

Categories and components of soyasaponin in the Chinese wild soybean (*Glycine soja*) genetic resource collection

Yuya Takahashi · Xiang-Hua Li · Chigen Tsukamoto · Ke-Jing Wang

Received: 28 October 2016 / Accepted: 27 February 2017 / Published online: 16 March 2017
© Springer Science+Business Media Dordrecht 2017

Abstract Soyasaponins, natural chemical compounds in soybean seeds, have important health benefits and also influence soyfood taste. A total of 3795 wild soybean (*Glycine soja*) accessions, collected from almost all the areas where wild soybean grows in China, were analysed using thin-layer chromatography to clarify saponin composition polymorphisms within Chinese wild soybeans. Twenty-three saponin components were distinguished, including four new components. The four new saponin components were a *Sg-6* saponin K- α g [29-acetyl H- α g] in 18 accessions, an HAb- α g [29-hydroxy Ab] in an Ab type accession, an A- α g [3-*O*-(Glc-Gal-GlcUA)-soyasapogenol A (SS-A)] which is a biosynthetic precursor compound of the Aa and Ab types and its derivative KA- α g [29-*O*-acetyl A- α g] in the *sg-7* mutant. Eight *sg-1⁰* variant accessions without the second sugar (xylose or glucose) at the C-22 position and one high-content group-E saponin Bd and one high-content group-A saponin Ae accessions were detected. Our present results can be expected to play vital roles in clarifying the biosynthetic pathway for

soyasaponins. We suggest that the A- α g is the precursors for group-A and Bd components, and the *Sg-6* saponin is biosynthesized from the Bd component in the relationship between the biosyntheses of group-A and *Sg-6* saponins. The above new components and variants contribute to develop healthy and functional foods in soybean breeding programs.

Keywords *Glycine soja* · Soyasaponin · Thin layer chromatography · Wild soybean

Introduction

Wild soybean (*Glycine soja* Sieb. et Zucc.), which is the progenitor of cultivated soybeans (*Glycine max* (L.) Merr.), contains many useful genes or characters, and is an important genetic resource for soybean breeding and genetic engineering. Wild soybean is distributed only in eastern Asia, China, the Far East of Russia, the Korean Peninsula, and Japan. Saponins are important secondary metabolites in soybeans, and comprise of a steroidal or triterpenoidal aglycone that combines with one or more sugar moieties by glycosylic linkage. Triterpenoidal saponins levels are high in soybean seeds (Price et al. 1987). A number of the chemical structures and soyasaponin compositions have been identified and characterized in cultivated soybean and wild soybean (Shiraiwa et al. 1991a, b; Kudou et al. 1992, 1993; Tsukamoto et al. 1993;

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

C. Tsukamoto (✉)
Department of Applied Biological Chemistry, Iwate
University, Morioka, Japan
e-mail: chigen@iwate-u.ac.jp

Krishnamurthy et al. 2014). Soyasaponins are known to belong to three groups: group-A, DDMP, and *Sg-6* (Honda et al. 2009), which are based on the chemical structures of the aglycones and the composition of the sugar chains that combine with the aglycones (Fig. 1). Group-A saponins contain soyasapogenol A (SS-A) as its aglycone with two sugar chains attached to the C-3 and C-22 hydroxyl positions (Shiraiwa et al. 1991b). DDMP saponins have soyasapogenol B (SS-B) as their aglycone with a DDMP molecular group (2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one) at the C-22 position and one sugar chain at the C-3 position (Shiraiwa et al. 1991a). DDMP saponins are easily decomposed by heat or by the lipoxygenase (radical) reaction into group B saponins and group E saponins, which have soyasapogenols B and E as their respective aglycones (Kudou et al. 1992, 1993). The group B and E saponins also easily decompose, even during high performance liquid chromatography (HPLC). Recently, third group saponins, *Sg-6* saponins, have been reported in wild soybeans in Japan and South Korea (Honda et al. 2009; Krishnamurthy et al. 2013). The *Sg-6* saponins include at least three components:

H, I, and J. They have soyasapogenols H, I, or J as their aglycones and one sugar chain at the C-3 position of the aglycones (Takahashi et al. 2013).

Soybean saponins involve in health and taste, and these effects are based on the chemical structures of their aglycones and their sugar chain composition (Tsukamoto and Yoshiki 2006). DDMP saponins and their decomposition products have many health benefits (Tsukamoto and Sugano 2006). Group-A saponins have recently been reported to have anti-obesity effects (Yang et al. 2015) and to prevent memory impairment (Hong et al. 2014). However, group-A saponins also have bitter and astringent tastes due to the acetylation of the terminal sugar in the sugar chain attached at the C-22 position of their SS-A aglycone (Okubo et al. 1992). Soybean taste breeding has therefore attempted to genetically eliminate or reduce the levels of group-A saponins in soybean seeds (Kato et al. 2007).

Soyasaponin composition and phenotypes have been investigated in Japan and South Korea. Several variants (A0- α g, A0- α a, and null Group-A) and a couple of rare types (AcAf and AuAeBc) have been

