Euphytica (2014) 198:413–424
DOI 10.1007/s10681-014-1118-0

The Sg-6 saponins, new components in wild soybean (*Glycine soja* Sieb. and Zucc.): polymorphism, geographical distribution and inheritance

Panneerselvam Krishnamurthy · Chigen Tsukamoto · Ram J. Singh · Jeong-Dong Lee · Hong-Sig Kim · Seung-Hwan Yang · Gyuhwa Chung

Received: 3 July 2013 / Accepted: 14 April 2014 / Published online: 26 April 2014

Abstract Hypocotyl saponin composition of 1,198 accessions of wild soybean (*Glycine soja*) collected from China, Korea, Japan and Russia Far East was analyzed by thin-layer chromatography to determine polymorphic variation and geographical distribution. Eight common distinguishable saponin phenotypes were identified: Aa, Ab, AaBc, AbBc, Aa+α, Ab+α, AaBc+α and AbBc+α. The latter four +α type were new. All eight types were identified in China. Type Ab+α was absent in Korea. Types Ab+α and AbBc+α, and Aa+α and Ab+α were not identified in Japan and Russia far east, respectively. Six new triterpene saponins were detected in +α type via LC-PDA/MS/MS analyses. They were, tentatively, designated as H-αg, H-αa, I-αg, I-αa, J-αg and J-αa. These saponins were inherited together by a single dominant allele. A gene symbol *Sg-6* was assigned. Hence, the new saponins were collectively named as Sg-6 saponins. The frequency of *Sg-6* allele was 17.6 % in Chinese, 10.0 % in Korean and 1.0 % in Japanese wild soybean. The wild soybeans having Sg-6 saponins can be utilized in soybean breeding programs as well as in saponin biosynthesis studies in soybean.

Keywords *Glycine soja* Sieb. and Zucc. · Soyasaponin polymorphism · Triterpene glycosides · Wild soybean · Soybean

Introduction

Soybean [*Glycine max* (L.) Merr.] is the most important grain legume for the food and feed and used by pharmaceutical, cosmeceutical and bio-fuel industries in the world (Chung and Singh 2008). More than one-third of the world's edible oils and two-third of the

Panneerselvam Krishnamurthy and Chigen Tsukamoto have contributed equally to this work.

P. Krishnamurthy · G. Chung (✉)
Department of Biotechnology, Chonnam National University, Chonnam, Yeosu 550-749, Korea
e-mail: chung@chonnam.ac.kr

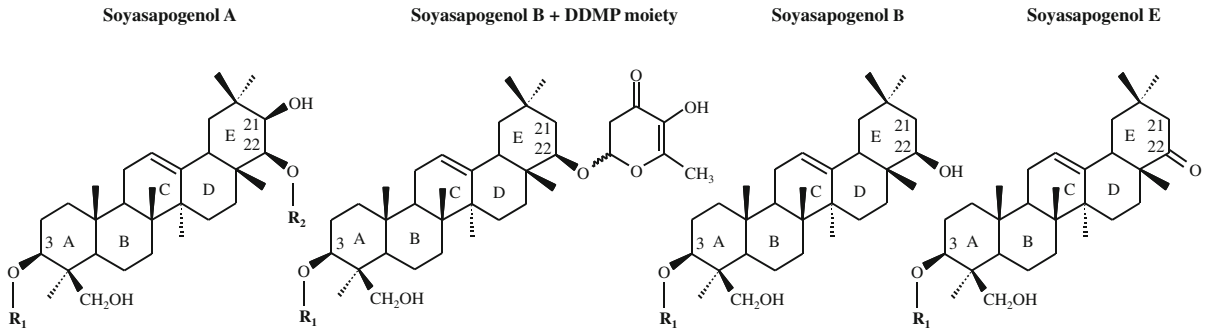
C. Tsukamoto
Department of Applied Biological Chemistry, Iwate University, Iwate, Morioka 020-8550, Japan

R. J. Singh
USDA/ARS, National Soybean Research Laboratory, University of Illinois, Urbana, IL 61801, USA

J.-D. Lee
Division of Plant Biosciences, Kyungpook National University, Daegu 702-701, Korea

H.-S. Kim
College of Agriculture, Life and Environment Sciences, Chungbuk National University, Cheongju 361-763, Korea

S.-H. Yang
Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Gyeonggi, Yongin 449-728, Korea



C-3 position (R_1)	C-22 position					
	Soyasapogenol A (R_2)			Soyasapogenol B		
	Group A			DDMP	Group B	Group E
	acetylXyl-Ara-	acetylGlc-Ara-	Ara-	DDMP-	HO-	O=
Glc-Gal-GlcUA-	Aa	Ab	A0-αg	αg	Ba	Bd
Rha-Gal-GlcUA-	Au	Ac	A0-βg	βg	Bb	Be
Gal-GlcUA-	Ae	Af	A0-γg	γg	Bb'	Be'
Glc-Ara-GlcUA-	Ax	Ad	A0-αa	αa	Bx	Bf
Rha-Ara-GlcUA-	Ay	Az	A0-βa	βa	Bc	Bg
Ara-GlcUA-	Ag	Ah	A0-γa	γa	Bc'	Bg'

Fig. 1 Chemical structure and nomenclature of soyasaponins. R_1 at the C-3 position of soyasapogenols A, B and E and R_2 at the C-22 position of soyasapogenol A are shown at the bottom of Fig. 1. The nomenclature of saponins used in this study is that of Shiraiwa et al. (1991a, b), Tsukamoto et al. (1993), Kudou et al. (1992, 1993), Kikuchi et al. (1999), and Takada et al. (2013)

world's protein meal are derived from soybean for human and animal diet as well as for non-edible uses, including industrial feedstock and combustible fuel (Masuda and Goldsmith 2009). Soybean also contains several health beneficial components such as isoflavones, saponins and lecithin (Sugano 2006). Wild soybean (*Glycine soja* Sieb. and Zucc.), the progenitor of soybean, is mainly distributed in China, the Korean Peninsula, Japan and Russia Far East (Chung and Singh 2008). This species is an excellent source of novel genes for soybean improvement; however, breeders have not extensively exploited wild soybeans in soybean breeding programs (Chung and Singh 2008).

