the frequency differences of alleles relevant to saponin phenotypes in different character types in morphology, agronomy and ecology, we partitioned the accessions into thirteen character groups including seven qualitative characters [seed-coat colours (yellow, green, brown, bicolour and black), cotyledon colours (yellow and green), stem erectness (erect, weak-erect, semi-prostrate and prostrate), pubescence colours (brown and grey), flower colours (purple and white), leaf shapes (rotundifolius, ovoid, elliptic, and lanceolate), and growth habits (definite, semi-definite and indefinite)], four quantitative characters [100-seed

Table 1 Saponin phenotypes and their geographical distribution in Chinese cultivated soybeans

Region	No. of accessions	Phenotype*				New saponin type**					Frequency (%)***		
		Aa ($Sg\text{-}1^{a}$ $sg\text{-}4$)	Ab ($Sg\text{-}1^{b}$ $sg\text{-}4$)	AaBc ($Sg\text{-}1^{a}$ $Sg\text{-}4$)	AbBc ($Sg\text{-}1^{b}$ $Sg\text{-}4$)	$Aa\beta_{KA}$	$Aa\alpha_{I}$	$Ab\alpha_{I}$	$Ab\alpha_{H}$	$Ab\alpha_{K}$	$Sg\text{-}1^{a}$	$Sg\text{-}1^{b}$	$Sg\text{-}4$
Northeast													
Heilongjiang	107	41	66								38.3	61.7	
Jilin	95	31	64								32.6	67.4	
Liaoning	92	21	71								22.8	77.2	
Subtotal	294	93	201								31.6	68.4	
Northwest													
Xinjiang	7	7									100.0		
Ningxia	76	65	11								85.7	14.3	
Gansu	99	32	64				1 (1.01)		1 (1.01)	1 (1.01)	33.3	66.7	
Shannxi	151	64	83					1 (0.66)	3 (1.99)		42.4	57.6	
Subtotal	333	168	158				1 (0.30)	1 (0.30)	4 (1.20)	1 (0.30)	50.8	49.2	
North													
Inner Mongolia	112	71	41								63.4	36.6	
Shanxi	189	117	72								61.9	38.1	
Beijing	29	21	8								72.4	27.6	
Hebei	106	55	51								51.9	48.1	
Subtotal	436	264	172								60.6	39.4	
Centre													
Henan	102	73	29								71.6	28.4	
Hubei	165	32	133								19.4	80.6	
Hunan	101	24	77								23.8	76.2	
Subtotal	368	129	239								35.1	64.9	
Eastern North													
Shandong	103	65	38								63.1	36.9	
Anhui	98	31	67								31.6	68.4	
Jiangsu	198	50	146					2 (1.01)			25.3	74.7	
Shanghai	53	9	44								17.0	83.0	
Subtotal	452	155	295					2 (0.44)			34.3	65.7	
Southeast													
Fujian	100	41	59			1 (0.99)					41.0	59.0	
Zhejiang	101	30	70								30.7	69.3	
Jiangxi	103	23	80								22.3	77.7	
Subtotal	304	94	209			1 (0.33)					30.3	69.7	

Table 1 continued

Region	No. of accessions	Phenotype*				New saponin type**					Frequency (%)***		
		Aa (Sg-1^a sg-4)	Ab (Sg-1^b sg-4)	AaBc (Sg-1^a Sg-4)	AbBc (Sg-1^b Sg-4)	Aaβ$_{KA}$	Aaα$_I$	Abα$_I$	Abα$_H$	Abα$_K$	Sg-1^a	Sg-1^b	Sg-4
Southwest													
Chongqing	102	42	59					1 (0.98)			41.2	58.8	
Sichuan	170	55	115								32.4	67.6	
Guizhou	190	68	121						1 (0.53)		35.8	64.2	
Tibet	12	3	8						1 (8.30)		25.0	75.0	
Yunnan	100	29	69	1 (1.00)	1 (1.00)						30.0	70.0	2.0
Subtotal	574	197	372	1 (0.17)	1 (1.00)			1 (0.17)	2 (0.33)		34.5	65.5	0.3
South													
Guangxi	21	11	10								52.4	47.6	
Guangdong	177	50	127								28.2	71.8	
Subtotal	198	61	137								30.8	69.2	
Total	2959	1161	1783	1 (0.03)	1 (0.03)	1 (0.03)	1 (0.03)	4 (0.14)	6 (0.20)	1 (0.03)	39.3	60.7	0.07

* Parentheses indicate principal alleles that control the phenotypes

** Parentheses indicate percent frequencies

*** Percent frequency of alleles

weight (g), plant height (cm), protein and oil contents (%)] and an ecological character (three sowing-ecotypes: spring-sowing, summer-sowing, and autumn-sowing) (Table 2).

The quantitative characters were divided into different classes. The 100-seed weight was grouped into four classes ($\leq$10.0, 10.01–20.0, 20.01–30.0, >30.0 g), plant height into five classes ($\leq$50.0, 50.01–100.0, 100.01–150.0, 150.01–200.0, >200.0 cm), protein content into four classes ($\leq$40.0, 40.01–45.0, 45.01–50.0, >50.0%), and oil content into five classes ($\leq$16.0, 16.01–18.0, 18.01–20.0, 20.01–22.0, >22.0%). All morphological and quantitative characters data were adopted from the soybean collection records preserved in the national crop gene bank, CAAS.

Japanese variety "Suzuyutaka" (saponin type: Ab) was used as the standard of saponin composition in thin layer chromatography (TLC). All chemicals (chloroform, methanol, acetonitrile, and sulfuric acid) were purchased from Beijing Chemical Works (Beijing, China).

Extraction of saponins

A mature dry seed from each accession was divided into hypocotyls, cotyledons and seed coats with a utility knife. The saponin components were extracted from the seed hypocotyl with 80% (v/v) methanol (MeOH) at room temperature for 12 h.

TLC and LC–MS analysis

The identification of saponin composition was carried out with TLC analysis according to our previous method (Takahashi et al. 2016a). Silica gel (SiO_2) plates were applied (catalog no. 105626, Merck Millipore, Germany). For further identification of variant saponin composition, LC-PDA/MS analysis was performed using a hybrid ion-trap mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific Incorporated, Waltham, MA, USA) under an UFLC system (Prominence UFLC system, Shimadzu Corporation, Kyoto, Japan) with a photodiode array (PDA) detector and a C30 reverse-phase column (Develosil C30 UG-3, 2.0 × 150 mm, 3 μm; Nomura Chemical, Okayama, Japan). The saponin extracts were diluted ten times with 80% (v/v) methanol. The solvent for elution were selected for acetonitrile including 0.1% formic acid. The initial solvent concentration was

adjusted to 20% (v/v) and linearly increased to 80% (v/v) in 60 min. An automatic full-scan mode over a mass of charge ratio (*m/z*) from 300 to 2000 was used to acquire MS data. The data of UV chromatograms and MS spectra were analyzed with Xcalibur software version 3.1 (Thermo Fisher Scientific Incorporated).