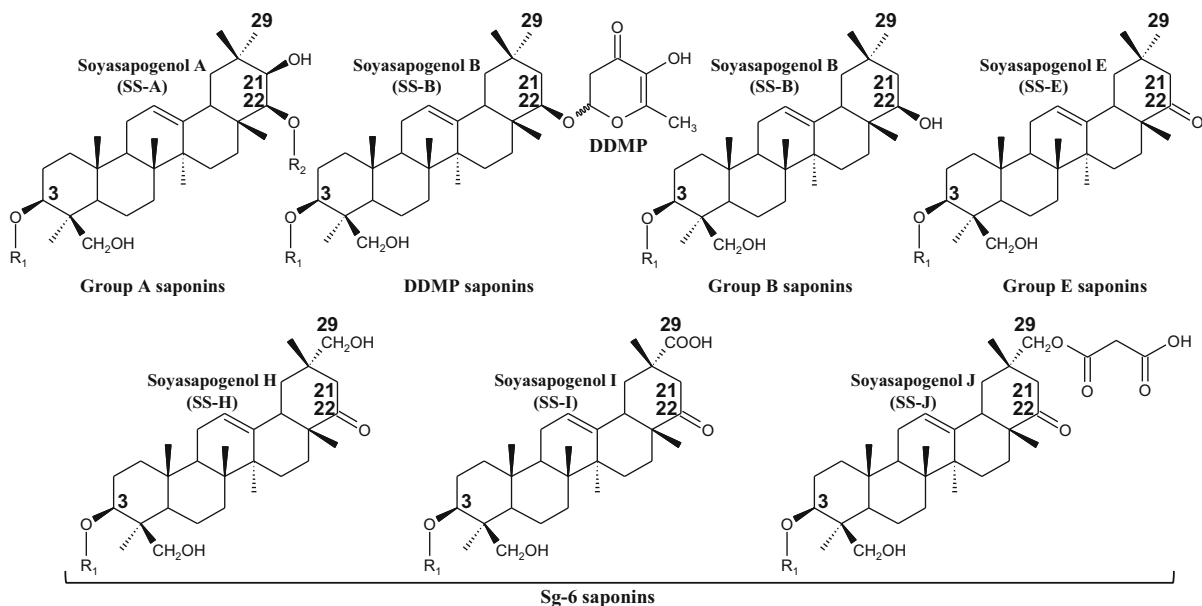


Fig. 1 The category groups and chemical structures of soybean saponins. There are several categories: group-A, DDMP, group-B, group-E, and *Sg-6* saponin (H, I, and J components). Group-A has two sugar chains, R₁ and R₂ at the C-3 and C-22 positions,

respectively, of the aglycone. The other groups have one sugar chain (R₁) at the C-3 position of the aglycone. Group-E and *Sg-6* saponins have a common chemical ketone group

detected in wild and cultivated soybean (Tsukamoto et al. 1993; Krishnamurthy et al. 2013). Some studies have shown that different geographical populations of wild soybean genetically differ in East Asia (Yu and Kiang 1993; Wang and Takahata 2007). Soybean originates in China and its progenitor, wild soybean, is widely distributed across China, except in the westernmost Xinjiang and Qinghai, and the southernmost Hainan region. Wild soybean is an important genetic resource for cultivated soybeans and has been extensively collected in China. There are about 9000 accessions conserved in the genebank. Therefore, Chinese wild soybeans are expected to mine useful genes or variants for soybean breeding.

In this study, our objective was to investigate saponin composition polymorphisms in Chinese wild soybean accessions and to understand the level of genetic variation of saponins for the future soybean breeding programs that aim to improve soybean nutritional levels and quality.

Materials and methods

Plant materials and chemicals

A total of 3795 accession samples from the Chinese wild soybean collection in the China genebank were analyzed for their saponin composition. The samples were randomly taken covering all the 28 administrative regions where wild soybean is distributed, and the sample number from each region relatively depended upon the number of wild soybean accessions preserved in the genebank. Several seeds of each accession were propagated into single plants (harvested as single plants). Then one random plant was used to analyze saponin composition. Two Japanese soybean varieties, Shiroseennari (Aa type) and Suzuyutaka (Ab type), were used as the standards. Commercial soyasapogenols A (SS-A) and B (SS-B) are the aglycones found in group A and DDMP saponins, respectively. The SS-A and SS-B reagents were purchased from ChromaDex Co. (Irvine, CA, USA).

Saponin extraction and composition analysis

Saponin extraction, TLC-analysis, LC-MS (LC-MS/MS) analysis were carried out according to our previous work (Takahashi et al. 2016).

Results

Identification of saponin composition by TLC analysis

The TLC analysis completely recognized 23 saponin components from 3795 wild soybean samples (Fig. 2; Table 1). These included known group-A saponins (purple) Aa, Au, Ae, Ax, Ab, Af, and Ad; DDMP saponins (brown) α g, β g, and β a; group-B saponin (brown) Ba; group-E saponin (blue) Bd; and Sg-6 group saponins (blue) H- α g, I- α g, and J- α g.

Four new components were detected in Chinese wild soybeans, which included two reported components (Takahashi et al. 2016): a precursor A- α g of group-A saponins and its derivative KA- α g (Fig. 2 lane 8, Fig. 3f) and other two were a HAb- α g (X_1) and Sg-6 group saponin K- α g (X_2).

New saponin component HAb- α g (X_1)

A new component X_1 was identified in accession No. 4922 (Fig. 2 lane 9, Fig. 3g). The ESI(+)-MS showed that X_1 had a molecular ion mass $[M + H]^+$ at m/z 1453.6467 (Fig. 4b), whose formula was deduced to be $C_{67}H_{104}O_{34}$ (calculated from $C_{67}H_{105}O_{34}^+$, 1453.6482). This was 16 larger than saponin Ab (m/z 1437.6536, $C_{67}H_{105}O_{33}^+$) (Fig. 4A), which implied that X_1 had an additional oxygen compared to saponin Ab. The prominent ion peak for X_1 appeared at m/z 991.2 for $[M\text{-tetraacetylGlc}(330)\text{-Ara}(132) + H(1)]^+$ (Fig. 4B), which was similar to that of the sugar chain at the C-22 position [22-*O*-(tetraacetylGlc-Ara)] of saponin Ab (Fig. 4A, m/z 975.2). Furthermore, other prominent peaks appeared at m/z 1291.2 for $[M\text{-Glc}(162) + H]^+$, m/z 829.3 for $[M\text{-tetraacetylGlc-Ara-Glc} + H]^+$, m/z 649.4 for $[M\text{-tetraacetylGlc-Ara-Glc-Gal}(\text{galactose})(162)\text{-H}_2\text{O}(18) + H]^+$, and m/z 473.4 for $[M\text{-tetraacetylGlc-Ara-Glc-Gal-GlcUA}(\text{glucuronic acid})(176)\text{-H}_2\text{O} + H]^+$ ($=[Aglycone(491)\text{-H}_2\text{O} + H]^+$) (Fig. 4B). These fragment patterns were also similar to those for the sugar chain at the C-3 position [3-*O*-(Glc-Gal-GlcUA)] of saponin Ab (m/z 1275.1, 813.2, 633.2, and 457.4, respectively) (Fig. 4A). The MS/MS results suggested that the structure of X_1 is similar to that of saponin Ab except one additional oxygen at the aglycone, whose molecular ion mass was 491 ($C_{30}H_{51}O_5^+$). Soybean saponins are active at the C-29 position of the aglycone and can easily insert chemical groups into the C-29