Soyasaponins are bioactive secondary metabolites consisting of a triterpene (C_{30}) aglycone attached with one or two oligosaccharide sugar chains via glycosylation linkages (Tsukamoto and Yoshiki 2006). Several structurally diverse groups of triterpene glycosides have been identified and characterized from the seeds of cultivated and wild soybean (Kudou et al. 1992, 1993; Shiraiwa et al. 1991a, c; Tsukamoto et al. 1993). Based on the aglycone structure, they have been classified into group A and DDMP (2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one) saponins. Group A saponins having soyasapogenol A (SS-A; 3β,21β,22β,24-tetrahydroxyolean-12-ene) are bisdesmosides while DDMP saponins having soyasapogenol B (SS-B; 3β,22β,24-trihydroxyolean-12-ene) are monodesmosides (Fig. 1). Group B saponins having SS-B and group E saponins having soyasapogenol E (SS-E; 3β,22-one,24-dihydroxyolean-12-ene) are also monodesmosides produced from DDMP saponins during the most commercial sample extraction and processing procedures (Kudou et al. 1992, 1993). Nomenclature of all soyasaponins was derived from these three soyasapogenols and their sugar moiety composition (Fig. 1).

Saponin polymorphism has been extensively studied in soybean to determine the structural diversity and to identify new saponin (Kikuchi et al. 1999; Kudou et al. 1992, 1993; Sayama et al. 2012; Shiraiwa et al. 1991a, b, c; Takada et al. 2010, 2012; Tsukamoto et al. 1993). So far, four common saponin phenotypes (Aa, Ab, AaBc and AbBc) and 2 mutant saponin phenotypes (AcAf and A0-αg) have been identified in soybean seed hypocotyl (Kikuchi et al. 1999; Takada et al. 2010, 2012; Tsukamoto et al. 1993). These

saponin phenotypes can be explained by the combination of dominant, codominant and recessive genes of the three gene loci namely *Sg-1*, *Sg-3* and *Sg-4* (Tsukamoto et al. 1993). Wild soybean accessions are evaluated when essential saponin mutants within soybean collections are not found. This resulted in identification of two Chinese wild soybean accessions (GD50029-2 and GD50326-2) with unknown saponins (α). These two accessions are previously designated as mutants (Honda et al. 2009).

When we analyzed the hypocotyl saponin composition of 3,720 accessions of Korean wild soybeans, we observed 10 % of the Korean wild soybeans contain α saponins which can produce four new saponin phenotypes (+α types: Aa+α, Ab+α, AaBc+α and AbBc+α) (Krishnamurthy et al. 2013, 2014). Based on this study, we hypothesized: i) accumulation of α saponins is not a mutational property, ii) the four new +α types are common saponin phenotypes, and iii) α saponins could also be frequently present in wild soybeans of China, Japan and Russia Far East. The objective of this study was to examine the hypocotyl saponin composition, polymorphism, geographical distribution and inheritance of α saponins (Sg-6 saponin) of wild soybeans from China, Korea, Japan and Russia Far East.

Materials and methods

Germplasm sources and chemicals

Of the 1,198 wild soybean accessions examined in this study, 526 accessions were from Japan, 373 accessions were from Korea (selected from our previous studies [Krishnamurthy et al. 2013, 2014]), 285 accessions were from China and 14 accessions were from Russia far east (Table 1). These accessions are being maintained either in Chung's wild legume germplasm collection (CWLGC) at the Chonnam National University, Yeosu, Chonnam, Korea or in the rural development administration (RDA), Suwon, Korea (list available from the authors). Two *G. max* cultivars Taekwang and Saeolkong, included in genetic study, were obtained from RDA. All chemicals used in this study were analytical grade and were purchased from Honeywell Burdick and Jackson, Seoul, Korea and Samchun Chemicals, Seoul, Korea.

Saponin extraction

Hypocotyl of five mature dry seeds from each accession was used in this study. Ten-fold volumes (v/w) of 80 % (v/v) aqueous methanol were added to extract saponins from intact hypocotyls. Extractions were carried out at room temperature for 24 h. The resulting extracts were stored at 4 °C and were directly analyzed by TLC (thin layer chromatography). When samples showed +α saponin phenotype in TLC, they were further analyzed by LC-PDA/MS/MS (liquid chromatography–photodiode array detector/mass spectrometry/mass spectrometry).

Thin layer chromatography analysis

Thin layer chromatography was performed according to Krishnamurthy et al. (2012). Briefly, 10 μL from each sample extract was directly applied on silica gel (SiO_2) coated TLC plates with an Eppendorf micropipette and slightly dried by using a hair drier. The plates were developed in a rectangular developing chamber which was saturated with the lower phase of chloroform:methanol:water (65:35:10, v/v/v) for 2 h. Plates were dried at 100 °C for 10 min and then developed with 10 % H_2SO_4 for 12 min in a closed chamber. Saponins were visualized by heating the plates at 115 °C for 13 min.