Results

Saponin composition and phenotypes in Chinese soybeans

Saponin analysis of 2959 Chinese soybean accessions distinguished 13 components in five separate saponin groups (group A: Aa, Ab, Ae, Ax, Ad, and Af; group DDMP: αg and βg; group B: Ba; group α: H-αg, I-αg and K-αg; group β: KA-αg) (Fig. 1).

Groups α and β were first found in cultivated soybeans, and the latter appeared in only one accession (sample ZDD13997 from the eastern south Zhejiang, an autumn-sowing type) and group α in 12 accessions, with separate components H-αg, I-αg and K-αg (Fig. 1). LC–MS analysis confirmed the existence of the new group-β and group-α saponins, which emerged at 22.5 min for KA-αg, 24.5 min for H-αg, 25.5 min for I-αg, and 35.9 min for K-αg (Fig. 2). These accessions with group α saponins comprised seven accessions from the northwest (ZDD10529, ZDD10728, ZDD10929 and ZDD11158 for Abα$_H$, ZDD10554 for Abα$_I$, ZDD11144 for Abα$_K$, and ZDD11182 for Aaα$_I$), three accessions from the southwest (ZDD17640 and ZDD15516 for Abα$_H$, and ZDD13221 for Abα$_I$), and two accessions from the eastern north areas (ZDD04515 and ZDD04584 for Abα$_I$).

Saponin composition includes many components, thus far more than 100 components have been detected according to published documents (Shiraiwa et al. 1991a, c; Kudou et al. 1992, 1993; Tsukamoto et al. 1993; Krishnamurthy et al. 2014b; Takahashi et al. 2016b, c). Usually, the combination of saponin components is divided into phenotypic types according to the occurrence of saponin composition in soybean variety individuals or accessions (Tsukamoto et al. 1993). In dividing saponin composition types for Korean wild soybeans, Krishnamurthy et al. (2013) only denoted presence or absence for the group α saponins (for example, Aa + α, AaBc + α, AbBc + α). Here we adopted idiographic components

Table 2 Frequency distribution of saponin phenotypes for botanically qualitative and quantitative character types in Chinese cultivated soybeans

Region	No. of accessions	Phenotype*				New saponin type					Frequency (%)**		
		Aa (Sg-1^a sg-4)	Ab (Sg-1^b sg-4)	AaBc (Sg-1^a Sg-4)	AbBc (Sg-1^b Sg-4)	Aaβ_{KA}	Aaα_I	Abα_I	Abα_H	Abα_K	Sg-1^a	Sg-1^b	Sg-4
Seed-coat color													
Green	376	124	251								33.0	67.0	
Yellow	1973	731	1231	1	1		1	4	3	1	37.2	62.8	0.1
Brown	187	72	115								38.5	61.5	
Bicolour	72	27	44						1		37.5	62.5	
Black	350	206	142		1				1	1	58.9	41.1	0.3
Cotyledon color													
Yellow	2878	1127	1736	1	1	1	1	4	6	1	39.3	60.7	0.1
Green	79	33	46								41.8	58.2	
Stem erectness													
Erect	2000	632	1356	1	1	1	2	2	6	1	31.7	68.3	0.1
Weak-erect	550	255	293				2	2			46.4	53.6	
Semi-prostrate	210	128	82								61.0	39.0	
Prostrate	197	144	52	1							73.6	26.4	0.5
Pubescence													
Brown	1567	601	957	1	1		2	2	4	1	38.5	61.5	0.1
Grey	1387	556	825		1	1	2	2	2	1	40.2	59.8	
No-pubescence	5	2	3								40.0	60.0	
Flower color													
Purple	1661	651	1000	1	1	1	3	3	3	1	39.3	60.7	0.1
White	1297	509	783				1	1	3		39.3	60.7	
Leaf shape													
Rotundifolious	128	58	70								45.3	54.7	
Ovoid	1132	495	626	1	1		2	2	4	1	44.0	56.0	0.2
Elliptic	1471	494	973					2	2		33.6	66.4	
Lanceolate	80	28	52								35.0	65.0	
Growth habit													
Definite	1621	503	1110	1	1	1	3	3	1	1	31.2	68.8	0.1
Semi-indefinite	499	173	322						4		34.7	65.3	
Indefinite	834	481	350	1			1		1	1	57.8	42.2	0.1
Ecotype													
Spring-sowing	1576	673	902						1		42.7	57.3	
Summer-sowing	1249	444	793	1	1		1	4	4	1	35.7	64.3	0.2

Table 2 continued

Region	No. of accessions	Phenotype*				New saponin type						Frequency (%)**		
		Aa ($Sg\text{-}1^a\ sg\text{-}4$)	Ab ($Sg\text{-}1^b\ sg\text{-}4$)	AaBc ($Sg\text{-}1^a\ Sg\text{-}4$)	AbBc ($Sg\text{-}1^b\ Sg\text{-}4$)	$Aa\beta_{KA}$	$Aa\alpha_I$	$Aa\alpha_I$	$Ab\alpha_I$	$Ab\alpha_H$	$Ab\alpha_K$	$Sg\text{-}1^a$	$Sg\text{-}1^b$	$Sg\text{-}4$
Autumn-sowing	133	43	88			1			1			33.1	66.9	
100-seed wt. (g)														
≤10.0	390	229	155	1	1				1	3		59.0	41.0	0.5
10.01–20.0	2093	798	1290				1		1	3		38.2	61.8	
20.01–30.0	429	113	310			1	1		1	2	1	26.8	73.2	
>30.0	46	17	28							1		37.0	63.0	
Plant height (cm)														
≤50.0	713	200	511	1							1	28.2	71.8	0.1
50.01–100.0	1617	584	1024				2		2	5		36.2	63.8	0.1
100.01–150.0	415	234	177		1				2	1		56.4	43.6	0.2
150.01–200.0	68	41	27									60.3	39.3	
>200.0	8	6	2									75.0	25.0	
Protein (%)														
≤40.0	111	33	78									29.7	70.3	
40.01–45.0	1432	618	809				1			4		43.2	56.8	
45.01–50.0	1198	417	772		1	1		3		3	1	34.9	65.1	0.1
>50.0	28	4	24									14.3	85.7	
Oil (%)														
≤16.0	355	157	196	1	1							44.5	55.5	0.3
16.01–18.0	1325	541	772			1	1	3	1	6		41.0	59.0	
18.01–20.0	948	331	616						1			34.9	65.1	
20.01–22.0	127	41	86									32.3	67.7	
>22.0	15	2	13									13.3	86.7	

