The Plant Cell, Vol. 24: 2123–2138, May 2012, www.plantcell.org

The *Sg-1* Glycosyltransferase Locus Regulates Structural Diversity of Triterpenoid Saponins of Soybean [W][OA]

Takashi Sayama,[a,1] Eiichiro Ono,[b,1] Kyoko Takagi,[a] Yoshitake Takada,[c] Manabu Horikawa,[d] Yumi Nakamoto,[e] Aya Hirose,[e] Hiroko Sasama,[a] Mihoko Ohashi,[e] Hisakazu Hasegawa,[f] Teruhiko Terakawa,[f] Akio Kikuchi,[g] Shin Kato,[g] Nana Tatsuzaki,[h] Chigen Tsukamoto,[h] and Masao Ishimoto[a,2]

[a] National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan
[b] Institute for Plant Science, Suntory Business Expert Ltd., Shimamoto, Mishima, Osaka 618-8503, Japan
[c] National Agricultural Research Center for Western Region, Zentsuji, Kagawa 765-8508, Japan
[d] Bioorganic Research Institute, Suntory Foundation for Life Sciences, Shimamoto, Mishima, Osaka 618-8503, Japan
[e] National Agricultural Research Center for Hokkaido Region, Toyohira, Sapporo, Hokkaido 062-8555, Japan
[f] Hokko Chemical Industry Company, Atsugi, Kanagawa 243-0023, Japan
[g] National Agricultural Research Center for Tohoku Region, Kariwano, Daisen, Akita 019-2112, Japan
[h] Graduate School of Agriculture, Iwate University, Morioka, Iwate 020-8550, Japan

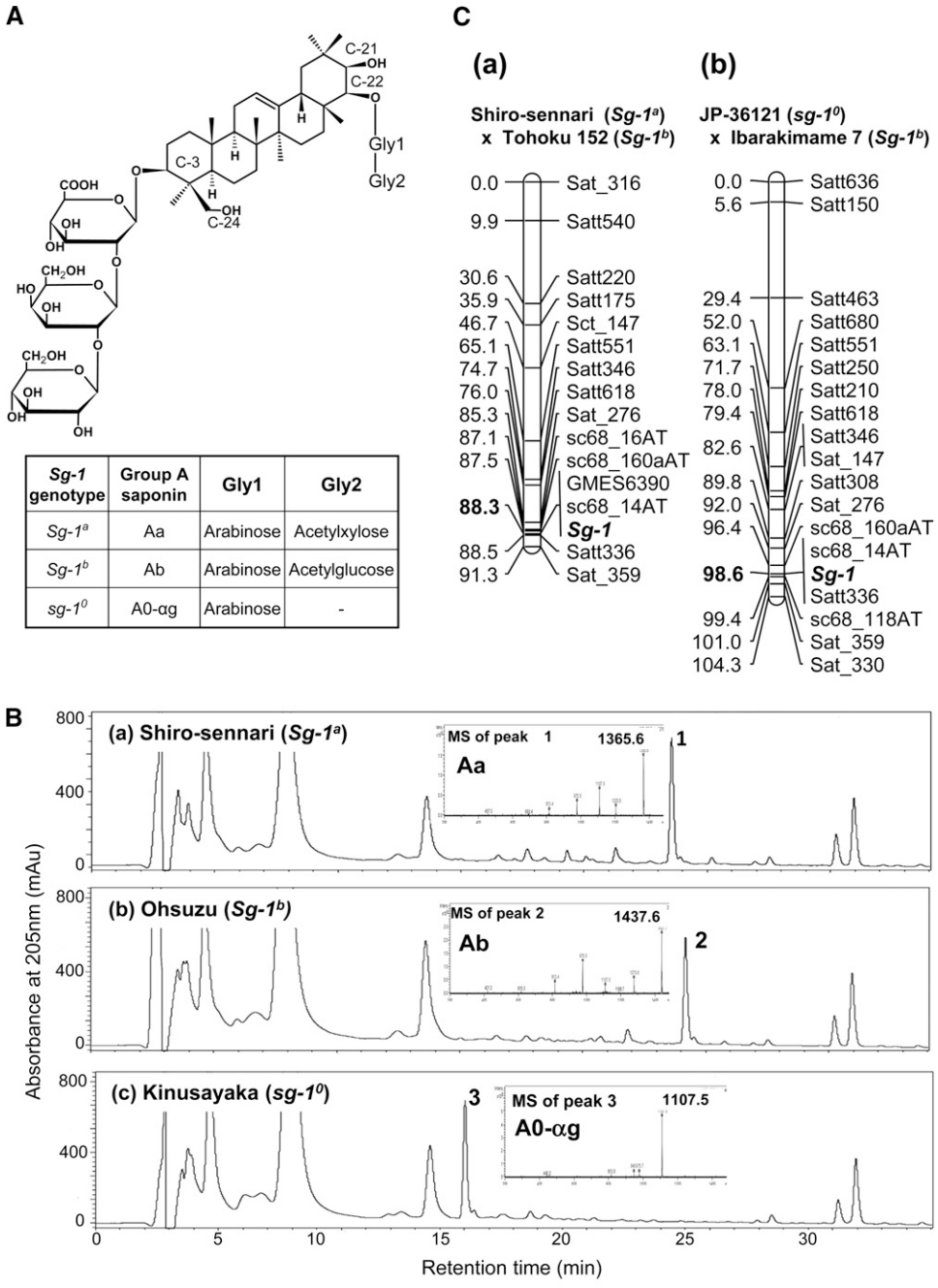
Triterpene saponins are a diverse group of biologically functional products in plants. Saponins usually are glycosylated, which gives rise to a wide diversity of structures and functions. In the group A saponins of soybean (*Glycine max*), differences in the terminal sugar species located on the C-22 sugar chain of an aglycone core, soyasapogenol A, were observed to be under genetic control. Further genetic analyses and mapping revealed that the structural diversity of glycosylation was determined by multiple alleles of a single locus, *Sg-1*, and led to identification of a UDP-sugar–dependent glycosyltransferase gene (Glyma07g38460). Although their sequences are highly similar and both glycosylate the nonacetylated saponin A0-αg, the *Sg-1*^{a} allele encodes the xylosyltransferase UGT73F4, whereas *Sg-1*^{b} encodes the glucosyltransferase UGT73F2. Homology models and site-directed mutagenesis analyses showed that Ser-138 in Sg-1^a and Gly-138 in Sg-1^b proteins are crucial residues for their respective sugar donor specificities. Transgenic complementation tests followed by recombinant enzyme assays in vitro demonstrated that *sg-1*^{0} is a loss-of-function allele of *Sg-1*. Considering that the terminal sugar species in the group A saponins are responsible for the strong bitterness and astringent aftertastes of soybean seeds, our findings herein provide useful tools to improve commercial properties of soybean products.

INTRODUCTION

High-quality proteins and fats are abundant in soybean (*Glycine max*) seeds. Global soybean demand is increasing not only for use as an oilseed crop and feed for livestock and aquaculture, but also as a nutritious food for human consumption and as a feedstock for industrial materials and biofuel (Masuda and Goldsmit, 2009). In addition to soy proteins, the seeds are also rich in physiologically active metabolites, such as isoflavones, lecithin, and saponins, and are used as economic sources of foods that promote and maintain health (Sugano, 2006). Daily intake of processed soybean foods appears to be one of the beneficial factors responsible for the health and longevity of Japanese people (Yamori, 2006).

Triterpene saponins are major components of these secondary metabolites in soybean seeds and exhibit wide structural diversity. Soybean saponins are divided into group A and 2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one (DDMP) saponins. Group A saponins are bisdesmoside-type saponins that have two sugar chains at the C-3 and C-22 position hydroxy groups of the aglycone moiety designated soyasapogenol A (3β,21β,22β,24-tetrahydroxyolean-12-ene) (Figure 1A). DDMP saponins conjugate the DDMP moiety at the C-22 position and a sugar chain at the C-3 position of soyasapogenol B (3β,22β,24-trihydroxyolean-12-ene) as the aglycone. Soyasapogenol B does not have the C-21 position hydroxy group. Degradation of DDMP saponins during processing for food use generates group B and E saponins (Kudou et al., 1992, 1993, 1994). DDMP and group B saponins seem to be ubiquitously distributed with some variation in sugar chain structure at the C-3 position (Tsukamoto et al., 1993; Takada et al., 2012), suggesting that DDMP and group B saponins might have a primary biological function in soybean. In general, saponins are considered to contribute to defense responses in plants because of their antimicrobial, antivirus, and anti-insect activities, although direct evidence in support of this notion is limited (Osbourn, 1996; Papadopoulou et al., 1999; Kuzina et al., 2009). Saponins also are thought to function as antioxidants by scavenging active oxygen species during growth and development (Tsujino et al., 1994; Yoshiki and Okubo, 1995). In legume nodules, superoxide radicals are generated in redox processes, such as respiration in mitochondria and bacteroids,

[1] These authors contributed equally to this work.
[2] Address correspondence to ishimoto@affrc.go.jp.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Masao Ishimoto (ishimoto@affrc.go.jp).
[W]Online version contains Web-only data.
[OA]Open Access articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.111.095174



Sg-1 genotype	Group A saponin	Gly1	Gly2
Sg-1^a	Aa	Arabinose	Acetylxylose
Sg-1^b	Ab	Arabinose	Acetylglucose
sg-1^0	A0-αg	Arabinose	-

Figure 1. Genetic Diversity of Group A Saponin Components in Soybean.