LC-PDA/MS/MS analysis

The crude hypocotyl extracts were diluted ten times with 80 % methanol prior to use. Ten micro liters from each extracts were analyzed in a UFLC system (Prominence UFLC system, Shimadzu, Kyoto, Japan) equipped with a photodiode array detector (PDA) and a tandem mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, Yokohama, Kanagawa, Japan) on a C30 reverse phase column (Develosil C30-UG-3, 2.0 mm I.D. × 150 mm, Nomura Chemical, Seto, Okayama, Japan) at 40 °C. Solvent A consisted of acetonitrile with 0.1 % (v/v) formic acid, and solvent B consisted of water with 0.1 % formic acid. A linear gradient elution of solvent A was performed at a flow rate of 0.15 mL/min: solvent A was initiated at 10 % (v/v) and increased to 90 % (v/v) over 80 min, and then increased to 100 % (v/v) for 5 min. The eluent composition was returned to the initial state of 10 % (v/v) solvent A for 15 min. The

Table 1 Saponin phenotypes and geographical locations of wild soybean accessions used in this study

Germplasm collection sites	No. of total accessions	Number of accessions								Saponin allele frequency[a] (%)			
		Aa *Sg-1*^{a} *sg-4* *sg-6*	Ab *Sg-1*^{b} *sg-4* *sg-6*	AaBc *Sg-1*^{a} *Sg-4* *sg-6*	AbBc *Sg-1*^{b} *Sg-4* *sg-6*	Aa + α *Sg-1*^{a} *sg-4* *Sg-6*	Ab + α *Sg-1*^{b} *sg-4* *Sg-6*	AaBc + α *Sg-1*^{a} *Sg-4* *Sg-6*	AbBc + α *Sg-1*^{b} *Sg-4* *Sg-6*	*Sg-1*^{a}	*Sg-1*^{b}	*Sg-4*	*Sg-6*
Japan													
Hokkaido	19	4	9	6									
Tohoku	136	33	6	94	2	1							
Kanto	14	1		9	2	2							
Chubu	60	20	1	36	2			1					
Kansai	71	2		69									
Chugoku	20	1		18	1								
Shikoku	6			6									
Kyushu	200	9		179	11			1					
Subtotal	526	70	16	417	18	3		2		93.6	6.4	83.1	1.0
Frequency (%)		13.3	3.0	79.3	3.4	0.6		0.4					
Korea[a]													
Gyeonggi-do	17					6		11					
Chungcheongbuk-do	9					6		3					
Chungcheongnam-do	33					14		14	5				
Jeollabuk-do	26							25	1				
Jeollanam-do	201					50		147	4				
Gyeongsangnam-do	47					16		31					
Gyeongsangbuk-do	13					4		9					
Gangwon-do	26					20		6					
Jeju-do	1							1					
Subtotal	373					116		247	10	98.5	1.1	61.8	10.0
Frequency[a]		34.7	0.4	54.0	0.5	3.1		6.6	0.3				
China													
Heilongjiang	28	10	4	14									
Jilin	112	35	11	14	11	15	10	7	9				
Liaoning	49	29	3	10		6		1					
Shandong	30	24	2	3	1								

Table 1 continued

Germplasm collection sites	No. of total accessions	Number of accessions								Saponin allele frequency[a] (%)			
		Aa *Sg-1*^{a} *sg-4* *sg-6*	Ab *Sg-1*^{b} *sg-4* *sg-6*	AaBc *Sg-1*^{a} *Sg-4* *sg-6*	AbBc *Sg-1*^{b} *Sg-4* *sg-6*	Aa + α *Sg-1*^{a} *sg-4* *Sg-6*	Ab + α *Sg-1*^{b} *sg-4* *Sg-6*	AaBc + α *Sg-1*^{a} *Sg-4* *Sg-6*	AbBc + α *Sg-1*^{b} *Sg-4* *Sg-6*	*Sg-1*^{a}	*Sg-1*^{b}	*Sg-4*	*Sg-6*
Jiangsu	11	5	1	2	3								
Zhejiang	11	4		7									
Hebei	8	8											
Henan	4	4											
Shanxi	26	20	2	3	1								
Ningxia	5	2	1				2						
Inner Mongolia	1	1											
Subtotal	285	141	24	54	16	21	12	8	9	78.6	21.4	30.5	17.6
Frequency (%)		49.5	8.4	18.9	5.6	7.4	4.2	2.8	3.2				
Russia far east													
Khabarovsk		1		1									
Primorsky		3		1				2	2				
Amur		3		1									
Subtotal	14	7		3				2	2				
Total	1,198												

[a] 373 Korean wild soybean accessions, saponin phenotype frequency and saponin allele frequency of Korean wild soybeans ($n = 3{,}720$) are utilized from our previous studies (Krishnamurthy et al. 2013, 2014)

Table 2 Segregation of unknown saponin components (*Sg-6* allele) in two F_2 populations

Parents and their progenies	Predicted genotype for the *Sg-6* locus	Number of seeds		χ^2 value	Probability
		Observed	Expected		
Cross 1					
P_1: Taekwang	*sg-6*				
P_2: CWS0857	*Sg-6*				
F_1 individuals	*Sg-6*	5			
F_2 individuals	*sg-6*	35	35.75	0.02	0.88
	Sg-6	108	107.25		
Cross 2					
P_1: Saeolkong	*sg-6*				
P_2: CWS0857	*Sg-6*				
F_1 individuals	*Sg-6*	5			
F_2 individuals	*sg-6*	25	23.75	0.09	0.77
	Sg-6	70	71.25		

eluate from the column was monitored by a PDA detector at UV 205 and 292 nm and by a tandem mass spectrometer in the positive ion mode of electrospray ionization [ESI(+)] method. An automatic full scan mode over a mass-to-charge ratio (m/z) range from 300 to 1,800 and the top three ion-trap mode were used to acquire MS and MS/MS data, respectively. The UV and MS spectra were recorded and analyzed with Xcalibur software version 2.1 (Thermo Fisher Scientific, Yokohama, Kanagawa, Japan).