* Parentheses indicate principal alleles that control the phenotypes

** Percent frequency of alleles
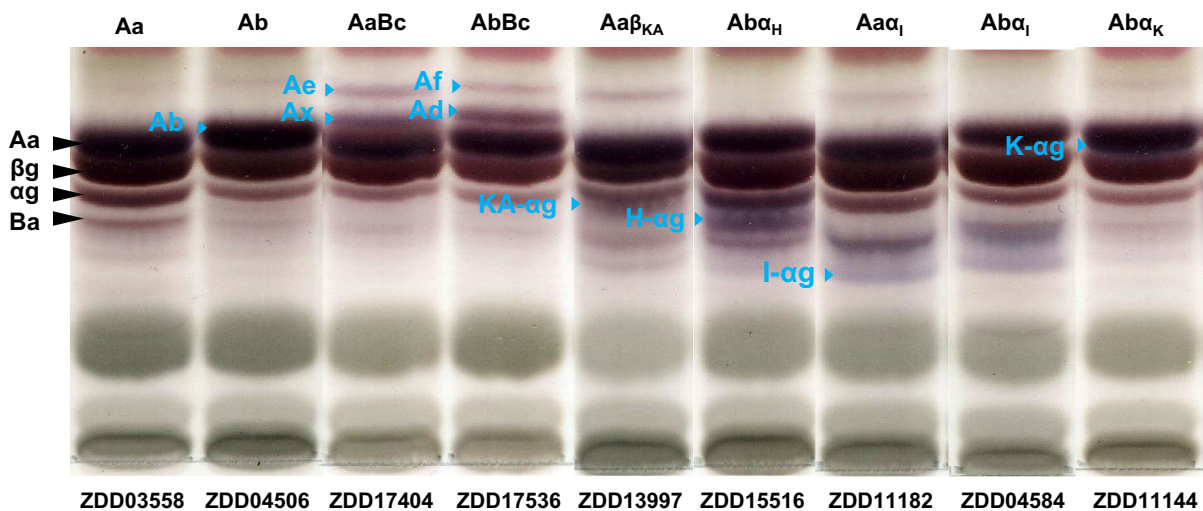
Fig. 1 Polymorphism of saponin composition detected in seed hypocotyls of Chinese soybeans by TLC analysis. There were five groups of saponin components: group-A (Aa, Ab, Ae, Ax, Af, and Ad), DDMP (αg and βg), group-α (H-αg, I-αg, and K-αg) and β (KA-αg), group B (Ba) and nine phenotypes of saponin composition: Aa, Ab, AaBc, AbBc, Aaβ_{KA}, Abα_H, Aaα_I, Abα_I, and Abα_K. The group-α saponins were characterized by *blue* colour

of group α in addition to Krishnamurthy et al.'s phenotype grouping and identified four common types Aa, Ab, AaBc, AbBc and five new types Aaβ_{KA}, Abα_H, Aaα_I, Abα_I, and Abα_K (Fig. 1). The Ab + α type did not appear in Korean wild soybeans but occurred in Chinese soybeans.

Two common saponin phenotypes Aa (involved with *Sg-1^a*, *Sg-3*, *sg-4*, *Sg-5*, and *sg-6* genes) and Ab (*Sg-1^b*, *Sg-3*, *sg-4*, *Sg-5*, and *sg-6*) were observed at a frequency of 39.2 and 60.8%, respectively, in the total accession samples. The AaBc type (*Sg-1^a*, *Sg-3*, *Sg-4*, *Sg-5*, *sg-6*) and AbBc type (*Sg-1^b*, *Sg-3*, *Sg-4*, *Sg-5*, *sg-6*), which rested with the *Sg-4* dominant allele, each appeared in only one accession (ZDD17404, AaBc and ZDD17536, AbBc).

Frequencies and geographical distribution of saponin phenotypes in Chinese soybeans

The analysis showed that two types Aa and Ab were the main types of nine saponin phenotypes in Chinese cultivated soybeans, with frequencies of 39.2 and 60.3% for Aa and Ab, respectively (Table 1). Aa and Ab type are controlled by two codominant alleles of *Sg-1^a* and *Sg-1^b*, respectively, at the *Sg-1* locus for Xyl and Glc at C-22 terminal sugar component (Sayama et al. 2012), and our current results demonstrated that there were regional differences in the distribution of allelic

frequencies between Aa (*Sg-1^a* allele) and Ab (*Sg-1^b* allele) in Chinese soybeans. One region (Northwest) had almost the same frequencies of both *Sg-1^a* and *Sg-1^b* alleles (50.8 and 49.2%, respectively) and one region (North) showed a higher frequency (60.6%) for the *Sg-1^a* allele than that for *Sg-1^b* (39.4%); while the other six areas all showed higher and similar *Sg-1^b* frequencies (64.9–69.7%) than *Sg-1^a* allele (30.3–35.1%). By and large, the northern regions (the Yellow River valley) had a higher *Sg-1^a* frequency, and the Northeast and the Yangtze River valley and its southern regions had a higher *Sg-1^b* frequency in Chinese soybeans (Table 1; Fig. 3).

There were only two accessions from southwestern Yunnan Province that were found to carry dominant *Sg-4* allele, with a very low 0.33 frequency in all the tested southwestern soybean accessions (Table 1). The sole group β saponin component KA-αg was found in one accession from Zhejiang Province in the Southeast region. Only twelve accessions have group α saponins, existing at a very low frequency of 0.41%, in Chinese cultivated soybeans, which were distributed in three regions, the Northwest, Eastern North, and Southwest. Particularly the Northwest, Gansu and Shaanxi provinces produced seven accessions and contributed a relatively higher proportion (58.33%), with more than half the number of all the group α saponin accessions (Table 1).