(A) Structural variation of sugar chains at the C-22 position in group A saponins. Group A saponins (Aa, Ab, and A0-αg) are bisdesmoside glycosides having two sugar chains attached at the C-3 and C-22 position of soyasapogenol A, which has four hydroxy groups at the C-3, 21, 22, and 24 positions of the oleane-12-ene molecule. Saponin Aa has acetylxylose and saponin Ab has acetylglucose as the terminal acetylated sugar at the C-22 position. Saponin A0-αg does not have the terminal acetylated sugar. The presence and composition of the terminal acetylated sugar is determined genetically by alleles of the *Sg-1* locus.

(B) Separation and detection of group A saponins in hypocotyls of three different genotypes of the *Sg-1* locus. HPLC chromatograms were obtained at 205 nm and the MS analysis of fragment ion peaks in 80% methanol extracts was prepared from the hypocotyls of three genotypes: **(a)**, 'Shiro-sennari' (*Sg-1*^{a}); **(b)**, 'Ohsuzu '(*Sg-1*^{b}); **(c)**, 'Kinusayaka' (*sg-1*^{0}). Peak numbers indicate group A saponin components: 1, Aa; 2, Ab; 3, A0-αg. Insets in each chromatogram show MS analysis data of the indicated peak components. Peak 1 (saponin Aa, m/z = 1365.6) or peak 2 (saponin Ab, m/z = 1437.6) was

as well as direct reduction of O_2 by nitrogenase, hydrogenase, and ferredoxin in bacteroids (Dalton, 1995). Ectopic expression of the β-amyrin synthase gene (*OXA1*) of *Aster sedifolius* resulted in increased accumulation of triterpene saponins and enhanced root nodulation in barrel medic (*Medicago truncatula*) (Confalonieri et al., 2009). In addition, exogenous treatments with group B saponins stimulate root growth in lettuce (*Lactuca sativa*) and *Arabidopsis thaliana* (Tsurumi and Ishizawa, 1997; Tsurumi et al., 2000).

Soybean saponins are undesirable components of food because they are the main cause of bitterness and astringent aftertastes (Okubo et al., 1992) and of foaming in tofu production. According to previous studies, group B and E saponins have a less bitter astringent aftertaste than the group A saponins (Okubo et al., 1992). Furthermore, DDMP saponins and their derivatives are expected to show beneficial human health effects, such as prevention of dietary hypercholesterolemia (Fenwick et al., 1991; Murata et al., 2005, 2006), suppression of colon cancer cell proliferation (Ellington et al., 2005, 2006), and antiperoxidation of lipids and liver-protecting action by acceleration of secretion of thyroid hormones (Ishii and Tanizawa, 2006). Thus, manipulation of saponin composition and content is required for improvement of soybean quality and function; however, the responsible biosynthetic enzymes are mostly unknown.

Saponin composition in soybean seeds differs among varieties and seed tissues. Hypocotyls, which represent only 2% of the seed weight, contain >30% of the total saponin and all of the group A saponins (Taniyama et al., 1988a; Shimoyamada et al., 1990; Shiraiwa et al., 1991a, 1991b). Removal of seed hypocotyls during food processing is one practical solution to reduce undesirable tastes from soy foods; however, this process also discards many components benefiting health. Group A saponins are subdivided by a difference in the C-22 sugar chains (Tsukamoto et al., 1993). Saponin Aa has a 2,3,4-tri-*O*-acetyl-β-D-xylopyranosyl(1→3)-α-L-arabinopyranosyl sugar chain, and saponin Ab has a 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl (1→3)-α-L-arabinopyranosyl sugar chain. Both have an acetylated sugar at the terminal position (Figure 1A) that causes the undesirable aftertaste (Kitagawa et al., 1988; Taniyama et al., 1988b; Okubo et al., 1992). By screening more than 1000 germplasm stocks, a soybean cultivar and a wild accession were found to accumulate saponin A0-αg, which lacks the acetylated terminal sugar at the C-22 position (Figure 1A) (Kikuchi et al., 1999; Takada et al., 2010). These findings led to breeding of cultivars useful for human food, such as Kinusayaka, which reduces the bitter taste and astringent effects (Kato et al., 2007).

Codominant alleles, *Sg-1*^{a} and *Sg-1*^{b}, at a single locus designated *Sg-1* control accumulation of saponins Aa and Ab, respectively (Shiraiwa et al., 1990; Tsukamoto et al., 1993). The biosynthesis of saponin A0-αg results from a recessive allele, *sg-1*^{0}, at the same locus (Kikuchi et al., 1999; Takada et al., 2010). Here, we describe the identification of multiple alleles of the *Sg-1* locus that are responsible for the structural diversity of the terminal sugar at the C-22 position of group A saponins. Biochemical analyses of the allelic gene products revealed that *Sg-1*^{a} and *Sg-1*^{b} encode UDP-sugar–dependent glycosyltransferases, UGT73F4 and UGT73F2, catalyzing the addition of Xyl and Glc, respectively, to the Ara residue at the C-22 position. Analyses of site-directed mutations in the highly homologous UDP-sugar–dependent glycosyltransferase (UGT) genes identified the critical residues for sugar donor specificity. Furthermore, in vivo transgenic complementation tests followed by in vitro recombinant enzyme assays confirmed that *sg-1*^{0} alleles isolated from two backgrounds (*Sg-1*^{a} and *Sg-1*^{b}) are loss-of-function alleles of *Sg-1*.

RESULTS

Genetic Identification of Multiple Alleles at the *Sg-1* Locus

Soybean hypocotyls mainly accumulate one of the group A saponins (Figure 1A) (see Supplemental Figure 1 online). The acetylated components Aa and Ab were readily distinguished from each other by HPLC and liquid chromatography–mass spectrometry (LC-MS) analysis (Figure 1B). Instead of one of the two acetylated saponins, 'Kinusayaka' and a wild accession 'JP-36121' accumulate a typical saponin, A0-αg, which lacks the acetylated terminal sugar at the C-22 position (Figure 1A) (see Supplemental Table 1 online). The retention time for A0-αg was clearly different from those of Aa and Ab obtained by HPLC analysis (Figure 1B). Segregating progenies of the ST and JI populations (see Methods) were developed from the crosses between soybean genotypes with different group A saponin components (Figure 1C) (see Supplemental Table 1 online). Each F2 seed was divided into cotyledon and hypocotyl, which were used for genotyping of molecular markers and identification of group A saponin components, respectively. Previous linkage analyses had positioned the *Sg-1* locus between simple sequence repeat (SSR) markers Sat_276 and Sat_359 on chromosome (Chr) 7 (linkage group M) (Takada et al., 2010). Consequently, new SSR markers were designed based on the soybean genomic sequence Glyma1.0 (http://www.phytozome.net/soybean; Schmutz et al., 2010) between these two SSR markers (see Supplemental Table 2 online). For the fine-scale identification of the *Sg-1* locus, 284 F2 individuals of the ST and JI populations were analyzed with these SSR markers as well as the common SSR markers on linkage group M (Cregan et al.,

Figure 1. (continued).

detected in each seed of **(a)** *Sg-1*^{a} or **(b)** *Sg-1*^{b} genotype, respectively. Neither Aa nor Ab was detected in seeds of the **(c)** *sg-1*^{0} genotype where peak 3 (saponin A0-αg, m/z = 1107.5) was detected.

(C) Genetic linkage maps of the *Sg-1* locus determining group A saponin components and DNA markers in soybean Chr 7 (linkage group M). Genetic distances of DNA markers and the *Sg-1* locus are shown in centimorgans. DNA marker information is shown in Supplemental Table 2 online. **(a)** F2 population derived from a cross between 'Shiro-sennari' (*Sg-1*^{a}) and 'Tohoku 152' (*Sg-1*^{b}). **(b)** F2 population derived from a cross between 'JP-36121' (*sg-1*^{0}) and 'Ibarakimame 7' (*Sg-1*^{b}).