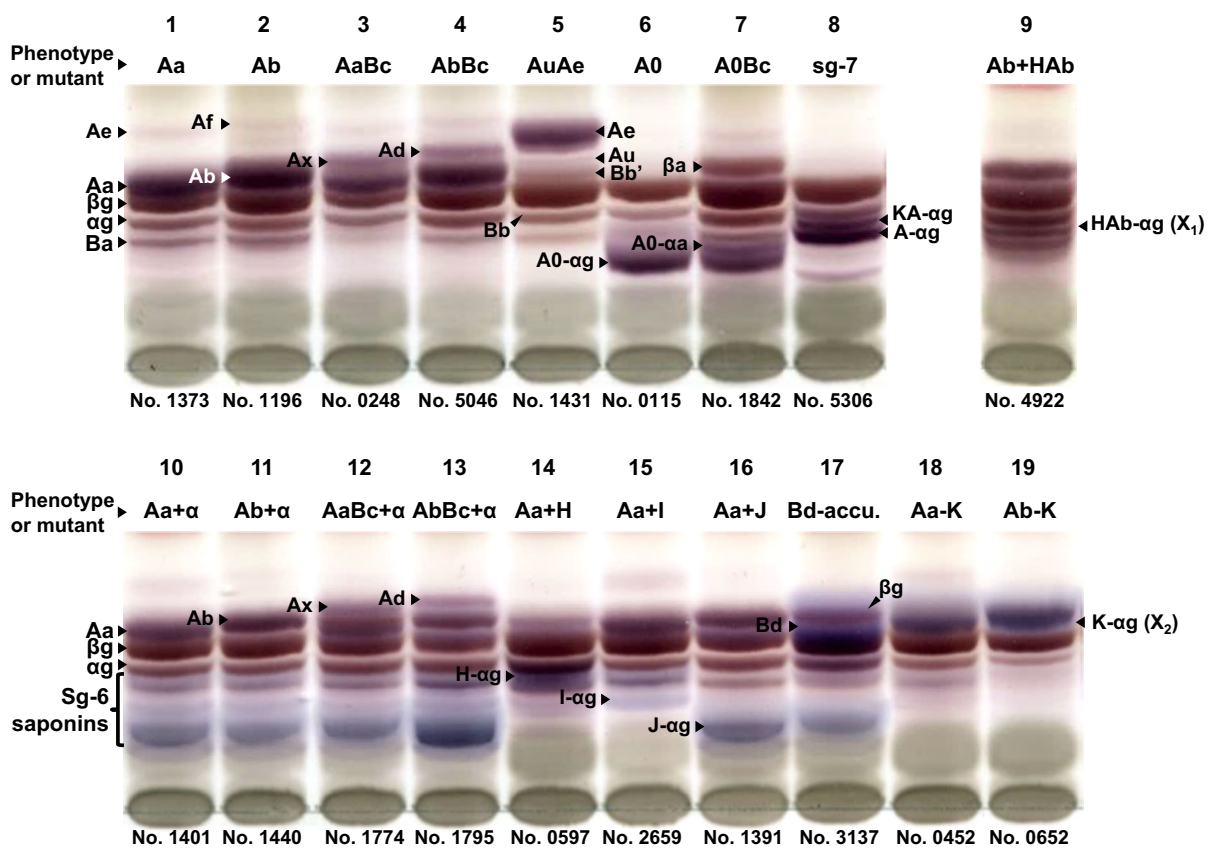


Fig. 2 TLC patterns for saponins in Chinese wild soybeans. Twenty-three components were distinguished, which included four new components: an A- α g (C-3 glycosylated SS-A), and its derivative KA- α g (lane 8), a HAb- α g (lane 9) and a K- α g in Sg-6 saponins (lanes 18 and 19). Lane 5 shows Ae-accumulation

(AuAe) caused by *sg-3/sg-3*; lanes 6 and 7 show variant A0- α g (or A0- α a), which is caused by *sg-1⁰/sg-1⁰*; lane 8 illustrates the null-group-A mutant (*sg-7/sg-7*) and lane 17 shows Bd accumulation

position like Sg-6 saponins (Fig. 1) (Takahashi et al. 2013). Therefore, this additional oxygen was estimated to lie in the C-29 position like soyasapogenol H (Fig. 1). Based on these chemical profiles, the X₁ and its aglycone were named to HAb- α g and soyasapogenol HA, respectively (Fig. 5A).

A new recruit member K- α g component of Sg-6 saponin

Sg-6 saponins are colored blue in the TLC analysis (Honda et al. 2009). This characteristic may be due to the ketone (C=O) group at the C-22 position (Takahashi et al. 2013; Krishnamurthy et al. 2014). A component X₂ was newly identified. It had a mobility between the Aa (Ab) and β g components and appeared blue in color (Fig. 2, lanes 18 and 19). The UV

chromatogram revealed that a unique peak X₂ clearly appeared in No. 0652 (Fig. 6f). The X₂ had a molecular ion [M + H]⁺ peak at *m/z* 1015.51117 (Fig. 4C), whose molecular formula was estimated to C₅₀H₇₈O₂₁ (calculated from C₅₀H₇₉O₂₁⁺, 1015.510841). The C-3 sugar chain was identified as Glc-Gal-GlcUA by the fragment ions at *m/z* 853.1 for [M-Glc(162) + H]⁺, *m/z* 673.0 for [M-Glc(162)-Gal(162)-H₂O(18) + H]⁺ and *m/z* 497.3 for [M-Glc(162)-Gal(162)-GlcUA(176)-H₂O(18) + H]⁺ in the MS/MS spectrum (Fig. 4B). The sugar chain composition at the C-3 position was identical with that of group-A saponins Aa and Ab, and DDMP saponin α g. The MS/MS fragmentations suggested that X₂ had an unknown aglycone with a molecular ion mass of 515 (C₃₂H₅₁O₅⁺), which was 42 larger than that of soyasapogenol H (473, C₃₀H₄₉O₄⁺). This implied that the

Table 1 Categories and frequencies (%) of 23 soyasaponin components identified by TLC analysis in Chinese wild soybean collection