Fig. 2 HPLC chromatograms of saponin extracts from soybean hypocotyls under UV 205 nm by LC–MS analysis. The results distinctly revealed the occurrence of group-α saponin components H-αg, I-αg, and K-αg and KA-αg in soybeans (**c–f**). **a**, **b** were the standard saponin components of Aa and Ab types
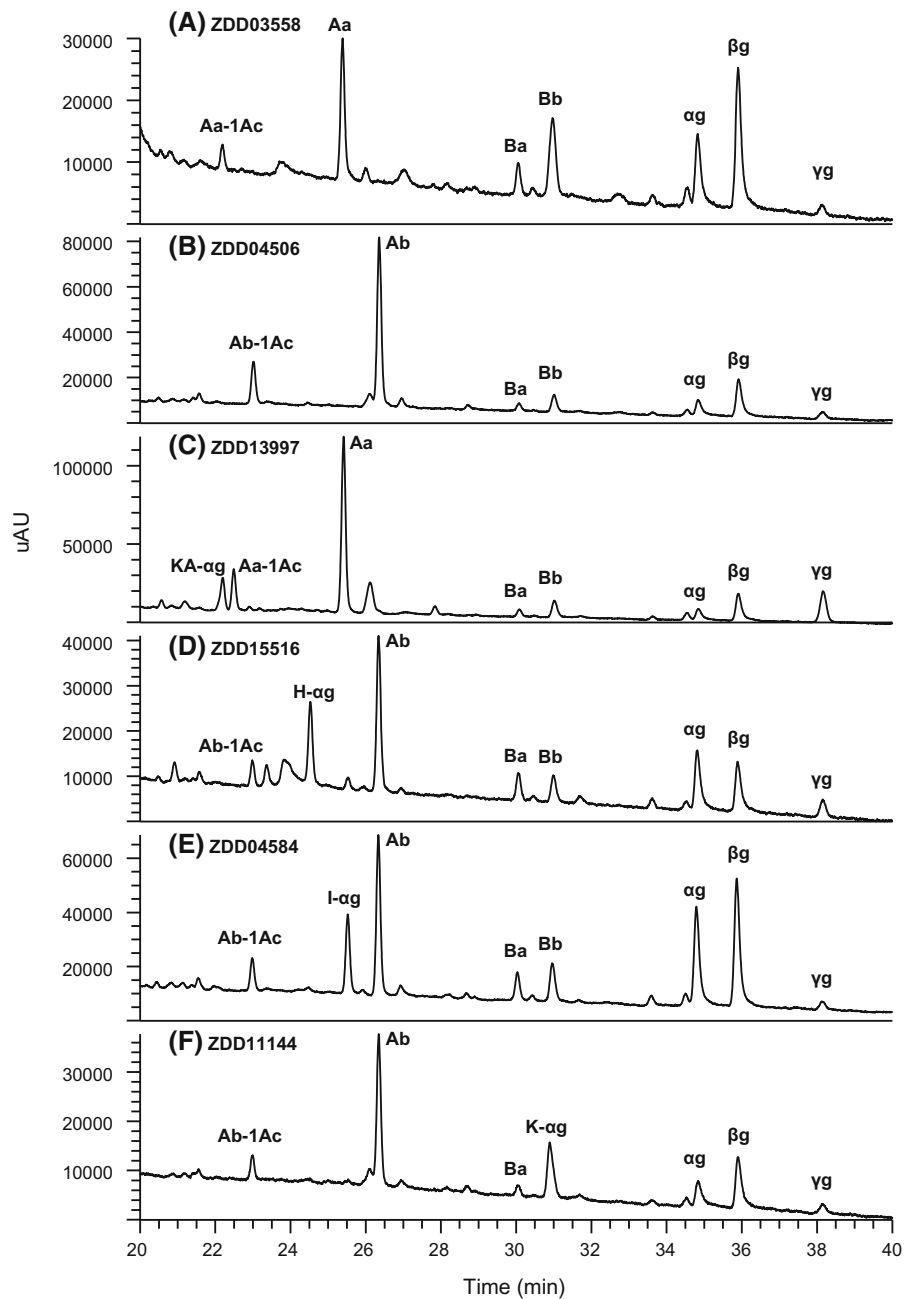
Allelic frequency at the *Sg-1* locus among different botanically qualitative character types in Chinese soybeans

Seed-coat colour, cotyledon colour, stem erectness, pubescence colour, flower colour, leaf shape and growth habits are main botanical qualitative traits for the morphological description of soybean varieties. Some characters could be related to evolutionary degrees or ecological types of soybeans. Aa and Ab phenotypes were validated to two main saponin types in Chinese soybeans (Table 1). Here, we analyzed the allelic distribution of two main Aa and Ab phenotypes with these characters.
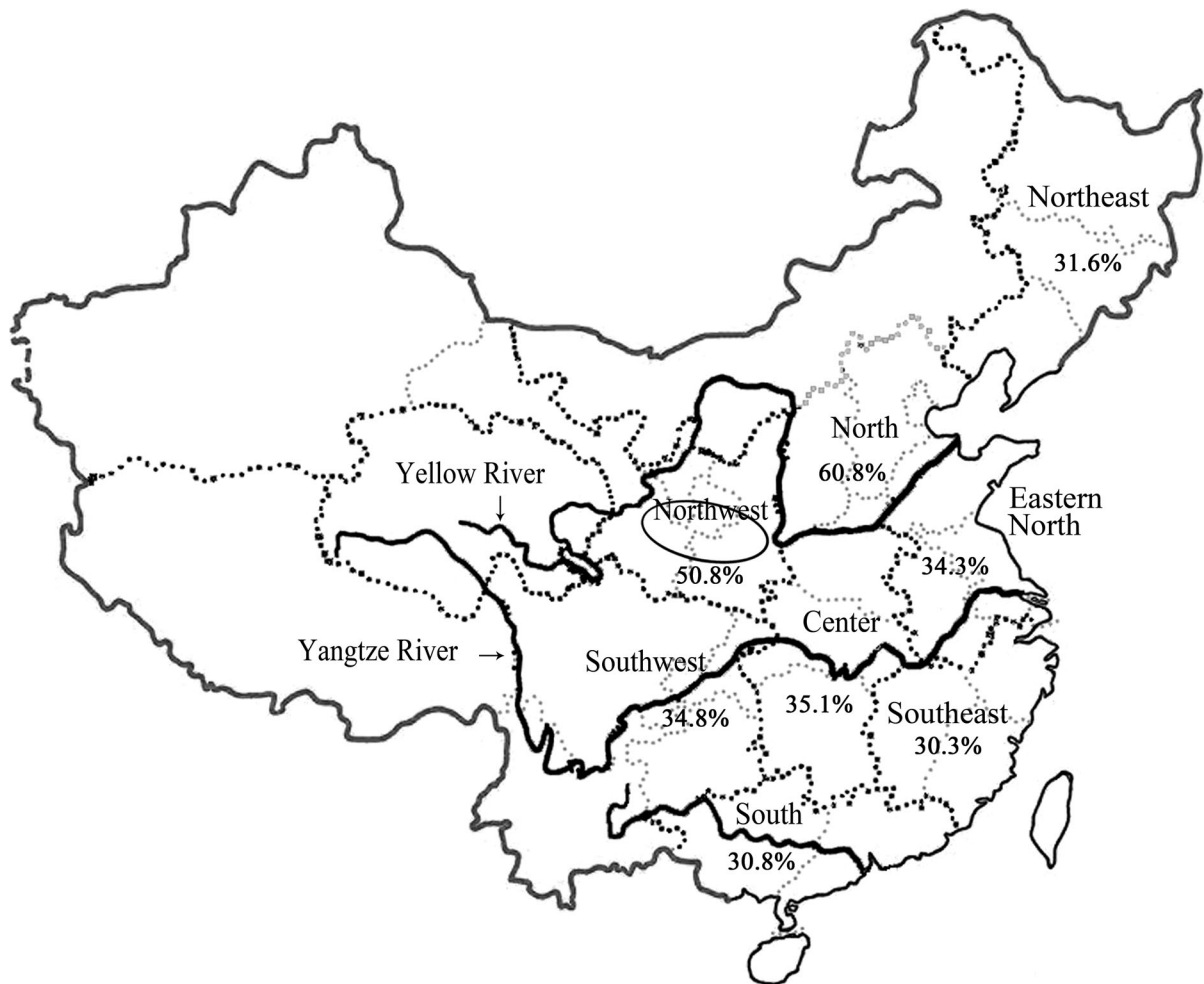
Fig. 3 Percentage frequency distribution of the $Sg-1^a$ allele in eight geographical regions, northeast, Northwest, North, Centre, Eastern North, Southeast, Southwest and South. The northwest and north China along Yellow River had the higher frequencies (>50%) of $Sg-1^a$. The *ellipse* denotes the area of the middle and lower reaches of the Wei River valley

Cotyledon, pubescence and flower colours

The results showed that there were no obvious differences in the frequency of $Sg-1^a$ or $Sg-1^b$ between yellow (39.3% $Sg-1^a$) and green (41.8% $Sg-1^a$) types of cotyledon colours, among grey (40.2% $Sg-1^a$), brown (38.5% $Sg-1^a$) and no-pubescence (40.0% $Sg-1^a$) of pubescence colours, and between purple (39.3% $Sg-1^a$) and white (39.3% $Sg-1^a$) of flower colours. However, $Sg-1^b$ frequency (58.2–60.7%) was higher than $Sg-1^a$ (38.5–41.8%) in each type of the three characters (Table 2).