1999; Hwang et al., 2009). The *Sg-1* locus was eventually mapped along with GMES6390 and sc68_14AT between Satt336 and sc68_160aAT (Figure 1C). These results indicated that the location of the *Sg-1* locus was restricted physically to the 168-kb region that included sc68_160aAT (43,058 kb on Chr 7) and Satt336 (43,226 kb) (see Supplemental Table 3 online). There were 21 genes (Glyma07g38260 to Glyma07g38500) predicted in this region between sc68_160aAT and Satt336 (see Supplemental Table 4 online). Among the predicted genes, Glyma07g38460 and Glyma07g38470 were expected to encode sugar transferases and were highly similar (Figure 2A).

Gene Structure of Multiple Alleles of the *Sg-1* Locus

Genomic regions (43,197,140 to 43,202,444 bp on Chr 7) containing Glyma07g38470 (*Sg-1-like* in Figure 2A) were amplified and cloned from 'Shiro-sennari,' 'Moshidou Gong 503,' 'Ohsuzu,' 'Ibarakimame 7,' 'Kinusayaka,' and 'JP-36121' (see Supplemental Figures 2 and 3 online). The sequences of this region were nearly identical among these six genotypes and with the 'Williams 82' sequence in Glyma1.0. 'Ohsuzu,' 'Ibarakimame 7,' and 'Kinusayaka' had an identical sequence (AB628093) to 'Williams 82.' Although 18 single nucleotide polymorphisms and two insertion/deletion mutations were observed among the sequences, these polymorphisms were not related to differences in the respective compositions of group A saponins.

In contrast with Glyma07g38470, the corresponding sequences (43,193,410 to 43,196,571 bp on Chr 7) of Glyma07g38460 (*Sg-1* in Figure 2A) demonstrated a close relationship to group A saponin composition. The predicted coding sequences were completely identical among three *Sg-1^b* genotype varieties (AB628089 in the DDBJ/EMBL/GenBank databases) 'Ohsuzu,' 'Ibarakimame 7,' and 'Williams 82,' whereas the *Sg-1^a* genotype varieties (AB628091) 'Shiro-sennari' and 'Moshidou Gong 503' exhibited substitutions of 19 nucleotides compared with *Sg-1^b* (see Supplemental Figure 4A online). Furthermore, two varieties, 'Kinusayaka' (AB628090) and 'JP-36121' (AB628092), which accumulate the nonacetylated saponin A0-αg, were found to have deletions in the sequences corresponding to *Sg-1^b* and *Sg-1^a*, respectively. A cleaved amplified polymorphic sequences marker and two insertion/deletion markers were developed to examine the identity of these structural variations in soybean germplasm (see Supplemental Figure 4B and Supplemental Table 2 online). The group A saponin compositions of all the genetic resources tested (Hwang et al., 2008) coincided with the genotypes of these diagnostic markers (see Supplemental Table 5 online). Moreover, the primary structure of Glyma07g38460 was predicted to encode a UGT. Taken together, these results indicated that Glyma07g38460 was the most likely candidate for the *Sg-1* locus. Thus, Glyma07g38460 and Glyma07g38470 are hereafter referred to as *Sg-1* and *Sg-1-like*, respectively (Figure 2A).

Transcripts of *Sg-1* were detected in hypocotyls of developing seeds with different *Sg-1* genotypes (Figure 2B). It is important to note that truncated transcripts were observed in 'Kinusayaka' and 'JP-36121,' which lack saponins Aa and Ab. *Sg-1* also was expressed in developing cotyledons, where group A saponins are not detected (Shimoyamada et al., 1990; Shiraiwa et al., 1991a, 1991b). To clarify the primary structures of the *Sg-1* alleles, the complete coding regions were isolated from different *Sg-1* genotypes. The corresponding sequences of *Sg-1^a* and *Sg-1^b* were 1431 bp in length and had no intron (Figure 2C). In addition, a sequence (GMFL02-39-F07) identical to *Sg-1^a* was found in the Soybean Full-Length cDNA Database of 'Norin 2' (Umezawa et al., 2008). *Sg-1^a* and *Sg-1^b* showed significant structural similarity (98.3% amino acid sequence identity), and only nine amino acids were different between the two genes. Both *sg-1^0* alleles showed truncated amino acid sequences; a deletion of 10 amino acids from the *Sg-1^a* allele was found for *sg-1$^{0\text{-}a}$* from 'JP-36121' and a deletion of 16 amino acids from the *Sg-1^b* allele was found for *sg-1$^{0\text{-}b}$* from 'Kinusayaka.' In addition, the regions adjacent to the coding sequences were confirmed by DNA gel blot analysis (see Supplemental Figure 5 online). The *Hin*dIII and *Eco*RI digestion patterns were in accordance with that of the coding sequences of the *Sg-1* locus, supporting the contention that the two haplotype variations, *Sg-1^a* and *Sg-1^b*, are responsible for group A saponin composition.

Complementation Test of *sg-1^0* Allele by *Sg-1^b*

To confirm the biological functions of *Sg-1* genes in vivo, a genomic sequence of *Sg-1^b* ('Williams 82') including 2.5-kb 5′-end and 1-kb 3′-end flanking regions (see Supplemental Figure 6 online) was introduced into >400 explants of 'Kinusayaka,' which has the natural *sg-1$^{0\text{-}b}$* allele. By selection for herbicide resistance and red fluorescent luminescence, two transgenic plants were regenerated and produced seeds. However, one transgenic plant failed to transmit the transgene into T_1 progenies as described elsewhere (Yamada et al., 2010). The other transgenic plant produced seeds harboring an exogenous *Sg-1^b* gene (see Supplemental Figure 7 online). T_1 seeds segregated for the presence of *Sg-1^b* so that transgenic seeds with *Sg-1^b* and siblings without the transgene could be used for analysis of group A saponin composition. The transgenic seeds produced saponin Ab as did 'Williams 82' instead of saponin A0-αg (Figure 3), demonstrating that the exogenously introduced *Sg-1^b* successfully complements the deficiency in terminal acetylglucose at the C-22 position. These results strongly indicated that the *sg-1$^{0\text{-}b}$* allele is a loss-of-function allele of the *Sg-1* locus and that *Sg-1^b* encodes an UDP-glucosyltransferase catalyzing the addition of Glc to saponin A0-αg.

Biochemical Characterization of Recombinant Sg-1 Proteins

The disappearance of the group A acetylated saponin and the accumulation of the precursor, A0-αg, in soybean homozygous for the *sg-1^0* allele is strong evidence in support of a role for Sg-1 in glycosylation of A0-αg in soyasaponin biosynthesis. To validate the biochemical properties of Sg-1^a and Sg-1^b proteins, His-tagged recombinant proteins corresponding to *Sg-1^a* and *Sg-1^b* were heterologously expressed as soluble proteins in *Escherichia coli* cells. These recombinant proteins were purified with a nickel-charged column (see Supplemental Figure 8 online) and then subjected to analysis of substrate specificity and catalytic properties using A0-αg as a glycosyl acceptor and

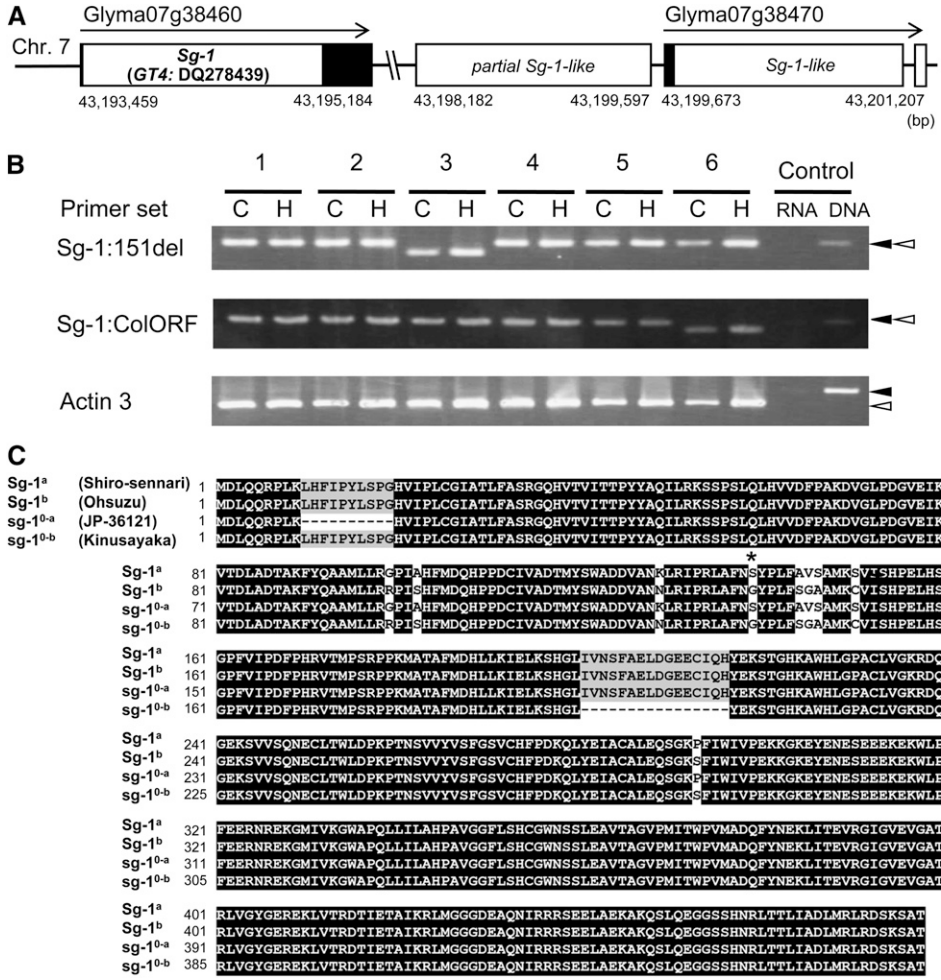


Figure 2. Extraction of Glycosyltransferase Genes of the *Sg-1* Locus.