Genetic study

Two *G. max* cultivars Taekwang (having allele *sg-6*) and Saeolkong (*sg-6*) and one *G. soja* accession CWS0857 (*Sg-6*) were hybridized for developing segregating populations for genetic study (Table 2). In each cross, the hypocotyl of randomly selected F_2 seeds (143 seeds from Taekwang × CWS0857 and 95 seeds from Saeolkong × CWS0857) were analyzed separately in LC-PDA/MS/MS. 50-fold volumes (v/w) of 80 % (v/v) aqueous methanol were used to extract saponins from individual hypocotyl.

Results

Polymorphism in hypocotyl saponin composition of wild soybean

Hypocotyl saponin composition of wild soybeans exhibited eight common distinguishable phenotypes: Aa, Ab, AaBc, AbBc, Aa+α, Ab+α, AaBc+α and AbBc+α (Fig. 2). The frequency of these 8 types in Chinese wild soybean was 49.5, 8.4, 18.9, 5.6, 7.4, 4.2, 2.8 and 3.2 %, respectively (Table 1). In Japanese wild soybean, AaBc (79.3 %) was predominant, Aa (13.3 %) was moderate, Ab (3.0 %) and AbBc (3.4 %) were subordinate, Aa+α (0.6 %) and AaBc+α (0.4 %) were rare, and Ab+α and AbBc+α were not detected (Table 1). In Chinese wild soybean, allele *Sg-1*^{a} was predominantly detected (78.6 %); alleles *Sg-4*, *Sg-1*^{b} and *Sg-6* were moderately detected at the frequency of 30.5, 21.4 and 17.6 %, respectively (Table 1). Alleles *Sg-1*^{a} (93.6 %) and *Sg-4* (83.1 %) were predominant, *Sg-1*^{b} (6.4 %) was subordinate and *Sg-6* (1.0 %) was rare in Japanese wild soybean (Table 1). We did not examine the saponin phenotype frequencies of Russian Far East wild soybeans since we have only 14 accessions. Of the 14 accessions of Russian far east, seven accessions showed Aa, three accessions showed AaBc, two accessions showed AaBc+α and two accessions showed AbBc+α (Table 1).

Detection of unknown saponins and soyasapogenols

Of the eight common saponin phenotypes, Aa+α, Ab+α, AaBc+α and AaBc+α types showed blue color saponin bands (α) in TLC analysis, which were retained just below the DDMP saponins (Fig. 2). In LC-PDA/MS/MS analysis, six unknown saponin components (1–6) were detected in the seed hypocotyl

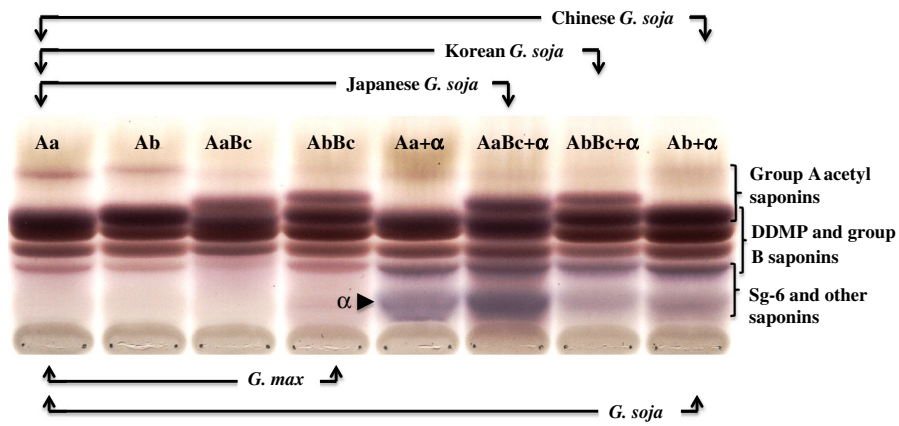


Fig. 2 Common saponin phenotypes of seed hypocotyls of *G. soja* populations by thin layer chromatography (TLC). TLC patterns of saponin components are mainly divided into three groups, group A acetyl saponins (*upper area*), DDMP and group B saponins (*middle*), and other saponins (*bottom*), which contains Sg-6 saponins (α) in this condition. Phenotypes were indicated at the *top* of each lane