Seed-coat colour

Seed-coat colours presented differences in allelic frequencies among seed-coat types, i.e., black seed type showed the highest $Sg-1^a$ frequency (58.9%), brown (38.5%), bicolour (37.5%) and yellow (37.2%) colour types had similar lower frequencies, and the green had the lowest $Sg-1^a$ frequency (33.0%). In addition, only the black seed type had a higher $Sg-1^a$ than $Sg-1^b$ frequency, whereas all other types were contrariwise (Table 2).

Leaf shape

The soybean landrace accessions had four leaf shapes, and all showed higher frequencies (54.7–66.4%) for the $Sg-1^b$ allele than for $Sg-1^a$ (33.6–45.3%) (Table 2). There were two obvious groups in $Sg-1^a$ frequency: rotundifolius and ovoid leaf types accumulated higher frequencies of $Sg-1^a$, with 45.3 and 44.0% frequencies, respectively. Elliptic and lanceolate leaf types had lower frequencies for the $Sg-1^a$ allele, giving 33.6 and 35.0% frequencies, respectively.

Stem erectness and growth habit

There were four types of stem erectness in the accessions. $Sg-1^a$ or $Sg-1^b$ allelic frequency in the stem type groups depended upon the degrees of stem erectness (Table 2). $Sg-1^a$ frequency varied from high to low with the degree of erectness: prostrate (73.6%) > semi-prostrate (61.0%) > weak-erect (46.4%) > erect (31.7%).

Stem growth habit of soybeans included three types: definite, semi-definite and indefinite growth. Similarly to stem erectness, growth habit also showed that the frequency of $Sg-1^a$ or $Sg-1^b$ varied with growth habit types (Table 2); the order of frequency fluctuation for $Sg-1^a$ accumulation was: indefinite type (57.8%) > semi-indefinite (34.7%) > definite (31.2%). The indefinite type also accumulated a higher frequency of the $Sg-1^a$ allele than of $Sg-1^b$.

Ecotype

The analyzed accessions belonged to three sowing-ecotypes: spring, summer, and autumn. The frequencies of $Sg-1^a$ among the three types reduced gradually in the order spring (42.7%), summer (35.7%), and autumn (33.1%) (Table 2). $Sg-1^b$ frequency was higher than $Sg-1^a$ in each sowing-type.

Allelic frequency at the Sg-1 locus among different types of agronomically quantitative characters in Chinese soybeans

We investigated the degree of accumulation of the $Sg-1^a$ or $Sg-1^b$ allele in four important agronomically quantitative characters: 100-seed weight (g), plant height (cm), protein and oil contents (%). Some of these characters clearly exhibited differences in the frequency of $Sg-1$ accumulation (Table 2).

100-seed weight

In four different groups of 100-seed weight, the group ≤10.0 g had an obviously higher frequency (59.0%) of $Sg-1^a$ than the other three groups (38.2% for 10.01–20.0 g, 26.8% for 20.01–30.0 g, and 37.0% for >30.0 g) (Table 2). Furthermore, $Sg-1^a$ frequency was higher than $Sg-1^b$ in the group with the smallest seeds.

Plant height

Plant height was divided into five classes from ≤50.0 to >200.0 cm. The results displayed that $Sg-1^a$ or $Sg-1^b$ frequency varied with plant height (Table 2). The degression of $Sg-1^a$ frequency occurred from the tallest plant type (>200.0 cm) to the shortest one (≤50.0 cm), i.e., frequency gradually was reduced in turn from 75.0, 60.3, 56.4, 36.2 to 28.2% for the tallest plant type (>200.0 cm), high plant height (150.01–200.0 cm), medium height (100.01–150.0 cm), low plant height (50.01–100.0 cm) and the shortest plant (≤50.0 cm), respectively.

Protein and oil content

Protein and oil content are two important characters in soybeans, and are related to the degree of evolution. The highest and lowest content of protein in soybeans are the consequence of artificial selection. We observed that soybeans with a protein content of 40.01–45.0% had a higher $Sg-1^a$ frequency (43.2%), followed by 45.01–50.0 (34.9%) protein content. The classes of soybeans with the highest and lowest content of protein had the lowest frequencies of the $Sg-1^a$ allele, 14.3 and 29.7%, respectively. However, all the soybean protein content classes contained higher frequencies of the $Sg-1^b$ allele than the $Sg-1^a$ allele (Table 2).

In oil content types, the degression of $Sg-1^a$ frequency occurred from the lowest content type (≤16.0%) to the highest content type (>22.0%), i.e., frequency gradually was reduced in turn from 44.5, 41.0, 34.9, 32.3 to 13.2% in five content class types, the lowest content (≤16.0%), low content

(16.01–18.0%), medium content (18.01–20.0%), high content (20.01–22.0%) and the highest content (>22.0%), respectively.

Discussion

Saponin composition and the significance of finding group α saponins in Chinese soybeans

Soybean group A saponins Aa, Ab, Ae, Ax, Af, Ad, DDMP βg, αg and group B Ba components are common main components (Kudou et al. 1992, 1993; Shiraiwa et al. 1990, 1991a, c; Tsukamoto et al. 1993). Honda et al. (2009) identified a new group of α saponins (i.e., Sg-6 saponins) in Japanese wild soybean accessions. Subsequently the group α was confirmed to exist in Korean and Chinese wild soybeans (Krishnamurthy et al. 2014a; Takahashi et al. 2016c). Takahashi et al. (2016c) reported further research of group α saponins, where four kinds of single components H-αg, I-αg and J-αg, and K-αg were recognized definitely in Chinese wild soybeans. However, thus far, no group α saponins have been found in cultivated soybeans. In our present investigation, a total of 13 components were detected in Chinese soybeans, including two new groups (α and β) in addition to common group A (Aa, Ab, Ae, Ax, Af, Ad), DDMP (βg, αg) and group B (Ba) components. Three components of group α saponins, H-αg, I-αg and K-αg, were first identified in cultivated soybeans. Another component KA-αg was also detected, which was a new recruit to saponins and belonged to group β. This KA-αg saponin component has recently been found in a Chinese semi-wild soybean (Takahashi et al. 2016a).