(A) Schematic map of glycosyltransferase genes in the proximal regions of the *Sg-1* locus on Chr 7. The two predicted glycosyltransferase genes, Glyma07g38460 and Glyma07g38470, were isolated based on the transcript information of 'Williams 82' in Glyma1.0 (Schmutz et al., 2010) according to the linkage maps of the *Sg-1* locus (Figure 1C; see Supplemental Table 3 online).

(B) RT-PCR analysis of Glyma07g38460 expression in cotyledon (C) and hypocotyl (H) in developing seeds of six genotypes, 1, 'Ohsuzu' (*Sg-1^b*); 2, 'Shiro-sennari' (*Sg-1^a*); 3, 'Kinusayaka' (*sg-1^{0-b}*); 4, 'Tohoku 152' (*Sg-1^b*); 5, 'Ibarakimame 7' (*Sg-1^b*); 6, JP-36121 (*sg-1^{0-a}*). The synthesized cDNA was amplified by the two primer sets Sg-1:151del and Sg-1:ColORF, which amplify partial sequences of Glyma07g38460, including the deleted sequences detected in 'Kinusayaka' (*sg-1^{0-b}*) and 'JP-36121' (*sg-1^{0-a}*), respectively (see Supplemental Figure 4A online). Expression of the gene for Actin 3 was examined as an internal control. Controls of total RNA and DNA prepared from the hypocotyls of 'Shiro-sennari' were also amplified by the three primer sets. Open arrowheads indicate the positions of the cDNA amplicons corresponding to each target sequence, whereas the closed arrowheads indicate those of genomic amplicons of control (DNA lane).

(C) Amino acid alignments of the candidate Sg-1 genes of four different genotypes: 'Shiro-sennari' (*Sg-1^a*), 'Ohsuzu' (*Sg-1^b*), 'JP-36121' (*sg-1^{0-a}*), and 'Kinusayaka' (*sg-1^{0-b}*). 'JP-36121' and 'Kinusayaka' were identified as the *sg-1^0* genotype with A0-αg, but they gave rise to different amino acid sequences with truncated forms of Sg-1^a and Sg-1^b, respectively. Asterisk indicates the crucial amino acids for the sugar donor specificity (Figure 5).

UDP-Glc, UDP-Xyl, UDP-Gal, and UDP-GlcUA as glycosyl donors. Reaction mixtures were subsequently analyzed by LC-MS (Figure 4). Sg-1^a protein showed marked xylosyl transfer activity from UDP-Xyl to A0-αg (Figure 4B). The mass spectrometry (MS) fragment pattern of the product was consistent with that of deacetyl Aa. The estimated apparent K_m values for A0-αg and UDP-Xyl were 19.4 ± 6.2 mM and 112.5 ± 29.8 mM, respectively (Table 1). The k_{cat} and the specificity constants values (i.e., the K_{cat}/K_m values) for A0-αg were determined to be 0.108 ± 0.010 s^{-1} and 5.57 s^{-1} mM^{-1}, respectively. By sharp contrast, UDP-Glc, UDP-Gal, and UDP-GlcA were inert as glycosyl donors for the catalysis of Sg-1^a (Figure 5A; see Supplemental Figure 9 online). Thus, these results clearly demonstrate that Sg-1^a is the previously unidentified UDP-Xyl–dependent xylosyltransferase for A0-αg and are consistent with the accumulation of Aa (acetylxylosyl-A0-αg) in hypocotyls of soybean plants expressing the *Sg-1^a* allele. By contrast, Sg-1^b protein showed glucosyl transfer activity for A0-αg, resulting in

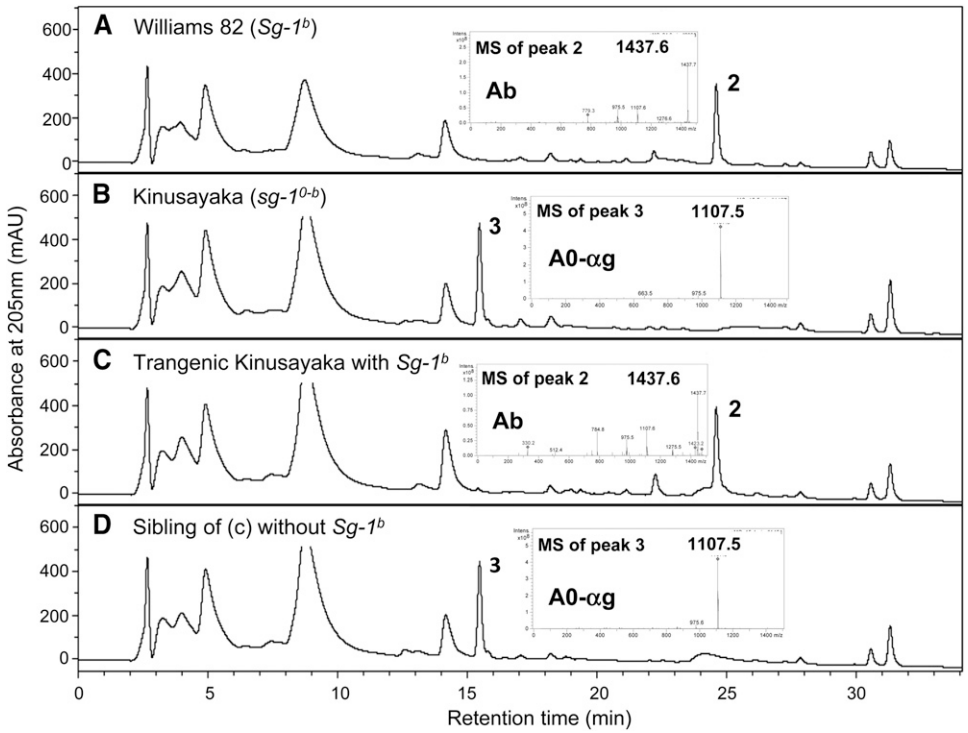


Figure 3. Complementation Analysis of the *Sg-1* Locus for Glycosylation of the Terminal Sugar at the C-22 Position.

LC-MS analyses of 80% methanol extracts obtained from the hypocotyls of transgenic seed and corresponding nontransgenic seeds. 'Williams 82' (*Sg-1^b*) **(A)**, 'Kinusayaka' (*sg-1^{0-b}*) **(B)**, T_1 seed of the transgenic 'Kinusayaka' with the genomic *Sg-1^b* gene of 'Williams 82' **(C)**, and nontransgenic seed separated from their transgenic 'Kinusayaka' siblings **(D)**. Peak numbers for saponin components are the same as in Figure 1B. mAU, milliabsorbance unit.

production of deacetyl Ab (Figure 4D). The estimated apparent K_m values for A0-αg and UDP-Glc were 82.1 ± 27.2 mM and 350.7 ± 89.8 mM, respectively (Table 1). The k_{cat} and the specificity constants values for A0-αg were determined to be 1.215 ± 0.137 s^{-1} and 14.8 s^{-1} mM^{-1}, respectively. UDP-Xyl, UDP-Gal, and UDP-GlcA were not used as glycosyl donors for catalysis by Sg-1^b (Figure 5A; see Supplemental Figure 9 online), showing that Sg-1^b is the UDP-Glc–dependent glucosyltransferase for A0-αg. This result also agrees with the observation that soybean plants expressing the *Sg-1^b* allele accumulate Ab (acetylglucosyl-A0-αg). However, the process for acetylation of the terminal sugar is still unknown. Other possible substrates (soybean saponin aglycones, soyasapogenol A and soyasapogenol B; a *Medicago* saponin, hederagenin; and a soybean isoflavone, daidzein) were tested. However, neither Sg-1^a nor Sg-1^b reacted with these compounds. These results clearly demonstrated that Sg-1 is a UGT specific to A0-αg.