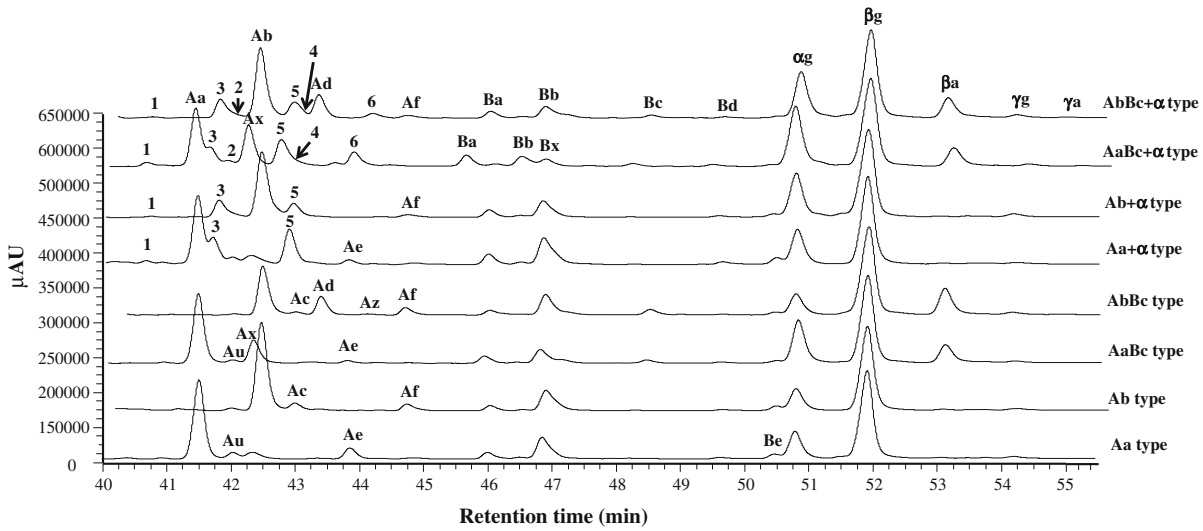
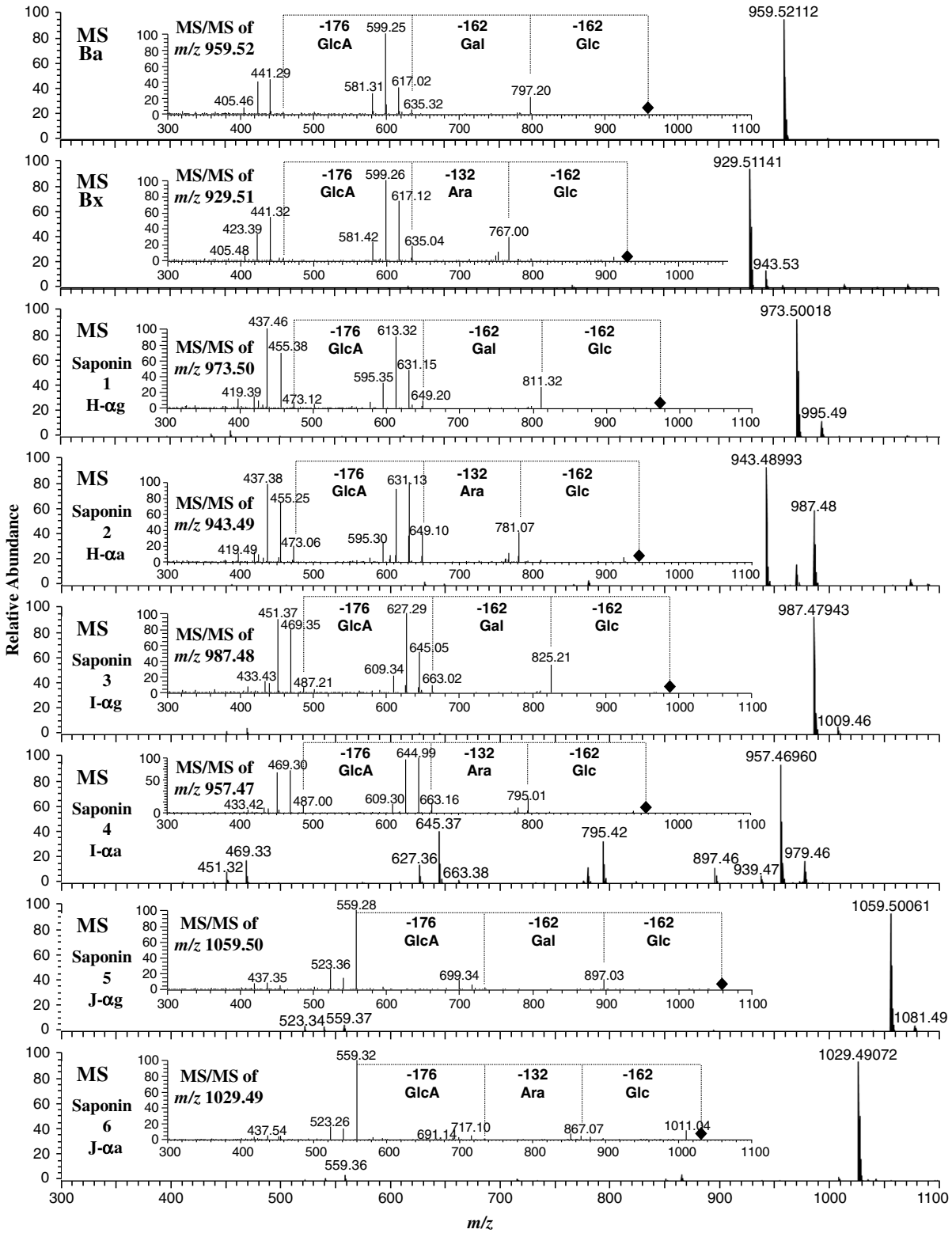