Tsukamoto et al. (1993) showed six phenotypes (Aa, Ab, AcAf, AaAb, AaBc, AbBc) of saponin composition in Japanese soybeans. In view of finding the existence of new saponins (groups α and β) in Chinese soybeans, a total of nine phenotypes of saponins were identified, four of which were common phenotypes of Aa, Ab, AaBc, AbBc, and five were new types of Aaβ$_{KA}$, Abα$_H$, Aaα$_I$, Abα$_I$, and Abα$_K$.

Up to the present, group α saponins are known to exist only in wild soybeans (Honda et al. 2009; Krishnamurthy et al. 2014a). The finding that group α saponins occurred in soybeans was of a far-reaching significance; the fact that group α saponins existed in cultivated soybeans testifies to the safety of this group of α saponins to human health because these saponins have been consumed for several 1000 years in China.

Alleles for soybean saponins in east Asia

Eight loci are proposed to be involved in saponin biosynthesis, GmSGT2, GmSGT3 (Shibuya et al. 2010), Sg-1 (Shiraiwa et al.1990; Tsukamoto et al. 1993), Sg-3, Sg-4 (Tsukamoto et al. 1993; Takada et al. 2012), Sg-5 (Tsukamoto et al. 1998; Takada et al. 2013), Sg-6 (Krishnamurthy et al. 2014b), and Sg-7 (Takahashi et al. 2016a). Recessive alleles at loci Sg-3 and Sg-5 were found in Japanese and Korean wild soybeans (Tsukamoto et al. 1993, 1998; Krishnamurthy et al. 2013); a recessive sg-3 would lead to the absence of Aa, or Ab and Ax or Ad components (Tsukamoto et al. 1993), and recessive sg-5, to none of the whole group A (Tsukamoto et al. 1998). The sg-3 allele only appeared in one Japanese soybean (Tsukamoto et al. 1993) and in very few Korean wild soybeans, and recessive sg-5 allele occurred in very few Japanese and Korean wild soybeans (Tsukamoto et al. 1998; Krishnamurthy et al. 2013). The Sg-4 allele is rare in soybeans but abundant in wild soybeans; about 2% of Japanese soybeans, 67.5% of Japanese wild soybeans, and about 61.4% of Korean wild soybeans possess the Sg-4 allele (Tsukamoto et al. 1993; Krishnamurthy et al. 2014a). There were a small number of Sg-4 allele in Chinese soybeans and it existed with a very low frequency (0.067%), only two of 2959 accessions. The recessive sg-7 allele was found only in one semi-wild type wild soybean (Takahashi et al. 2016a).

The locus Sg-1 is responsible for the formation of group A saponins (Sayama et al. 2012) and there are three alleles Sg-1^a, Sg-1^b and sg-1^0 (recessive allele for no terminal sugar at C-3 position) in this locus. The Sg-1^a allelic frequency was higher in Korean and Japanese wild soybeans, with 98.4 and 85.4% frequencies, respectively; whereas cultivated soybeans had a lower Sg-1^a frequency of about 26.6% (Tsukamoto et al. 1993; Krishnamurthy et al. 2014a). The recessive sg-1^0 allele also rarely existed with a frequency of 0.3% in both Korean and Japanese wild soybeans (Tsukamoto et al. 1993; Krishnamurthy et al. 2014a). We detected two co-dominant alleles Sg-1^a and Sg-1^b in Chinese soybeans, and the Sg-1^a allele showed a frequency of 39.3%, which was slightly

higher than the frequency of 26.6% in Japanese soybeans (Tsukamoto et al. 1993) but was much lower than the frequency of $Sg-1^a$ in *Glycine soja* (98.4% in South Korea and 85.4% in Japan). We do not know the reasons for the accumulation of the $Sg-1^b$ allele in cultivated soybean during the domestication from wild soybeans. A possible explanation might be that the $Sg-1^b$ allele could be of some benefit to cultivation adaptation or human preference, and/or to some physiological function, stress resistance, or other characters of soybeans by direct or indirect artificial selection during the domestication of soybeans.

The group α saponins were genetically thought to be controlled by a single $Sg-6$ locus (Krishnamurthy et al. 2014b), and approximately 10% of Korean wild soybeans carried the $Sg-6$ allele and 90% were recessive ($sg-6$), i.e., without group α (Krishnamurthy et al. 2013). However, our investigation detected three different components (H-αg, I-αg, and K-αg) of group α saponins in twelve Chinese soybeans, which suggests that there could be different genes or loci controlling these respective components of group α saponins. The accumulation of no α saponins in *G. max* agreed with the domestication of soybeans because group α saponins also exist at a relatively lower level in *G. soja* (Krishnamurthy et al. 2014a).

Unexpectedly, a new group, β saponin (KA-αg component), was first found in a Chinese cultivated soybean accession. This KA-αg component was only recently found in a semi-wild soybean lacking group A saponins in China (Takahashi et al. 2016a), which was proposed to be derived from the A-αg precursor of group A by acetylation at the C-29 position. The present discovery that a KA-αg component occurred in a normal Aa type accession (ZDD13997) suggested that this ZDD13997 accession possessed the acetylating enzyme and a small part of the A-αg precursor molecules become KA-αg by acetylation at the C-29 position, and that the acetylation or acetyl of KA-αg could prevent KA-αg from attaching arabinose (Ara) to the OH group at the C-22 position. As a result, both Aa and KA-αg components were detected in soybean ZDD13997.

Accumulation of $Sg-1^a$ allele, geographical distribution and origin of soybean

The $Sg-1^a$ allele accumulates at high frequency in the wild soybean species (Tsukamoto et al. 1993;

Krishnamurthy et al. 2014a); this peculiarity was also characterized by the evolutionarily degree-low or some particular character types in the soybeans analyzed here (Table 2). Smaller seeds (≤10.0 g/100-seed wt.), the black seed-coat, prostrate stem, indefinite growth habit, and spring-sowing type are thought to be primordial characters of soybeans, and accordingly these characters exhibited the highest accumulation level of the $Sg-1^a$ allele. The summer and autumn-sowing types are derived from spring-sowing types, therefore, they should have relatively lower frequencies of the $Sg-1^a$ allele. There was no frequency differentiation in the $Sg-1^a$ locus between types of cotyledon, pubescence and flower colours.

Stem erectness and growth habit characters represented a decreased inclination of $Sg-1^a$ allelic frequency with the degrees of character intensity, prostrate > semi-prostrate > weak-erect > erect in stem erectness and indefinite > semi-indefinite > definite in growth habit. These phenomena could be understood in consideration of the frequency decrease of $Sg-1^a$ allele incidental to the evolution of soybeans from the wild soybean. Three quantitative characters (plant height, protein and oil contents) appeared to have some correlations with the $Sg-1$ locus, showing gradually decreased frequencies of the $Sg-1^a$ allele from the biggest to the smallest classes (Table 2); we could not definitely explain the reasons for this; however, from a domestication perspective, the correlations between the $Sg-1$ locus and these quantitative characters accorded with the evolution of soybeans because the progenitor wild species *G. soja* always has higher plant height and lower oil content. The wild species has a higher average protein content of 40–45% than soybeans, and this protein content could belong to a primitive trait of soybeans, then the conscious or unconscious artificial selection towards a high or low content of protein in soybean domestication and breeding would bring about the decrease of $Sg-1^a$ allelic frequency from the primordial to evolutionarily acquired class types according to the logic observed here that the frequency of the $Sg-1$ locus varied regularly with the degree of evolution (Table 2).