Furthermore, the deletion mutants, *sg-1^{0-a}* from 'JP-36121' (10–amino acid deletion allele of *Sg-1^a*) and *sg-1^{0-b}* from 'Kinusayaka' (16–amino acid deletion allele of *Sg-1^b*) were heterologously expressed in *E. coli* and subjected to enzyme assays. Neither showed any glycosylating activity for A0-αg (*sg-1^{0-a}* result is shown in Supplemental Figure 9E online). The absence of biochemical activity is consistent with lack of deacetyl Aa in *sg-1^{0-a}* and deacetyl Ab in *sg-1^{0-b}* and again strongly supports the notion that both *sg-1^0* alleles are loss-of-function alleles of the *Sg-1* locus.

Homology Model-Based Site-Directed Mutagenesis of Sg-1 Proteins

Although *Sg-1^a* and *Sg-1^b* alleles encode highly similar proteins sharing 476 amino acids (98.3% identity), they clearly exhibited different sugar donor specificity (Figures 4 and 5A). Nine pairs of amino acids differed between the two alleles and only a few residues were expected to be related to the unique sugar donor specificities of the enzymes. To identify crucial amino acids involved in discrimination of sugar donors, homology structure models of Sg-1^a and Sg-1^b were constructed based on the crystal structure of the grapevine (*Vitis vinifera*) anthocyanidin 3-*O*-glucosyltranferase, Vv GT1/UGT78A5 (Protein Data Bank code: 2c1z) (Ford et al., 1998; Offen et al., 2006) (Figure 5B). Each model assumed binding of the appropriate sugar donor and was developed according to molecular dynamics minimization using CHARMm force field for Accelrys Discovery Studio 2.5. Structural comparisons predicted that Ser-138 in the Sg-1^a protein and Gly-138 in the Sg-1^b protein are most likely to be crucial amino acids for specificity of sugar donors. Among the nine pairs of amino acids, Ser-138 and Gly-138 were predicted to be closest to the sugar moieties of the respective sugar donor, UDP-Xyl for Sg-1^a and UDP-Glc for Sg-1^b. Moreover, it was presumed that an OH/π interaction existed between a hydroxy group of Ser-138 and an aromatic ring of Tyr-139 in UDP-Xyl-bound Sg-1^a. On the other hand, in UDP-Glc-bound Sg-1^b, it was realized that Tyr-139 formed an OH/π interaction with a 6-OH group of the Glc moiety

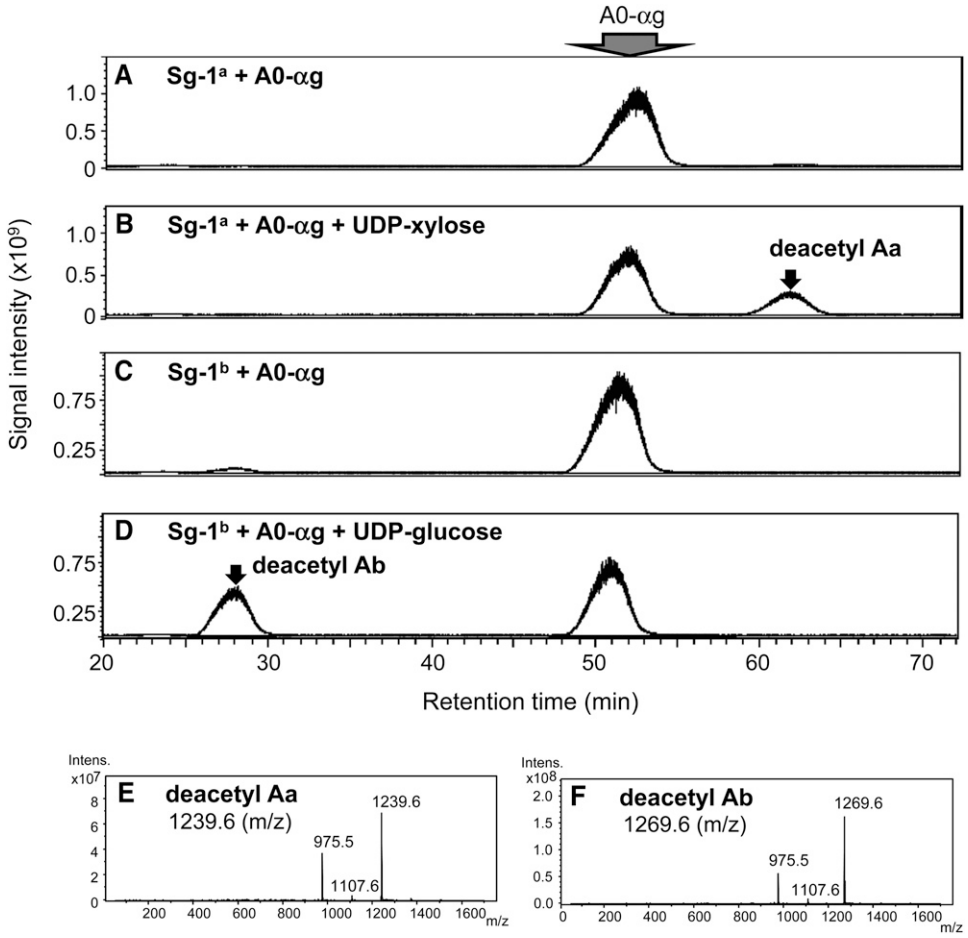


Figure 4. Enzymatic Activity of Sg-1 Proteins for Glycosylation of the Terminal Sugar of the C-22 Position.

LC-MS chart for the enzyme reaction with A0-αg. Sg-1^a **(A)**, Sg-1^a with UDP-Xyl **(B)**, Sg-1^b **(C)**, and Sg-1^b with UDP-Glc **(D)**. MS fragmentations of deacetyl Aa **(E)** and deacetyl Ab **(F)** saponins.

in UDP-Glc and a hydrogen bond interaction was also exhibited between its 6-OH group of UDP-Glc and an amide proton of Gly-138 in Sg-1^b. These two different interactions likely contribute to the sugar donor specificities.

To determine the biochemical impact of these amino acids predicted by the homology models, we constructed mutant forms in which the Ser-138 residue of Sg-1^a was replaced with a Gly residue (S138G) and the Gly-138 residue was replaced with a Ser residue in Sg-1^b (G138S). Enzyme assays showed that, in addition to its original xylosylating activity, Sg-1^a-S138G also displayed new glucosylating and galactosylating activities, both of which were scarcely observed in wild-type Sg-1^a (Figure 5A). By contrast, Sg-1^b-G138S dominantly displayed xylosylating activity instead of the original glucosylating activity. These results showed that Ser-138 in Sg-1^a and Gly-138 in Sg-1^b are crucial residues in specificity for UDP-Xyl and UDP-Glc, respectively.

Phylogenetics of Sg-1 Proteins

Sg-1^a and *Sg-1^b* were genetically and biochemically determined to encode a previously unidentified soyasapogenol xylosyltransferase and glucosyltransferase, respectively, and have been designated as UGT73F4 and UGT74F2 by the committee for UDP-glucuronosyltransferase nomenclature (Mackenzie et al., 1997). Phylogenetic analysis indicated that Sg-1^a and Sg-1^b are classified in a phylogenetic group known as cluster IIIa represented by flavonoid 7-*O*-glycosyltransferases (Noguchi et al., 2008) (Figure 5C). More importantly, this cluster was found to include a subcluster composed of various triterpene/phytosterol-related glycosyltransferases from Fabaceae and Solanaceae (e.g., *M. truncatula* UGT73K1 and UGT73F3; soybean SGT2/UGT73P2 for saponin; tomato [*Solanum lycopersicum*] GAME1 [UGT73L5], GAME2 [UGT73L4], and GAME3 [UGT73L6]; potato [*Solanum tuberosum*] SGT1, SGT2, and SGT3; and *Solanum aculeatissimum* GT4A for steroidal saponin) (Moehs et al., 1997; Achnine et al., 2005; Kohara et al., 2005; McCue et al., 2005, 2006, 2007; Naoumkina et al., 2010; Shibuya et al., 2010; Itkin et al., 2011). Sg-1^a (UGT73F4) and Sg-1^b (UGT73F2) genes apparently show structural similarity to *M. truncatula* saponin glucosyltransferase UGT73F3 (Naoumkina et al., 2010). Thus, Sg-1^a and Sg-1^b are new members of this triterpene/phytosterol subcluster.

Table 1. Kinetic Parameters of Sg-1 Proteins

	K_m (μM)			k_{cat}/K_m (s^{-1} mM^{-1})	
Allele (UGT)	A0-αg	UDP-Sugar	k_{cat} (s^{-1})	A0-αg	UDP-Sugar
Sg-1^a (UGT73F4)	19.4 ± 6.2	112.5 ± 29.8	0.108 ± 0.010	5.57	0.96
Sg-1^b (UGT73F2)	82.1 ± 27.2	350.7 ± 89.8	1.215 ± 0.137	14.80	3.46

Kinetic parameters were determined at pH 7.5 and 30°C with UDP-sugar (2 mM) as described in Methods. A0-αg (100 μM) was used for determining kinetic parameters of UDP-sugar donors. UDP-Xyl and UDP-Glc were used as a UDP-sugar donor for UGT73F4 and UGT73F2, respectively.