Fig. 3 LC-PDA/MS/MS analysis of the hypocotyl extracts of eight common saponin phenotypes. Saponin components were monitored by a PDA detector at UV 205 nm. Six unknown saponin components 1, 2, 3, 4, 5, and 6 were detected within a range of group A saponins. Phenotypes AaBc+α and AbBc+α contained all of six components while Aa+α and Ab+α contained only components 1, 3 and 5. Saponin composition is specific to each accession and phenotype

extracts of wild soybean accessions showing blue color bands in TLC. These saponins were detected within a range of group A saponins (Fig. 3). Phenotypes AaBc+α and AbBc+α contained all of six components while Aa+α and Ab+α contained only components 1, 3 and 5 (Fig. 3). ESI(+) MS of saponins 1–6 showed molecular ion $[M+H]^+$ peaks at m/z 973.5001, 943.4899, 987.4794, 957.4696, 1059.5006 and 1029.4907, respectively (Fig. 4). The deduced molecular mass of these saponins were $C_{48}H_{77}O_{20}$ (calcd. 973), $C_{47}H_{75}O_{19}$ (calcd. 943), $C_{48}H_{75}O_{21}$ (calcd. 987), $C_{47}H_{73}O_{20}$ (calcd. 957), $C_{51}H_{79}O_{23}$ (calcd. 1059), and $C_{50}H_{77}O_{22}$ (calcd. 1029), respectively.

MS/MS fragments of molecular ion peaks of saponins 1–6 are inserted in each MS spectrum


MS Ba
MS/MS of m/z 959.52
-176 GlcA
-162 Gal
-162 Glc
959.52112
MS Bx
MS/MS of m/z 929.51
-176 GlcA
-132 Ara
-162 Glc
929.51141
MS Saponin 1 H-αg
MS/MS of m/z 973.50
973.50018
MS Saponin 2 H-αa
MS/MS of m/z 943.49
943.48993
MS Saponin 3 I-αg
MS/MS of m/z 987.48
987.47943
MS Saponin 4 I-αa
MS/MS of m/z 957.47
957.46960
MS Saponin 5 J-αg
MS/MS of m/z 1059.50
1059.50061
MS Saponin 6 J-αa
MS/MS of m/z 1029.49
1029.49072
Relative Abundance
m/z

◀**Fig. 4** ESI(+)/MS spectrum of group B (Ba and Bx) and unknown saponin components (1–6). Sugar chain sequence of Ba and Bx are identical with their genuine saponins DDMP-αg and DDMP-αa, respectively. Saponins 1–6 were tentatively named as saponins H-αg, H-αa, I-αg, I-αa, J-αg and J-αa, respectively (see "Discussion" section). MS/MS fragments of molecular ion peaks of all saponins were inserted in each MS spectrum

(Fig. 4). MS/MS fragments of saponin 1 showed the sequence-specific prominent ions for the loss of $[M\text{-glc}(162)+H]^+$ at m/z 811.32, $[M\text{-glc}(162)\text{-gal}(162)\text{-}2H_2O(36)+H]^+$ at m/z 627.29 and $[M\text{-glc}(162)\text{-gal}(162)\text{-glcUA}(176)\text{-}2H_2O(36)+H]^+$ at 451.37 from the sugar chain attached at the C-3 position of the unknown aglycone. Similarly, MS/MS fragments of saponin 2 showed the sequence-specific prominent ions for the loss of $[M\text{-glc}(162)+H]^+$ at m/z 975.01, $[M\text{-glc}(162)\text{-ara}(132)\text{-}H_2O(18)+H]^+$ at m/z 644.99 and $[M\text{-glc}(162)\text{-ara}(132)\text{-glcUA}(176)\text{-}H_2O(18)+H]^+$ at 469.30 from the C-3 position sugar chain of the unknown aglycone. These sequence-specific prominent ions found in saponins 1 and 2 suggested that they have the same aglycone molecule, tentatively named soyasapogenol H (SS-H), with the molecular mass $[M+H]^+$ of 473 ($C_{30}H_{48}O_4$). While saponins 3 and 4 showed similar sequence-specific prominent ions to those of saponins 1 and 2, respectively. Thus, saponins 3 and 4 have the same aglycone, tentatively named soyasapogenol I (SS-I), whose molecular mass $[M+H]^+$ was 487 ($C_{30}H_{47}O_5$). Saponins 5 and 6 also showed similar sequence-specific prominent ions to those of saponins 1 and 2, respectively. Hence, saponins 5 and 6 have the same aglycone, tentatively named soyasapogenol J (SS-J), with the molecular mass $[M+H]^+$ of 559 ($C_{33}H_{51}O_7$).

Inheritance of unknown saponin components

The genetic inheritance of unknown saponins was examined in the F_1 and F_2 population derived from the two crosses: (i) Taekwang (*sg-6*) × CWS0857 (*Sg-6*), (ii) Saeolkong (*sg-6*) × CWS0857. All F_1 seeds contained unknown saponins. In F_2 seeds from each cross, a 3:1 ratio was observed (Table 2). This suggests the unknown saponins were all inherited together by a single dominant gene. A gene symbol *Sg-6* was assigned. Hereafter, the newly identified unknown saponins were collectively designated as Sg-6 saponins.