DDMP saponins (and their decompositions)																			
Group-A saponins						Group-B						Group-E							
Aa-series		Ab-series		A0-series		A-series		DDMP		Component		Freq.%		Component		Freq.%			
Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%	Component	Freq.%		
Aa	78.8	Ab	20.9	A0- α g	0.21	A- α g	0.03	α g	99.9	Ba	99.9	Bd	0.03						
Au	0.1	Ac		A0- β g		A- β g		β g	100	Bb	0.1	Be							
Ae	78.9	Af	20.9	A0- γ g		A- γ g		γ g		Bb'	0.1	Be'							
Ax	25.0	Ad	5.7	A0- α a	0.08	A- α a		α a		Bx		Bf							
Ay		Az		A0- β a		A- β a		β a	30.8	Bc		Bg							
Ag		Ah		A0- γ a		A- γ a		γ a		Bc'		Bg'							
Sg-6 saponins																			
H-series						J-series						K-series							
Component		Freq.%		Component		Freq.%		Component		Freq.%		Component		Freq.%		Component		Freq.%	
H- α g	0.6	I- α g	11.4	J- α g	10.9	K- α g	0.5	KA- α g	0.03	HAb- α g	0.03								
H- β g		I- β g		J- β g		K- β g		KA- β g		HAb- β g									
H- γ g		I- γ g		J- γ g		K- γ g		KA- γ g		HAb- γ g									
H- α a		I- α a		J- α a		K- α a		KA- α a		HAb- α a									
H- β a		I- β a		J- β a		K- β a		KA- β a		HAb- β a									
H- γ a		I- γ a		J- γ a		K- γ a		KA- γ a		HAb- γ a									

Except for the 23 TLC-detected components that showed frequencies (%), other components have not yet been detected by TLC analysis

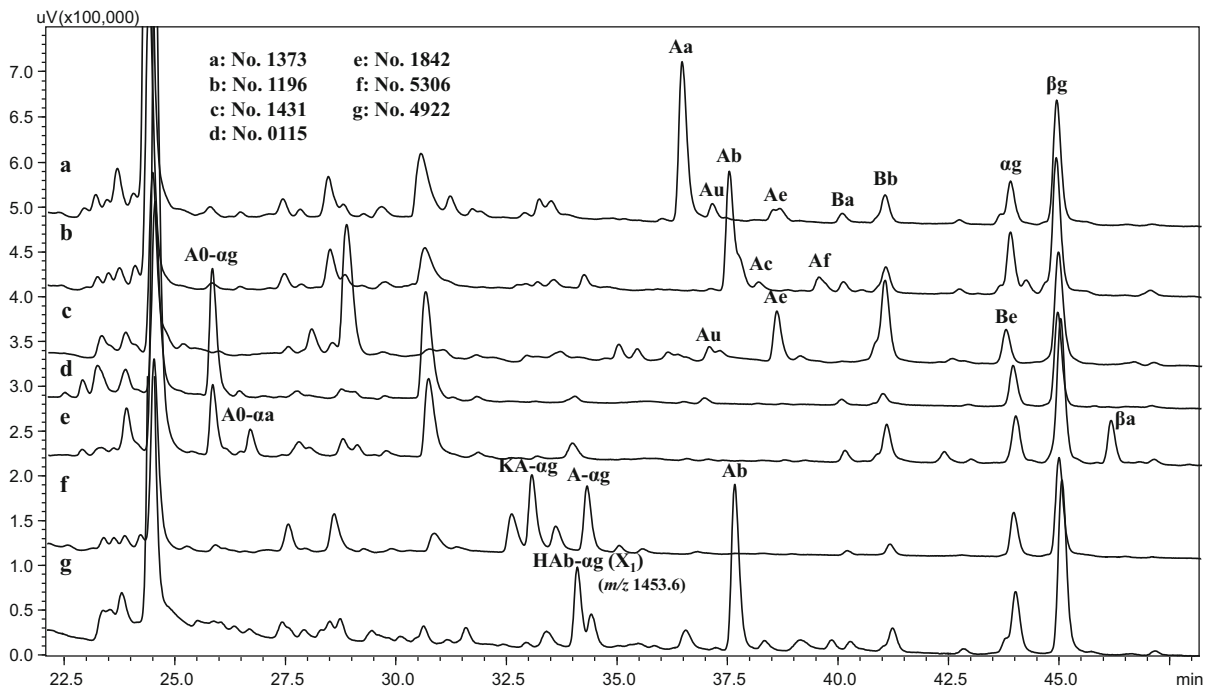


Fig. 3 LC–MS analysis of saponins from several representative accessions. *a* Aa type (No. 1373), *b* Ab (No. 1196), *c* AuAe type (No. 1431), *d* A0 type (No. 0115), *e* A0Bc type (No. 1842),

f null-group-A (No. 5306), and *g* HAB- α g (X_1) component (No. 4922). The results were obtained from an analysis by a PDA detector at UV 205 nm

aglycone of X_2 could have an additional acetyl group ($-\text{COCH}_3$) in the C-29 position of soyasapogenol H against the saponin HAB- α g (X_1). The X_2 and its aglycone were named to K- α g and soyasapogenol K, respectively (Fig. 5B), which might belong to *Sg-6* saponin. Eighteen K- α g accessions were identified (Table 1).

Separate three single components of *Sg-6* saponin

Sg-6 saponin has three components: H- α g, I- α g, and J- α g (Honda et al. 2009). Krishnamurthy et al. (2013) reported that the three components behaved together as a congener controlled by a dominant allele *Sg-6* in *Glycine soja*. However, we observed that H- α g, I- α g, and J- α g could exist singly and separately in different accessions (Fig. 2, lanes 14, 15, and 16). The LC–MS data showed that accession No. 1401 simultaneously carried H- α g, I- α g, and J- α g (Fig. 6a). However, these saponins could occur in separate individuals (Fig. 6b, c, d). In the Chinese accessions, about 13.1% contained *Sg-6* saponins, and the frequencies for H- α g, I- α g, and J- α g were about 0.6, 11.4, and 10.9% in the accessions, respectively (Table 1).

A0 variant of group-A saponin

Eight samples were found to carry an A0- α g component or one additional A0- α a component (Fig. 2, lanes 6 and 7). Five accessions (A0 type) carried A0- α g (Fig. 2, lane 6), and other three accessions (A0Bc type) carried both A0- α g and A0- α a (Fig. 2, lane 7). Eight of the A0 variant accessions were collected from northeastern China and one was from Henan Province, central China.