DISCUSSION

Soybean saponins are a diverse group of natural products and show a wide variety of biological activities. Saponin structural diversity is generated by sequential oxidations and, especially, by glycosylation. The soybean genome revealed as many as 260 annotated UGT genes, and some of them share high nucleotide sequence similarity, but most of their biochemical functions remain to be clarified. Recently, a few advanced studies identified UGT genes involved in soyasaponin biosynthesis from soybean (Shibuya et al., 2006, 2010).

In an effort to improve soybean seed quality by breeding, extensive genetic resources were screened to obtain genetic variants differing in saponin composition (Shiraiwa et al., 1991a; Tsukamoto et al., 1993). These genetic variants combined with the well-organized soybean genome database facilitated work to elucidate the biosynthetic pathway of soyasaponins. In this study, we successfully identified two UDP-glycosyltrasferases, Sg-1^a (UGT73F4) and Sg-1^b (UGT73F2), for soyasaponin by genetic analysis of multiple alleles of the *Sg-1* locus and subsequent reciprocal biochemical analyses. Sg-1^a (UGT73F4) and Sg-1^b (UGT73F2) catalyze glycosylation at the sugar moiety of the C-22 position of A0-αg (glycoside) saponin, whereas the structurally related UGT73F3 of *M. truncatula* catalyzes glucosylation mainly at the C-28 position of sapogenin aglycone. In consideration of the empirical observation that both recessive *sg-1^0* alleles accumulate A0-αg, the Sg-1 protein is unlikely to be involved in glycosylation of sapogenin aglycones in vivo. In terms of sugar acceptor specificity, Sg-1 proteins are previously uncharacterized sugar-sugar glycosyltransferases that specifically glycosylate a sugar moiety of phytochemical glycosides. Gm SGT2 (UGT73P2) of soybean was recently identified to be a soyasapogenol B monoglucoside 2′′-*O*-galactosyltransferase (Shibuya et al., 2010). It also is a sugar-sugar glycosyltransferase of this triterpene/phytosterol subcluster, although its regiospecificity for the sugar moiety is different from that of Sg-1. Moreover, another member of this subcluster, potato SGT3 also is known to catalyze sugar-sugar rhamnosylation of steroidal saponins, although it shows high sequence similarity to SGT1 and SGT2 that catalyze glycosylation of a steroidal sapogenin, solanidine (McCue et al., 2005, 2006, 2007). In general, flavonoid UGTs form distinct functional clusters based on their regiospecificity, suggesting that the regio-specificity arose prior to speciation (Noguchi et al., 2009). Flavonoid UGTs catalyzing glycosylation of flavonoid aglycones and the sugar moiety of flavonoid glycosides are especially clearly divided into different phylogenetic clusters (Noguchi et al., 2008). In fact, sugar-sugar glycosyltransferases for flavonoids are considered to form a phylogenetically distinct cluster known as cluster IV. From an evolutionary perspective, it is interesting to ask why triterpene/phytosterol-related UGTs are structurally similar irrespective of their regio-specificity. This structural similarity suggests that, compared with flavonoid UGTs, triterpene/phytosterol-related UGTs might have evolved more recently from an ancestral UGT and adapted rapidly to the current glycosylation positions and specific sugar donors. This hypothesis implies that their regiospecificity also could be determined by a relatively small number of amino acids as is the case of sugar donor specificity demonstrated in this study. In addition, this hypothesis should also be examined with other classes of triterpene/phytosterol-related UGTs, such as *M. truncatula* UGT71G1, *Saponaria vaccaria* UGT74M1, and soybean UGT91H4, which are not members of the triterpene/phytosterol-related UGT73 subcluster described in this study (Shao et al., 2005; Meesapyodsuk et al., 2007; Shibuya et al., 2010). Future crystal structural studies of triterpene UGTs should clarify questions about the diverse regiospecificities of triterpene UGT enzymes.

Naturally occurring loss-of-function alleles *sg-1$^{0\text{-}a}$* and *sg-1$^{0\text{-}b}$* have distinct in-frame deletion mutations near the N terminus and middle of the UGT73F4 and UGT73F2 proteins, respectively. Since the deletion of 10 amino acids (LHFIPYLSPG) found in sg-1$^{0\text{-}a}$ from 'JP-36121' is adjacent to the His-20 residue that is considered to be catalytically important as a base for nucleophilic attack in Vv-GT1 (Offen et al., 2006), this deletion seems to be functionally critical for catalysis. On the other hand, the 16 amino acids (IVNSFAELDGEECIQH) deleted in sg-1$^{0\text{-}b}$ from 'Kinusayaka' are predicted to form an α-helix-and-turn structure, which is located outside of the crucial Gly-138 residue in the Sg-1^b homology model, suggesting a structural impact on the correct conformation of the substrate pocket for sugar donor recognition. It is important to note that most of the recombinant protein of sg-1$^{0\text{-}a}$ and sg-1$^{0\text{-}b}$ expressed in *E. coli* was found in the pellet fraction, suggesting that these mutations cause protein aggregation by incorrect folding (see Supplemental Figure 8 online). This could be the major reason for their loss of function.

Sg-1^a and Sg-1^b proteins exhibit alternative sugar donor specificity. The conformational homology models facilitated identification of the critical residues (Ser-138 for Sg-1^a and Gly-138 for Sg-1^b) as the basis for their sugar donor specificity. Interconversion of sugar donor specificity by only one amino acid substitution at this position not only highlights functional plasticity of the UGT enzyme, but also strongly suggests that the two alleles differentiated recently. Previous research indicated that there are three distinguishable regions considered to be

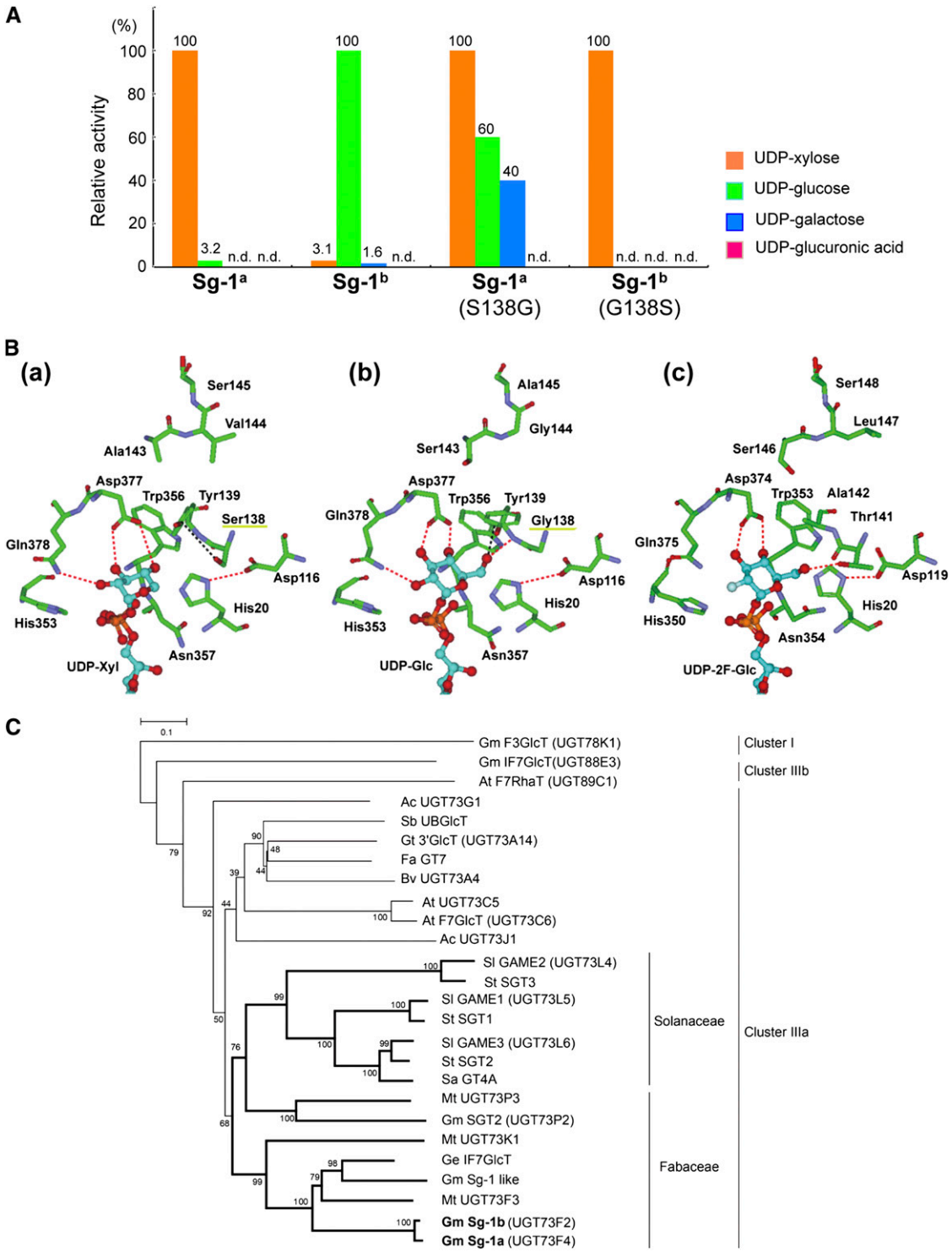


Figure 5. Phylogenetics of Sg-1 Proteins and Their Related UDP-Glycosyltransferases.