Discussion

Geographical distribution of hypocotyl saponin composition in wild soybean populations

Twelve allelic genes including *Sg-6* (*Sg-1*^{a}*/Sg-1*^{b}*/sg-1*$^{0-a}$*/sg-1*$^{0-b}$, *Sg-3/sg-3*, *Sg-4/sg-4*, *Sg-5/sg-5* and *Sg-6/sg-6* on five loci *Sg-1*, *Sg-3*, *Sg-4*, *Sg-5* and *Sg-6*, respectively), have been reported in the biosynthesis of soyasaponins (Sasama et al. 2010; Sayama et al. 2012; Takada et al. 2010, 2012, 2013; Tsukamoto et al. 1993). Alleles *Sg-1*^{a} and *Sg-1*^{b} are co-dominant at the *Sg-1* locus while *sg-1*$^{0-a}$ and *sg-1*$^{0-b}$ are recessive at the same locus (Kikuchi et al. 1999; Sayama et al. 2012; Takada et al. 2010). Allele *Sg-1*^{a}, which controls the addition of xylose sugar moiety at the terminal position of the C-22 sugar chain of SS-A, was predominant in Korean wild soybean (98.5 %) while the frequency of that in Japanese wild soybean (93.6 %) and Chinese wild soybean (78.6 %) was relatively low. Only Chinese wild soybean had high frequency of *Sg-1*^{b} allele (21.4 %), which adds glucose sugar moiety at the terminal position of the C-22 sugar chain of SS-A (Table 1). Allele *Sg-4* was more frequently found in Japanese wild soybean (83.1 %) and Korean wild soybean (61.8 %) than in Chinese wild soybean (30.5 %). New allele *Sg-6* was found 17.6 % in Chinese, 10.0 % in Korean and 1.0 % in Japanese wild soybean (Table 1). In conclusion, the frequency of *Sg-1*^{a} was high in Korea, *Sg-4* was high in Japan, and *Sg-1*^{b} and *Sg-6* was high in China. Concurrently, the existence of *Sg-1*^{b} was very rare in Korea and *Sg-6* was very rare in Japan. These results show that the distribution of saponin alleles has significant geographical differences.

Although twelve allelic genes accounted for the soyasaponin polymorphism, only the combination of six allelic genes (*Sg-1*^{a}*/Sg-1*^{b}, *Sg-4/sg-4* and *Sg-6/sg-6*) contributed in the production of eight common hypocotyl saponin phenotypes. Alleles *Sg-3* and *Sg-5* are found to be dominant in all the analyzed accessions except the mutants (Krishnamurthy et al. 2013, 2014; Takada et al. 2013). Extensive saponin analysis of Korean wild soybean identified seven common phenotypes: Aa (34.7 %), Ab (0.4 %), AaBc (54.0 %), AbBc (0.5 %), Aa+α (3.1 %), AaBc+α (6.6 %) and AbBc+α (0.3 %) (Krishnamurthy et al. 2013, 2014). Type Ab+α has not been detected in Korea. In this study, all of the eight common phenotypes were

identified in Chinese wild soybean. Types Ab+α and AbBc+α, and Aa+α and Ab+α were not identified in Japanese and Russia Far East accessions, respectively (Table 1; Fig. 2). We believe that, if we analyze more accessions from Russia Far East, we may detect the missing Aa+α and Ab+α type. However, the possibility of existence of Ab+α and AbBc+α type in Japanese wild soybean is very low. Because, in Japan, extensive saponin analysis of wild soybean identified only two accessions [GD50029-2 (Aa+α) and GD50326-2 (AaBc+α)] carrying *Sg-6* allele (Honda et al. 2009). Tsukamoto et al. (1993) reported AaBc (58.4 %) and Aa (21.6 %) types were dominant followed by Ab (9.7 %) and AbBc (4.6 %) types in Japanese wild soybeans. These frequencies were quite different from our results (Table 1). It may because the geographical places of Japanese wild soybean accessions used in this study were different from those examined by Tsukamoto et al. (1993).

In this study, though we screened a small number of wild soybean accessions, we detected all four forms of +α phenotypes (Aa+α, Ab+α, AaBc+α and AaBc+α) in China, two forms (Aa+α and AaBc+α) in Japan and two forms (AaBc+α and AbBc+α) in Russia Far East (Table 1; Fig. 2). This shows that wild soybeans with +α phenotypes (Sg-6 saponins) are not mutants.

Partial characteristics of Sg-6 saponins

Soyasapogenols A, B and E, are the three aglycones so far reported in the cultivated and wild soybeans (Kudou et al. 1992, 1993; Shiraiwa et al. 1991a, c; Tsukamoto et al. 1993). They contain five triterpene rings (A, B, C, D, and E) and differ from each other at the C-21 and C-22 positions of soyasapogenols (Fig. 1). Though the chemical structures of SS-H, -I and -J are not elucidated yet, preliminary structure analysis by the combination of ^{1}H-NMR, ^{13}C-NMR, UV and IR spectrum, and high resolution MS and MS/MS analysis suggested that these soyasapogenols are different from one another at the part of their E-ring (Takahashi et al. 2013). The sugar chain sequence attached at the C-3 position of saponins 1, 3, and 5 is same with those of DDMP saponin αg, while the C-3 sugar chain of saponins 2, 4, and 6 is same with those of DDMP saponin αa (Fig. 4). Based on the sugar chain composition and the molecular mass of aglycone, saponins 1, 2, 3, 4, 5, and 6 have been, tentatively, designated as H-αg, H-αa, I-αg, I-αa, J-αg and J-αa, respectively.