Rare Ae-accumulation (AuAe) type

Usually Ae component is a minor one in some plants. However, rare types with a high Ae content were identified by TLC in four accessions (Fig. 2, lane 5). Sample No. 1431 showed that there was no saponin Aa, but Ae was accumulated as a major saponin. The Ae-accumulation (AuAe) showed inheritability in later offspring generations (data not shown). LC–MS detection showed that the saponin Ae content in No. 1431 (20.2 mg/g) was 14 times higher than in the normal Aa accession (No. 1373, 1.45 mg/g). All four accessions containing the rare AuAe type originated from northeastern China.

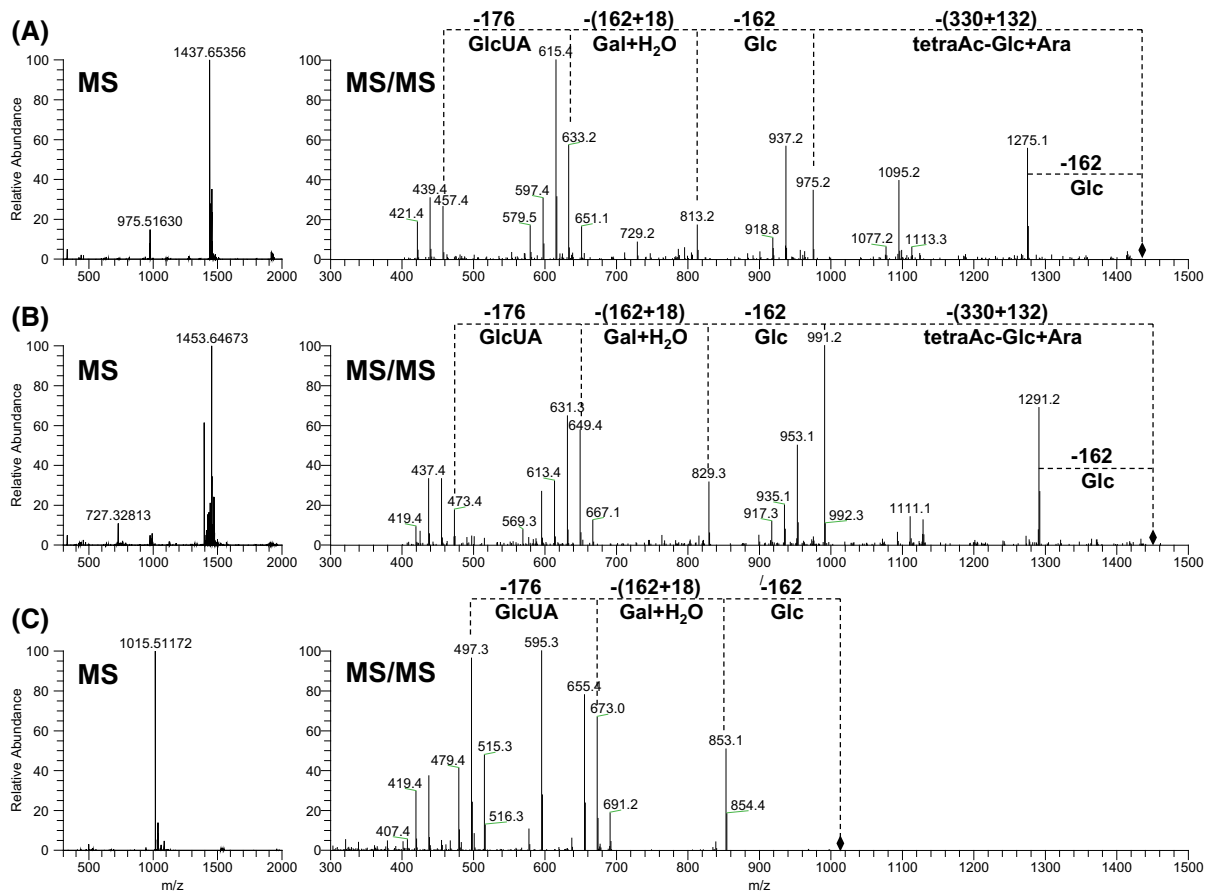


Fig. 4 MS and MS/MS spectra of the Ab (A), HAb- α g (B), and K- α g (C) components. MS/MS fragment patterns of the molecular ion peaks for the saponins showed that the three saponin components had the same three sugar moiety sequence at the C-3 position, which were in the order Glc-Gal-GlcUA, and no sugar chain at the C-22 position in the K- α g (X_2) component.

Rare Bd-accumulation type

Commonly, the Bd component is invisible in TLC because it is present at very low levels, but HPLC can identify it. A rare Bd-accumulating accession was identified by TLC (Fig. 2, lane 17). The LC-MS analysis determined a main peak that appeared at 42.8 min in accession No. 3137, which was the Bd component of group E saponins (Fig. 6e). This accession also had a J- α g component in the TLC results (Fig. 2, lane 17) and had a small amount of Aa component in the HPLC results (Fig. 6e). The Bd content in this accession was very high with 17.3 mg/1 g hypocotyl, and was 37.2 times higher than that in Aa-type accession No. 1373 (0.47 mg/1 g hypocotyl).

This was different to the Ab and HAb- α g (X_1) components, which had a two sugar moiety, Glc-Ara, at the C-22 position. K- α g (X_2) had a molecular ion $[M + H]^+$ peak at m/z 1015.5 and contained a new aglycone, tentatively named as soyasapogenol K (m/z 515.4). It is suggested that K- α g (X_2) has an acetylation at the C-29 position, called a 29-acetyl H- α g

This accession was propagated for three generations and each generation still accumulated Bd according to the HPLC analysis (data not shown). Therefore, the results suggested that the Bd-accumulating accession was a genetic variant with higher Bd levels.