We observed correlations between the $Sg-1$ locus and the domestication or evolution of soybeans within the cultivated species, therefore, the $Sg-1$ locus may be used as an indicator or a marker to probe the area of origin for soybean. With regard to

the area of origin of soybean in China, the clear area is unknown and there are long-standing disputes around several hypotheses, such as origination in the northeast (Fuguda 1933; Li 1994), north (Hymowitz and Newell 1981; Xu et al. 1986; Chang 1989; Li et al. 2013), south (Gai et al. 2000; Xu and Gai 2003; Zhao et al. 2004; Wen et al. 2009; Guo et al. 2010), and multiple centres in China (Lu 1978; Zhou 1998; Xu et al. 2002).

Around the area of origin of soybeans, the aforementioned mainly focus on whether the area of origin is in north (the Yellow River valley including the Huang-Huai plain) or in south China (the Yangtze River and its south). However, archaeological remains of wild and soybean seeds unearthed from over 30 sites of the Neolithic cultures in China in context dating from 7000 BC to the Longshan period (2350–1900 BC) are all located in the Yellow River valley, i.e., northern regions (Lee et al. 2007; Wu et al. 2013). Our present results obtained here exhibited higher frequencies of Sg-1^a allele in north and northwest China along the Yellow River valley. A recent report indicated a geographical area of the highest genetic diversity for Chinese soybeans on the basis of an SSR assessment, which lies 30°–35°N × 105°–110°E, including the valley of the middle and lower reaches of the Wei River and the valley of the upper reaches of the Hanjiang River (Wang et al. 2016). This area belongs to the southeastern Gansu Province and southern Shaanxi, and the middle-upper reaches of the Yellow River valley. In this investigation, group α saponins were first found in cultivated soybeans, which also implies that the group α saponin genes introgressed into the cultivars when the early domestication of soybeans started from wild soybeans. As a result of integrating the documents published for the archaeological remains of wild and cultivated soybean, the higher frequency distribution of the Sg-1^a allele observed here along the Yellow River valley and the appearance of group α soybean germplasm with a relatively higher frequency (seven of 13 group α landrace accessions) in northwestern Gansu and Shaanxi provinces, we more specifically propose that the area in the middle and lower reaches of the Wei River valley (Shaanxi and Gansu provinces), i.e., the middle-upper reaches of the Yellow River valley, is the possible area of origin of soybeans in China.

References

Chang RZ (1989) Studies on the origin of cultivated soybean. Oil Crop China 1:1–6 **(in Chinese)**

Ellington AA, Berhow MA, Singletary KW (2006) Inhibition of Akt signaling and enhanced ERK1/2 activity are involved in induction of macroautophagy by triterpenoid B-group soyasaponins in colon cancer cells. Carcinogenesis 27:298–306

Fenwick GR, Price KR, Tsukamoto C, Okubo K (1991) Saponins. In: D'Mello JPF, Duffus CM, Duffus JH (eds) Toxic Substances in crop plants. The Royal Society of Chemistry, Cambridge, pp 285–327

Fuguda Y (1933) Cytogenetical studies on the wild and cultivated Manchurian soybean (*Glycine* L.). Jpn J Bot 6:489–506

Gai JY, Xu DH, Shimamoto Y, Abe J, Hukushi H, Kitajima S (2000) Studies on the evolutionary relationship among ecotypes of *G. max* and *G. soja* in China. Acta Agron Sin 26:513–520 **(in Chinese)**

Guo J, Wang YS, Song C, Zhou JF, Qiu LJ, Huang HW, Wang Y (2010) A single origin and moderate bottleneck during domestication of soybean (*Glycien max*): implications from microsatellites and nucleotide sequences. Ann Bot 106:05–514

Honda N, Tsukamoto C, Maehara Y, Tayama I, Kitamura K, Singh RJ, Chung G (2009) Saponin composition of *Glycine soja* Sieb. & Zucc. mutants having new soyasapogenol aglycones and geographical distribution in South Korea. In: Proc World Soybean Res Conf VIII, Beijing, p 114

Hong SW, Yoo DH, Woo JY, Jeong JJ, Yang JH, Kim DH (2014) Soyasaponins Ab and Bb prevent scopolamine-induced memory impairment in mice without the inhibition of acetylcholinesterase. J Agric Food Chem 62:2062–2068

Hymowitz T, Newell C (1981) Taxonomy of genus *Glycine*, domestication and uses of soybeans. Econ Bot 35:272–288

Ishii Y, Tanizawa H (2006) Effects of soyasaponins on lipid peroxidation through the secretion of thyroid hormones. Biol Pharm Bull 29:1759–1763

Kang JH, Sung MK, Kawada T, Yoo H, Kim YK, Kim JS, Yu R (2005) Soybean saponins suppress the release of proinflammatory mediators by LPS-stimulated peritoneal macrophages. Cancer Lett 230:219–227

Kato S, Yumoto S, Takada Y, Kono Y, Shimada S, Sakai T, Shimada H, Takahashi K, Adachi T, Tabuchi K, Kikuchi A (2007) A new soybean cultivar 'Kinusayaka' lacking three lipoxygenase isozymes and group A acetyl saponin. Bull Natl Agric Res Cent Tohoku Reg 107:29–42 **(in Japanese)**

Krishnamurthy P, Tsukamoto C, Honda N, Kikuchi A, Lee JD, Yang SH, Chung G (2013) Saponin polymorphism in the Korean wild soybean (*Glycine soja* Sieb. and Zucc.). Plant Breed 132:121–126

Krishnamurthy P, Lee CM, Tsukamoto C, Takahashi Y, Singh RJ, Lee JD, Chung G (2014a) Evaluation of genetic structure of Korean wild soybean (*Glycine soja*) based on saponin allele polymorphism. Gene Resour Crop Evol 61:1121–1130

Krishnamurthy P, Tsukamoto C, Singh RJ, Lee JD, Kim HS, Yang SH, Chung G (2014b) The Sg-6 saponins, new components in wild soybean (*Glycine soja* Sieb. and Zucc.): polymorphism, geographical distribution and inheritance. Euphytica 198:413–424

Kudou S, Tonomura M, Tsukamato C, Shimoyamada M, Uchida T, Okubo K (1992) Isolation and structural elucidation of the major genuine soybean saponin. Biosci Biotechnol Biochem 56:142–143

Kudou SM, Tonomura C, Tsukamato T Uchida, Sakabe T, Tamura N, Okubo K (1993) Isolation and structural elucidation of DDMP-conjugated soyasaponins as genuine saponins from soybean seeds. Biosci Biotechnol Biochem 57:546–550