(A) Sugar donor specificity of wild-type and mutated Sg-1 proteins obtained by replacing Ser with Gly in Sg-1^a (S138G) and Gly with Ser in Sg-1^b (G138S) at amino acid position 138. The glycosylating activity of each enzyme with four types of sugar donor (UDP-Xyl, UDP-Glc, UDP-Gal, and UDP-GlcUA) was tested. A0-αg was used as a sugar acceptor for evaluating the sugar donor specificity. The highest activity for the four sugar donors is set as 100%. n.d., not detected.

associated with sugar donor specificity of UGT: (1) the N-terminal region represented as Arg-25 of daisy (*Bellis perennis*) UGT94B1 (glucuronosyltransferase) and Pro-19 of Vv-GT6 (bifunctional galactosyl/glucosyltransferase); (2) the middle region represented as Thr-141 of Vv-GT1 (glucosyltransferase) and Arg-140 of Vv-GT5 (glucuronosyltransferase); and (3) the C-terminal putative secondary product glycosyltransferase motif represented as His-374 of *Aralia cordata* GalT and Arg-350 of *Perilla frutescens* UGT88D7 (glucuronosyltransferase) (Kubo et al., 2004; Offen et al., 2006; Osmani et al., 2008; Noguchi et al., 2009; Wang, 2009; Ono et al., 2010). Obviously, Ser-138 of Sg-1^a is located within the middle region. These findings provide useful information for predicting sugar donor specificity and for engineering the biochemical activity of UGTs of interest.

The different OH/π interactions of Tyr-139 with the OH group of Ser-138 in Sg-1^a and the 6-OH group of the Glc moiety of UDP-Glc in Sg-1^b are both likely to contribute to the stabilization of active conformers during the process of glycosylation. Sg-1^b-G138S completely altered sugar donor specificity from UDP-Glc to UDP-Xyl, showing that Ser-138 is the crucial residue for specific recognition of UDP-Xyl. By contrast, Sg-1^a-S138G exhibited glucosylating activity comparable to the original xylosylating activity as well as moderate galactosylating activity. Given that Sg-1^a-S138G still possesses xylosylating activity, other amino acids also must be involved in recognition of UDP-Xyl. According to the model structures, Ala-143, Val-144, and Ser-145 for Sg-1^a or Ser-143, Gly-144, and Ala-145 for Sg-1^b are likely to be influential in the direction and conformation of amino acid residues around Ser-138 for Sg-1^a or Gly-138 for Sg-1^b.

Two recent studies on flavonoid UGTs reported that kiwifruit (*Actinidia deliciosa*) F3GGT1 and *Arabidopsis* UGT79B1 specifically catalyze xylosylation at the 2-OH of the sugar moiety of anthocyanidin 3-*O*-glycoside (Tohge et al., 2005; Montefiori et al., 2011; Yonekura-Sakakibara et al., 2012). Interestingly, they do not have a Ser residue at the position corresponding to Ser-138 in Sg-1^a but have an Ile residue instead. This difference suggests that sugar donor specificity for UDP-Xyl of these xylosyltransferases occurred locally and independently and strongly supports our hypothesis of convergent evolution of sugar donor specificity (Noguchi et al., 2009; Ono et al., 2010). Judging from the biochemical properties, the Ile residue may also be involved in recognition of UDP-Xyl by forming a hydrophobic interaction instead of an interaction between UDP-Xyl and Ser138-Tyr139 in Sg-1^a.

DDMP saponins and their derivative group B and E saponins are widely distributed in aerial, subterranean, and reproductive tissues of legume plants (Price et al., 1986; Shiraiwa et al., 1991a; Oleszek and Stochmal, 2002; Huhman et al., 2005). By contrast, group A saponins are found only in seed hypocotyls of soybean and its wild relatives, *Glycine soja* Sieb. and Zucc (Shimoyamada et al., 1990; Shiraiwa et al., 1990). The genus *Glycine* is divided into subgenera *Glycine* and *Soja*. The subgenus *Glycine* is composed of 16 wild perennial species mainly found in Australia and the Western Pacific region (Palmer et al., 1996; Hymowitz, 2004). The subgenus *Soja* includes cultivated soybean (*G. max*) and its wild annual progenitor (*G. soja*), and they are cross-compatible. With the exception of one observation of Aa saponin in a perennial species, *Glycine tabacina* (Labill.) Benth (Shiraiwa et al., 1990), the phylogenetic relationships in the two subgenera are in good agreement with the existence of group A saponins, suggesting that the biosynthetic pathway for group A saponins arose during the evolution of the subgenus *Soja* from the subgenus *Glycine*. Because Aa and Ab saponins are distributed in both cultivated and wild soybeans (see Supplemental Table 5 online), both alleles would have emerged in the ancestral wild soybean and would have been transmitted to cultivated soybean. It is important to note that the *Sg-1* gene expression was observed not only in developing seeds but also in the various organs we tested, such as leaf and stem, regardless of group A saponin distribution in soybean (see Supplemental Figure 10 online). This inconsistency between gene expression and metabolite distribution suggests the idea that Sg-1 or its ancestor may be involved in glycosylation of other classes of saponins and that it evolved to adapt to group A saponins occurring in an ancestral plant for *Soja* and *Glycine*. However, it is also possible that *Sg-1* is still involved in glycosylation of other saponins. Recent studies of *Medicago* saponins demonstrated that *Medicago* plants produce structurally diverse triterpenoid glycosides and several UGT73F3-like genes respond to a reduction of hemolytic saponins (Carelli et al., 2011; Tava et al., 2011). Thus, *Sg-1* may still be in the process of adapting to group A saponins.

Considering the existence of cultivated and wild mutants accumulating A0-αg, an intermediate compound in the biosynthesis of group A saponins, it may be that group A saponins are less constrained under selection pressure. In other words, the physiological roles of group A saponins may not be fully fixed. This could be one reason why group A saponins are highly diversified in composition and content as specified by different soybean genotypes, compared with DDMP and group B saponins (Shiraiwa et al., 1991a; Tsukamoto et al., 1993; Sasama et al., 2010). The genomic structures and locations of *Sg-1* and its related genes also support this idea. *Sg-1* was located in a tandem arrangement with *Sg-1* partial and *Sg-1-like* genes (Figure 2A). It should be noted that the *Sg-1-like* gene is predicted to have an intron unlike *Sg-1* in the Phytozome database

Figure 5. (continued).

(B) Homology models of Sg-1 proteins. A structure model of UDP-Xyl-bound Sg-1^a (a); a structure model of UDP-Glc-bound Sg-1^b (b); an x-ray crystal structure of grapevine UDP-2F-Glc-bound Vv_GT1 (Protein Data Bank code: 2c1z) (c).

(C) A phylogenetic tree of Sg-1^a (UGT73F4), Sg-1^b (UGT73F2), and related UGTs (available as Supplemental Data Set 1 online). An unrooted phylogenetic tree was constructed using MEGA5 software with the neighbor-joining method based on ClustalW multiple alignments (Thompson et al., 1994; Tamura et al., 2011). The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Bold line indicates a subcluster of triterpene/phytosterol-related UGTs.

(http://www.phytozome.net/), suggesting that the *Sg-1* gene may have emerged via retroposition of the *Sg-1-like* gene or its related UGT gene with an intron (see Supplemental Figure 11 online) (Matsuno et al., 2009). This is consistent with the hypothesis that group A saponins are newly developing metabolites. The biochemical properties of the Sg-1-like protein remain unknown; however, it is likely to be a glucosyltransferase since it has Gly-146 at the position corresponding to Gly-138 in Sg-1^b. In addition, most *Sg-1* homologous genes of Fabaceae plants, including soybean and *Medicago*, also have a Gly residue at the position Gly-138 of Sg-1^b, but not a Ser residue (Figure 5C). Therefore, it is plausible that Sg-1^b (GlcT) is the original allele and that Sg-1^a (XylT) differentiated afterward by acquiring the critical Ser-138 residue (see Supplemental Figure 11 online). The fact that Sg-1^a has trace glucosylating activity whereas Sg-1^b shows trace xylosylating activity also supports this evolutionary hypothesis (Figure 5; see Supplemental Figure 9 online). Further investigation of this repetitive region containing *Sg-1* and its related genes is needed to clarify this evolutionary issue.