Discussion

Abundant variability of saponins in the Chinese wild soybeans

Wild soybean saponins theoretically should have 54 components according to the known gene-coding (Shiraiwa et al. 1991a, b; Tsukamoto et al. 1993;

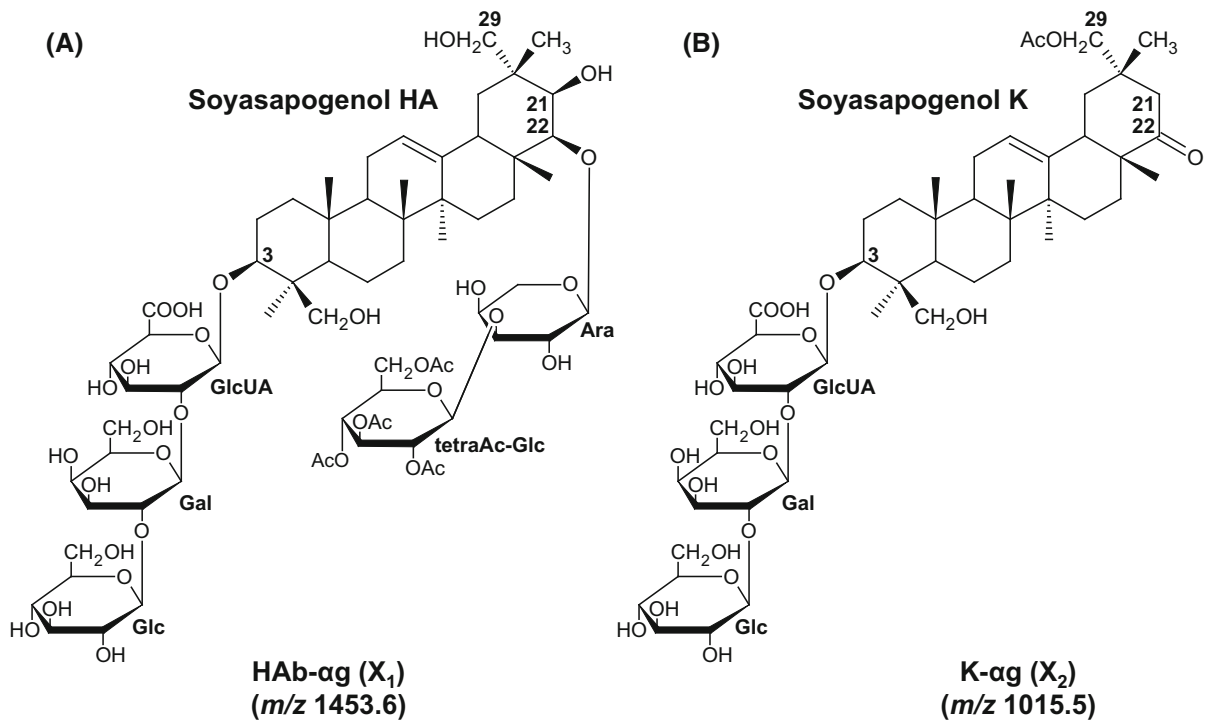


Fig. 5 Suggested chemical structures for X₁ and X₂, named HAB- α g and K- α g, respectively. **A** HAB- α g (X₁) has a new aglycone called soyasapogenol HA. It has an –OH group at the

C-29 position, i.e. a 29-hydroxyl SS-A. **B** K- α g (X₂) has a soyasapogenol K, which may be derived from SS-H by acetylation of the C-29 position, i.e. a 29-acetyl SS-H

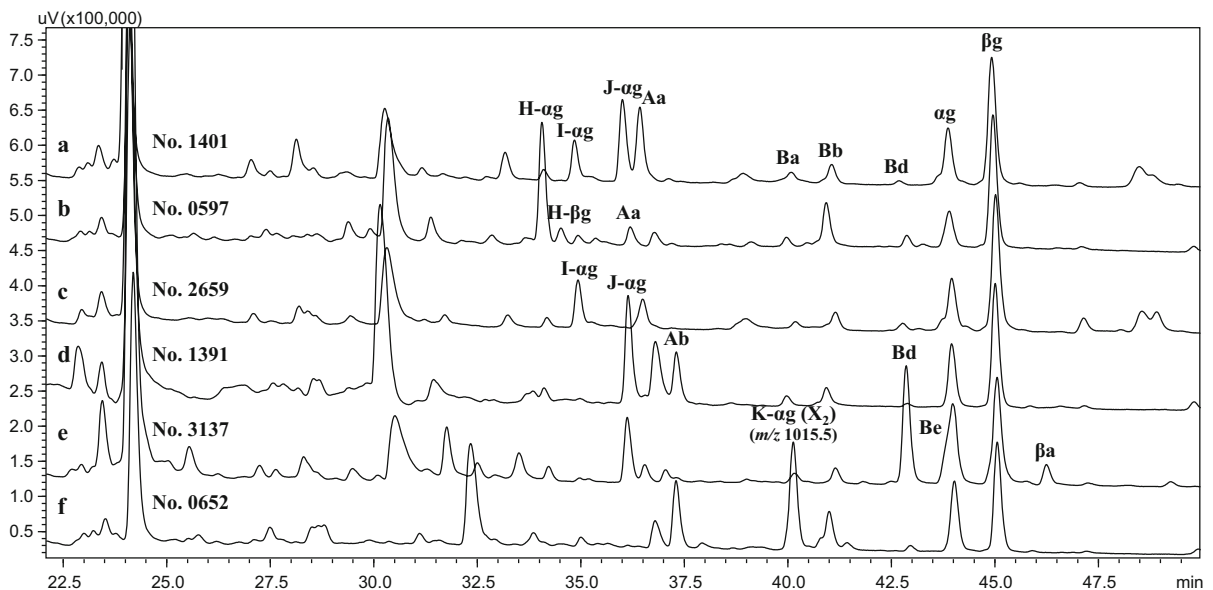


Fig. 6 LC–MS analysis of *Sg-6* saponins. **a** I- α g and J- α g (No. 1401), **b** H- α g (No. 0597), **c** I- α g (No. 2659), **d** J- α g (No. 1391), **e** group E Bd-accumulation (No. 3173), and **f** new component

X₂ (No. 0652). The results were obtained from an analysis by a PDA detector at UV 205 nm

Honda et al. 2009; Krishnamurthy et al. 2014), but actually there could be more than the number above if including the syntheses by genetic coding and by non-genetic coding for incomplete acetylation in the hydroxyl groups of the terminal sugar at C-22 position. However, according to our investigation, there are at least 78 gene-coded components, theoretically (Table 1). Nevertheless, the TLC analysis only distinguished 23 components, including four new saponin components in Chinese wild soybean (Table 1). Most of the saponins could not be identified because a large number of the saponin components completely and partly overlap in mobility, or have less content, or do not appear without their corresponding genes or genotypes.