Saponins H-αa, I-αa and J-αa, having Glc-Ara-GlcUA- sugar chain at the C-3 position of soyasapogenols, were not detected in the hypocotyl of wild soybeans showing Aa+α and Ab+α phenotypes. This agrees with the previous results that saponins having an arabinose (Ara) (arabinosyl saponins) at the second position of the C-3 sugar chain (Glc-Ara-GlcUA- or Rham-Ara-GlcUA or Ara-GlcUA-) of soyasapogenols (see Fig. 1) are not detected in the hypocotyl of soybean and wild soybean showing Aa and Ab phenotypes (Tsukamoto et al. 1993). This is because the hypocotyls of phenotypes Aa, Ab, Aa+α and Ab+α contain a recessive allele (*sg-4*) instead of dominant allele (*Sg-4*) at the *Sg-4* locus which controls the arabinosylation of the hydroxyl group of the C-2″ position of glucuronic acid attached at the C-3 position of soyasapogenols (Takada et al. 2012; Tsukamoto et al. 1993). Although saponins H-αa, I-αa, J-αa were not detected in the seed hypocotyls of Aa+α and Ab+α phenotypes (allele combination: *sg-4/Sg-6*), they were produced and detected in the hypocotyls of F_1 hybrid seeds (*Sg-4/Sg-6*) obtained from the crosses between AaBc type (*Sg-4/sg-6*) and Aa+α type (*sg-4/Sg-6*) (data not shown). This suggests, in F_1 hybrid seed, the gene locus *Sg-4* (arabinosyl transferase) from AaBc type used the soyasapogenols H, I and J from Aa+α type to produce saponins H-αa, I-αa and J-αa. It implies the fact that the soyasapogenols H, I and J can act as soyasaponin aglycones and that they can be utilize as the substrates for glycosyltransferases to produce soyasaponin glycosides in the biosynthetic pathway.

Saponins H-αg, H-αa, I-αg, I-αa, J-αg and J-αa were inherited together by *Sg-6* allele and were collectively named as Sg-6 saponins. Sg-6 saponins are made up of three different soyasapogenols H, I and J (Fig. 4). Hence, we presumed that the *Sg-6* allele directly and/or indirectly controls the presence of soyasapogenols H, I and J. Then, the soyasaponin glycosyltransferases use those soyasapogenols as substrates to produce Sg-6 saponins. It is quite difficult to explore clearly how a single gene controls the production of 3 different soyasapogenols (H, I and J) without further functional molecular research work. We propose two possibilities: (i) the product of *Sg-6* gene acts as a key component to produce one precursor from which all three soyasapogenols (H, I and J) are

produced, (ii) the presence of SS-H, SS-I and SS-J may depend on the presence of each other. More research is required to establish the chemical structure of these soyasapogenols and their relationship with the known soyasapogenols.

Future prospects

Since wild soybean is the progenitor of soybean (Chung and Singh 2008), Sg-6 saponins may also possibly exist in soybean. Saponin composition analysis was extensively conducted in Japanese soybean collection and found no soybean with Sg-6 saponins (Shiraiwa et al. 1991a, b, c, Tsukamoto et al. 1993). However, it was not extensively studied in Chinese and Korean soybean collection. Therefore, the comprehensive examination of saponin polymorphism in large number of soybean accessions from China and Korea is needed to provide a better understanding of the saponin relationship between soybean and wild soybean.

Acknowledgments We thank Dr. Sang-Mi Sun for maintaining the germplasm collection. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0013600).

References

Chung G, Singh RJ (2008) Broadening the genetic base of soybean: a multidisciplinary approach. Crit Rev Plant Sci 27:295–341

Honda N, Tsukamoto C, Maehara Y, Tayama I, Kitamura K, Singh RJ, Chung GH (2009) Proceedings of world soybean research conference VIII, Electronic Press on CD, No Pages, Beijing, China

Kikuchi A, Tsukamoto C, Tabuchi K, Adachi T, Okubo K (1999) Inheritance and characterization of a null allele for group A acetyl saponins found in a mutant soybean [*Glycine max* (L.) Merrill]. Breed Sci 49:167–171

Krishnamurthy P, Tsukamoto C, Yang SH, Lee JD, Chung G (2012) An improved method to resolve plant saponins and sugars by TLC. Chromatographia 75:1445–1449

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, Yuya T, Singh RJ, Lee JD, Chung G (2014) Evaluation of genetic structure of Korean wild soybean (Glycine soja) based on saponin allele polymorphism. Genet Resour Crop Evol. doi:10.1007/s10722-014-0095-4

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 S, Tonomura M, Tsukamato C, Uchida T, 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

Masuda T, Goldsmith PD (2009) World soybean production: area harvested, yield and long-term projections. Int Food Agribus Manag Rev 12:143–162

Sasama H, Takada Y, Ishimoto M, Kitamura K, Tsukamoto C (2010) Estimation of the mutation site of a soyasapogenol A-deficient soybean [Glycine max (L.) Merr.] by LC-MS/MS profile analysis. In: Cadwallader KR, Chang SKC (eds) Chemistry, texture, and flavor of Soy. Oxford University Press, UK, pp 91–102

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

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

Sugano M (2006) Soy in health and disease prevention. CRC Press, Taylor and Francis group, New York

Takada Y, Sayama T, Kikuchi A, Kato S, Tatsuzaki N, Nakamoto Y, Suzuki A, Tsukamoto C, Ishimoto M (2010) Genetic analysis of variation in sugar chain composition at the C-22 position of group A saponin in soybean, *Glycine max* (L.) Merrill. Breed Sci 60:3–8

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 Gen 126:721–731

Takahashi Y, Kon T, Muraoka H, Ishimoto M, Tsukamoto C (2013) The dominant Sg-6 synthesizes saponins with a keton function at oleanane aglycone C-22 position in soybean [Glycine max (L.) Merr.]. 11th international meeting on biosynthesis, function and biotechnology of isoprenoids in terrestrial and marine organisms (TERPNET), Kolymvari, Crete, Greece, p.197

Tsukamoto C, Yoshiki Y (2006) Soy saponin. In: Sugano M (ed) Soy in health and disease prevention. CRC Press, Taylor and Francis group, 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. Phytochem 34:1351–1356