Kuzuhara H, Nishiyama S, Minowa N, Sasaki K, Omoto S (2000) Protective effects of soyasapogenol A on liver injury mediated by immune response in a concanavalin A-induced hepatitis model. Eur J Pharmacol 391:175–181

Lee SO, Simons AL, Murphy PA, Hendrich S (2005) Soysaponins lowered plasma cholesterol and increased fecal bile acids in female golden Syrian hamsters. Exp Biol Med 230:472–478

Lee GY, Crawford GW, Liu L, Chen XC (2007) Plants and people from the early Neolithic to Shang periods in north China. PNAS 104:1087–1092

Li FS (1994) Study on origin and evolution of soybean. Soybean Sci 13:61–66

Li YH, Zhang C, Smulders Marinus JM, Li W, Ma YS, Xu Q, Chang RZ, Qiu LJ (2013) Analysis of average standardized SSR allele size supports domestication of soybean along the Yelloe River. Genet Resour Crop Evol 60:767–776

Lu SL (1978) Discuss about origin of cultivated soybean in China. China Agric Sci 4:90–94 **(in Chinese)**

Murata M, Houdai T, Yamamoto H, Matsumori M, Oishi T (2006) Membrane interaction of soyasaponins in association with their antioxidation effect-analysis of biomembrane interaction. Soybean Protein Res 9:82–86 **(in Japanese)**

Okubo K, Iijima M, Kobayashi Y, Yoshikoshi M, Uchida T, Kudou S (1992) Components responsible for the undesirable taste of soybean seeds. Biosci Biotech Biochem 56:99–103

Rowlands JC, Berhow MA, Badger TM (2002) Estrogenic and antiproliferative properties of soy sapogenols in human breast cancer cells in vitro. Food Chem Toxicol 40:1767–1774

Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012) The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24:2123–2138

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett 584(11):2258–2264

Shiraiwa M, Yamauchi F, Harada K, Okubo K (1990) Inheritance of ''group A saponin'' in soybean seed. Agric Biol Chem 54:1347–1352

Shiraiwa M, Kudo S, Shimoyamada M, Harada K, Okubo K (1991a) Composition and structure of group A saponin in soybean seed. Agric Biol Chem 55:315–322

Shiraiwa M, Harada K, Okubo K (1991b) Composition and content of saponins in soybean seed according to variety, cultivation year and maturity. Agric Biol Chem 55:323–331

Shiraiwa M, Harada K, Okubo K (1991c) Composition and structure of group B saponin in soy seed. Agric Biol Chem 55:911–917

Takada Y, Tayama I, Sayama T, Sasama H, Saruta M, Kikuchi A, Ishimoto M, Tsukamoto C (2012) Genetic analysis of variations in the sugar chain composition at the C-3 position of soybean seed saponins. Breed Sci 61:639–645

Takada Y, Sasama H, Sayama T, Kikuchi A, Kato S, Ishimoto M, Tsukamoto C (2013) Genetic and chemical analysis of a key biosynthetic step for soyasapogenol A, an aglycone of group A saponins that influence soymilk flavor. Theor Appl Genet 126:721–731

Takahashi Y, Li XH, Tsukamoto C, Wang KJ (2016a) Identification of a novel variant lacking group A soyasaponin in a Chinese wild soybean (*Glycine soja* Sieb. & Zucc.): implications for breeding significance. Plant breed 135:607–613

Takahashi Y, Chitisankul WT, Ota H, Muraoka H, Varanyanond W, Wang KJ, Tsukamoto C (2016b) Profile analysis to elucidate chemical structures of saponin components in soyfoods. In: 7th Int Soybean Proc Uti Conf (ISPUC VII), Harbin, pp 150–152

Takahashi Y, Tsukamoto C, Li XH, Wang KJ (2016c) Polymorphic studies of soybean saponin composition in Chinese wild soybean accessions (*Glycine soja* Sieb. & Zucc.). In: 7th Int Crop Sci Cong (7th ICSC), Beijing, p 109

Takahashi Y, Li XH, Tsukamoto C, Wang KJ (2017) Categories and components of soyasaponin in the Chinese wild soybean (*Glycine soja*) genetic resource collection. Genet Resour Crop Evol. doi:10.1007/s10722-017-0506-4

Tsukamoto C, Yoshiki Y (2006) Soy saponin. In: Sugano M (ed) Soy in health and disease prevention. CRC Press, New York, pp 155–172

Tsukamoto C, Kikuchi A, Harada K, Kitamura K, Okubo K (1993) Genetic and chemical polymorphisms of saponins in soybean seed. Phytochemistry 34:1351–1356

Tsukamoto C, Kikuchi A, Harada K, Kitamura K, Iwasaki T, Okubo K (1998) Genetic improvement of saponin components in soybean seed. In: Chainuvati C, Sarobol N (eds) Soybean feeds the world. Proc (Suppl) 5th World Soybean Res Conf Kasetsart University Press, Bangkok, pp 80–85

Tsukamoto C, Yoshitake T, Kikuchi A, Ishimoto M, Shimada K, Kitamura K (2009) Processing and utilization of soybeans modified saponin and isoflavones characteristics for functional soy-based foods. In: Proc World Soybean Res conf VIII, Beijing, p 112

Wang LX, Lin FY, Li LH, Li W, Yan Z, Luan WJ, Piao RH, Guan Y, Ning XC, Zhu L, Ma YS, Dong ZM, Zhang HY, Zhang YQ, Guan RX, Li YH, Liu ZX, Chang RZ, Qiu LJ (2016) Genetic diversity center of cultivated soybean (*Glycine max*) in China—new insight and evidence for the

diversity center of Chinese cultivated soybean. J Integr Agri 15:2481–2487

Wen ZX, Zhao TJ, Ding YL, Gai JY (2009) Genetic diversity, geographical differentiation and evolutionary relationship among ecotypes of *Glycine max* and *G. soja* in China. Chin Sci Bull 54:4393–4403

Wu WW, Jin GY, Wang HY, Wang CM (2013) Application and Domestication of *Glycine* L. in ancient China. Agri Archaeol 6:1–10 **(in Chinese)**

Xu DH, Gai JY (2003) Genetic diversity of wild and cultivated soybeans growing in China reveald by RAPD analysis. Plant Breed 122:503–506

Xu B, Zheng HY, Lu QH, Zhao SW, Zou SH, Hu ZA (1986) Three new evidences of the original area of soybean. Soybean Sci 5:123–130 **(in Chinese)**

Xu DH, Abe J, Gai JY, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–653

Yang SH, Ahn EK, Lee JA, Shin TS, Tsukamoto C, Suh W, Itabashi M, Chung G (2015) Soyasaponins Aa and Ab exert an anti-obesity effect in 3T3-L1 adipocytes through downregulation of PPARγ. Phytother Res 29:281–287

Zhao TJ, Gai JY (2004) The origin and evolution of cultivated soybean [*Glycine max* (L.) Merr.]. Sci Agri Sin 37:954-962 **(in Chinese)**

Zhou XA (1998) The first discussion about center of genetic diversity of cultivated soybean in China. Chin Agric Sci 31:37–43 **(in Chinese)**