In the South Korean accessions, 14 components and two rare phenotypes (AuAeBc and variant A0) were tested for wild soybeans by TLC analysis (Krishnamurthy et al. 2013). In this investigation, we identified twenty-two saponin components, eight A0-type accessions, four rare AuAe type accessions and one rare Bd-accumulation variant type, and one novel mutant (accession No. 5306) without the known group-A saponins which was caused by an invalid arabinosyl-transferase. This meant that the arabinose failed to bind to the C-22 hydroxyl group (Takahashi et al. 2016). We assigned a symbol *Sg-7* to the locus governing arabinose bonding to the C-22 hydroxyl group in accession No. 5306. These newly identified components, rare phenotypes, novel mutant(s), and variant types need to be cross-analysed to identify their genetic characteristics.

The significance of finding a rare, high Bd-accumulating accession

Group E saponins, such as Bd and Be, which have soyasapogenol E as the aglycone (Fig. 1), are thought to be the decomposed products of DDMP saponins during extracting. Generally, Bd appears as a minor component in the LC–MS analysis (Fig. 6). However, in this investigation, accession No. 3137 accumulated a high concentration of the group E (Bd) component (Fig. 6e). The Bd-accumulation in this accession was inheritable through three generations (data not shown). The important implication of this discovery is that not all the Bd in wild soybeans is the decomposed products of DDMP saponins. Our results suggest that the Bd in group E saponin can be

genetically biosynthesized in the hypocotyls of wild soybean.

Bd and *Sg-6* saponins all have a ketone group at the C-22 position of their aglycones (Fig. 1). Based on the chemical structures of Bd and *Sg-6* saponins, we suggest that Bd in developing hypocotyls might be associated with the biosynthesis of *Sg-6* saponins, i.e. Bd is the initial substrate. It may also be involved in the biosynthesis of the H- α g component of *Sg-6* saponins, where Bd is hydroxylated at the C-29 position into H- α g, followed by carboxylation into I- α g or malonylation into J- α g. Therefore, *Sg-6* saponins are derived from the Bd component, which strongly suggests that the Bd component is the precursor of *Sg-6* saponins. The reason of why Bd could be accumulated in this variant accession No. 3137 and did not been completely consumed are not clear, but we conjectured that the Bd gene in this accession could be over-expressed. This needs to be confirmed in the future.

The trade-off nexus between the biosynthesis of *Sg-6* saponins and group-A saponins

Soyasaponins consist of three major groups, DDMP, group-A, and *Sg-6*. The *Sg-6* saponins were first reported a few years ago (Honda et al. 2009). Usually, group-A Aa or Ab components in the accessions without the *Sg-6* group have higher saponin contents, as illustrated in Fig. 3a, g. However, we found that when *Sg-6* saponins, including Bd, appeared, group-A saponins would be more or less decrease, but DDMP did not change in content, as demonstrated by the HPLC analysis (Fig. 6b, c, d, e), which implies that their biosynthesis share the same precursor substrate at a certain step in their biosynthetic pathways. This substrate compound, which is a common component of group-A and *Sg-6* saponin biosynthesis, might be the group-A precursor component A- α g (C-3 glycosylated SS-A). A proportion of the A- α g was used to synthesize the Bd component through a dehydrase mediated dehydration reaction between two neighboring oxyhydrils at the C-21 and C-22 positions that led to a ketone group at the C-22 position. Therefore, when the *Sg-6* saponins are over-synthesized, group-A saponin levels decrease so as to become dim at the TLC analysis, and it is difficult for TLC analysis to identify them (Fig. 2, lanes 14, 15, 16, 17, 19). Generally, there appears to be a trade-off between

group-A and *Sg-6* saponin contents. Some genetic experiments are being made to confirm this relationship.

Novel saponin phenotypes in East Asian wild soybean populations

Two novel types, A0Bc and A0Bc-S, were detected in the wild soybean accessions from Japan (Tsukamoto et al. 1993, 1998) and South Korea (Krishnamurthy et al. 2013). An AuAeBc type was also observed in South Korea (Krishnamurthy et al. 2013), which is ascribed to the GmSGT-3 gene (Shibuya et al. 2010) in the plants with alleles *sg-3/Sg-4*. The A0Bc-S type detected in Japan and South Korea, containing the *Sg-4* gene (Takada et al. 2012), is caused by a recessive allele, *sg-5*, which means that there are no group-A saponins in the seeds (Tsukamoto et al. 1998). The A0Bc type, carrying the *Sg-4* gene, is controlled by recessive allele *sg-1⁰* at locus *Sg-1* for the terminal sugar (Xyl or Glc) in the sugar chain at the C-22 position (Kikuchi et al. 1999). Sayama et al. (2012) have sequenced the *Sg-1^a*, *Sg-1^b*, and *sg-1⁰* alleles at the *Sg-1* locus, and revealed that the recessive *sg-1⁰* allele that leads to the absence of the terminal Xyl or Glc sugars from the sugar chain at the C-22 position, and that *sg-1⁰* could have mutated from the *Sg-1^a* and *Sg-1^b* alleles in Japanese soybeans. The A0Bc type (with the *Sg-4* gene) in Japanese wild soybean was only controlled by *sg-1⁰*, which is an *Sg-1^a* mutant (Sayama et al. 2012). The A0Bc type detected in South Korea has also been reported to be caused by the mutant *sg-1⁰* replacing *Sg-1^a* (Krishnamurthy et al. 2015), but differed from the Japanese A0Bc type by a point mutation within the *sg-1⁰* gene.

In this investigation, three types of AuAe, A0, and A0Bc were identified in Chinese wild soybeans, including five A0 and three A0Bc accessions, and four AuAe accessions with recessive *sg-3* and *sg-4* genes. All the AuAe accessions were collected from northeast China, which implied that the recessive *sg-3* gene was distributed in northeastern China as well as neighboring regions. However, we do not know whether these A0 Chinese wild accessions were caused by the same point mutations within allele *sg-1⁰* as detected in South Korea (Krishnamurthy et al. 2015) and Japan (Sayama et al. 2012). In the eight A0 Chinese wild soybean accessions, one originated from

central China and seven from the northeast, and they need to be analysed for their mutational mechanisms.

The potential functional genes for soybean flavor breeding

The bitter and astringent tastes in soybean foods are to be caused by acetylation of the hydroxyl position of the C-22 terminal sugar moiety in group-A saponins (Okubo et al. 1992). In order to obtain better tasting soybean food, the acetylation of the hydroxyl position in group-A saponins needs to be genetically eliminated. One effective method is to introduce potentially useful genes, including genes that are unable to acetylate the hydroxyl position and to bind Xyl or Glc to the Ara at C-22 position. It may be possible to biosynthetically change the saponin compositions, such as strongly biosynthesizing *Sg-6* saponins to reduce group-A content. At present, genes that do not acetylate Xyl or Glc at the C-22 position have not been identified. However, many mutant germplasms or genes can be utilized to eliminate or decrease group-A contents. The *sg-5* gene has been transferred into variant ‘Tohoku 152’, that no longer produces group-A saponins (Sakai et al. 2002). The *sg-7* gene was involved in biosynthesizing incomplete group-A (A- α g) without the C-22 sugar moiety, which meant that no acetylation occurred. This suggests that it can be used to develop good-tasting soybeans. The *sg-1⁰* allele has no C-22 terminal sugar moiety. Therefore, it can not be acetylated, and it has been successfully transferred into variety ‘Kinusayaka’ (Kato et al. 2007). These eight Chinese wild soybean accessions are valuable germplasm resources for soybean taste breeding. The synthesis of the *Sg-6* group, including Bd-accumulating accessions, would decrease group-A content, which means that they could have great potential in soybean breeding. However, as a breeding parent, the accessions containing *Sg-6* saponins or Bd saponin must have a very low group-A saponin content, such as wild soybeans accessions No. 0597 and No. 3137, which have naturally very low group-A (Aa or Ab) saponin contents (Figs. 6b, e) and they should be useful breeding parents.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 31571697), the ‘‘Sci & Tech Innovation Program of Chinese Academy of Agricultural Sciences’’, and the ‘‘National Basic Research

Program” from the Ministry of Science and Technology of the People’s Republic of China (Item No. 2011FY110200).

References

- Honda N, Tsukamoto C, Maehara Y, Tayama I, Kitamura K, Singh RJ, Chung GH (2009) Saponin composition of *Glycine soja* (Sieb. et Zucc.) mutants having new soyasapogenol aglycones and geographical distribution in South Korea. Abstract of Proceedings of world soybean research conference VIII, Beijing, China, pp 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
- 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
- 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, Honda N, Kikuchi A, Lee JD, Yang SH, Chung G (2013) Saponin polymorphism in the Korean wild soybean (*Glycine soja* Sieb. et Zucc.). *Plant Breed* 132:121–126
- Krishnamurthy P, Tsukamoto C, Singh RJ, Lee JD, Kim HS, Yang SH, Chung G (2014) The Sg-6 saponins, new components in wild soybean (*Glycine soja* Sieb. et Zucc.): polymorphism, geographical distribution and inheritance. *Euphytica* 198:413–424
- Krishnamurthy P, Lee JD, Ha BK, Chae JH, Song JT, Tsukamoto C, Singh RJ, Chung G (2015) Genetic characterization of group A acetylsaponin-deficient mutants from wild soybean (*Glycine soja* Sieb. et Zucc.). *Plant Breed* 321:316–321
- Kudou S, Tonomura M, Tsukamoto C, Shimoyamada M, Uchida T, Okubo K (1992) Isolation and structural elucidation of the major genuine soybean saponin. *Biosci Biotech Biochem* 56:142–143
- Kudou S, Tonomura M, Tsukamoto 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 Biotech Biochem* 57:546–550
- Okubo K, Iizima 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
- Price KR, Johnson IT, Fenwick GR (1987) The chemistry and biological significance of saponins in foods and feedingstuffs. *Crit Rev Food Sci Nutr* 26:27–135
- Sakai T, Kikuchi A, Takada Y, Kono Y, Kunishi I, Tezuka M, Asao H, Shimada S (2002) Characteristics of soybean lines “Tohoku 151” and Tohoku 152” which are lacking all lipoxygenases and modified saponin composition in seeds. *Tohoku Agric Res* 55:59–60
- 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, Ishimono 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* 584:2258–2264
- Shiraiwa M, Harada K, Okubo K (1991a) Composition and structure of “group B saponin” in soybean seed. *Agric Biol Chem* 55:911–917
- Shiraiwa M, Kudou S, Shimoyamada M, Harada K, Okubo K (1991b) Composition and structure of “group A saponin” in soybean seed. *Agric Biol Chem* 55:315–322
- 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* 61:646–652
- Takahashi Y, Kon T, Muraoka H, Ishimoto M, Tsukamoto C (2013) The dominant *Sg-6* synthesizes saponins with a ketone function at oleanane aglycone C-22 position in soybean [*Glycine max* (L.) Merr.]. In: 11th International meeting on biosynthesis, function and biotechnology of isoprenoids in terrestrial and marine organisms (TERPNET), Kolyvari, Crete, Greece, p 197
- Takahashi Y, Li XH, Tsukamoto C, Wang KJ (2016) Identification of a novel variant lacking group-A soyasaponin in a Chinese wild soybean (*Glycine soja* Sieb. et Zucc.): implications for breeding significance. *Plant Breed* 135(5):607–613
- 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. *Phytochemistry* 34:1351–1356
- Tsukamoto C, Kikuchi A, Harada K, Kitamura K, Iwasaki T, Okubo K (1998) Soybean feeds the world. Proceedings (Suppl) of 5th world soybean research conference. In: Chainuvati C, Sarobol N (eds) *Genetic improvement of saponin components in soybean seed*. Kasetsart University press, Bangkok, pp 80–85
- Wang KJ, Takahata Y (2007) A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genet Resour Crop Evol* 54:157–165
- 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 PPAR γ . *Phytother Res* 29:281–287
- Yu H, Kiang YT (1993) Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). *Euphytica* 68:213–222