Recently, a spontaneous mutant lacking the ability to produce soyasapogenol A was identified through a comprehensive screening of wild soybean germplasm, and the soyasapogenol A deficiency trait is being incorporated into modern breeding lines (Sasama et al., 2010). Elucidation of the corresponding gene and reciprocal phenotyping of this mutant will not only show new insights into the evolution and physiological roles of the group A saponins in the subgenus *Soja*, but also provide useful molecular tools for improvements in the taste and function of soybean products.

METHODS

Plant Materials

Seven soybean (*Glycine max*) cultivars and one wild accession 'JP-36121' (see Supplemental Table 1 online) were used to examine the genomic sequence, gene expression, and gene function of *Sg-1*. The group A saponin composition of each variety also is shown in Supplemental Table 1 online. To identify and isolate the *Sg-1* locus from the soybean genome, genetic mapping of the *Sg-1* locus was conducted with segregating populations derived from the crosses 'JP-36121 (COL/Kumamoto/1984/Seriguchi-1)' × 'Ibarakimame 7' (Takada et al., 2010) and 'Shiro-sennari' × 'Tohoku 152.' The two populations named JI and ST, respectively, consisted of 94 and 190 F2 individuals.

HPLC and LC-MS Analysis of Soybean Saponins

Saponin components, including group A saponins, were extracted from seed hypocotyls with a 50-fold volume (v/w) of 80% (v/v) aqueous methanol for 12 h at room temperature. The extracts were passed through a 0.45-μm filter (Chromatodisk 4P; Kurabo), and a portion (10 μL) of the aliquot was applied onto a reverse-phase HPLC system (Agilent 1200 LC system; Agilent Technologies). The analytical column, Develosil C30-UG-5, 150 × 2.0 mm inner diameter (Nomura Chemical), was kept at 40°C in a thermal chamber. Solvent A consisted of acetonitrile containing 0.1% (by volume) formic acid, and solvent B was 0.1% formic acid solution. The gradient elution was performed at a flow rate of 0.15 mL/min: Solvent A was initiated at 20% (v/v) and maintained for 20 min, increased to 77% (v/v) for 18 min, and then increased to 100% (v/v) for 5 min. The eluent composition was returned to the initial state of 20% (v/v) solvent A for 20 min. The eluate from the column was monitored by a photodiode array detector (Agilent 1200 LC system) and ion-trap mass spectrometer (HCTultra PTM-HS; Bruker Daltonics). Source settings used for the ionization of saponins were as follows: nebulizer gas flow, 50.00 p.s.i.; dry gas flow, 10.00 L/min; capillary temperature, 250°C. Nitrogen (>99.99%) and He (>99.99%) were used as sheath and damping gas, respectively. LC-MS and tandem MS analyses were performed in the positive ion mode of electrospray ionization method using an automatic full scan mode over a mass-to-charge ratio (m/z) range from 50 to 1700. The UV and MS spectra were recorded and analyzed with DataAnalysis software version 3.4 (Bruker Daltonics).

Genetic Mapping of the *Sg-1* Locus

Genomic DNA extraction for ST and JI populations followed the previous report (Takada et al., 2010). Given that *Sg-1* locus was positioned between two SSR markers, Sat_276 and Sat_359, new SSR markers were designed based on the genomic sequence between the two SSR markers in the Glyma1.0 gene set (http://www.phytozome.net/soybean). The primer pairs were designed using the Simple Sequence Repeat Candidate Marker Search Tool in Comprehensive Phytopathogen Genomics Resource (http://cpgr.plantbiology.msu.edu/). Initially, markers were chosen on the condition that they had a total length of 100 to 300 bp and included more than 20 bp of SSR repeat region and that the primers were ~20 bp long with temperature values around 60°C (using default setting). PCR was performed using GoTaq Green Master Mix and its protocol, and the amplicons were observed by 10% nondenaturing polyacrylamide gel electrophoresis and ethidium bromide staining (Hwang et al., 2009). Finally, four markers with clear polymorphisms between parents were genotyped in the ST and JI populations along with the common SSR markers on linkage group M (Cregan et al., 1999; Hwang et al., 2009) (see Supplemental Tables 2 and 3 online). MAPMAKER/EXP version 3.0b was employed to analyze linkage between the markers (Lincoln et al., 1992). Genetic distance (centimorgans) was calculated using the Kosambi mapping function (Kosambi, 1943). Linkage maps were graphically visualized with MapChart (Voorrips, 2002).

Genomic Sequencing of the *Sg-1* Locus

Total DNA was extracted from young leaves with the use of an Automatic DNA Isolation System PI-50α (Kurabo) according to Plant DNA Extraction Protocol version 2. The DNA was used as template to determine the coding and proximal sequences of Glyma07g38460 and Glyma07g38470. Total regions of the genes were initially amplified by primer sets for cloning and sequencing (see Supplemental Table 2 online), and their entire sequences were determined using the primer walking method. Nucleotide sequences were analyzed with a 3130xL Genetic analyzer (Applied Biosystems) using a BigDye-Terminator ver.3.1 cycle sequencing kit (Applied Biosystems).

Expression Analysis of the *Sg-1* Locus

Total RNA was isolated from cotyledons and hypocotyls in developing seeds at mid to late maturation stages with the use of a NucleoSpin RNA plant kit (Macherey-Nagel) according to the procedure described previously (Nishizawa and Ishimoto, 2009). The isolated RNA was subjected to reverse transcription with an oligo(dT)$_{20}$ primer and ReverTra Ace reverse transcriptase (Toyobo). The synthesized cDNA was used as template for RT-PCR analysis with primer sets (see Supplemental Table 2 online) designed according to the *Sg-1* candidate gene (Glyma07g38460) and Actin 3 gene (V00450).

Quantitative RT- PCR of *Sg-1* and *Sg-1-like* in Various Soybean Tissues

Quantitative RT-PCR was performed as described in previous work (Noguchi et al., 2008). Total RNA was prepared from an array of organs

and tissues of cultivar 'William 82' using the RNeasy plant mini kit (Qiagen). Total RNA was treated with DNase I (RNase-free; Qiagen). cDNA synthesis was performed starting with 1 μg of total RNA using SuperScript II kit (Invitrogen). The *Sg-1*, *Sg-1-like*, and Gm-*GAPDH* cDNAs in the mixture were quantified by quantitative real-time PCR with specific primers (see Supplemental Table 2 online). Real-time PCR was performed using a 7500 Real-Time PCR system (Applied Biosystems) and a Power SYBR Green PCR kit (Qiagen). Relative transcription levels were analyzed by a ΔΔcycle threshold method (Applied Biosystems) after normalization to transcription of an internal standard (cDNA encoding Gm-GAPDH). The results are presented as the means ± SE of three independent determinations.

Complementation Analysis of *Sg-1* in Transgenic Soybeans

For the complementation of *sg-1*, we constructed a binary vector, pPZPPD:*Sg-1*^{b} (see Supplemental Figure 6 online), according to the following procedure. First, an expression cassette of a red fluorescent protein (DsRed2) was retrieved from pUR (Nishizawa et al., 2006) by *Xba*I and *Kpn*I digestion and then blunt-ended with T4 DNA polymerase. To generate a binary vector expressing *DsRed2* (pPZP:DsRed), the *DsRed2* expression cassette was inserted into pPZP2028 (Kuroda et al., 2010) that is a T-DNA binary vector based on pPZP202 (Hajdukiewicz et al., 1994), which had been digested with *Bam*HI and *Hin*dIII and blunt-ended. Subsequently, an expression cassette of a bacterial *N*-acetyltransferase gene (*hpat*) was retrieved from pUHG:hpat (Kita et al., 2009) by *Pst*I and *Eco*RI digestion and then blunt-ended. The *hpat* expression cassette was inserted into pPZP:DsRed, which had been digested with *Eco*RI and *Kpn*I and blunt-ended, to generate pPZP:DsRed:HPAT. A multiple cloning site was amplified by PCR with pBluescript SK+ (Stratagene) as a template and a primer set for the SK+ cloning site (see Supplemental Table 2 online) and then inserted into the blunt-ended *Eco*RI site of the pPZP:DsRed:HPAT to generate pPZPPD. A candidate gene (Glyma07g38460) was amplified by PCR with soybean genomic DNA ('Williams 82') as a template and a primer set for the *Sg-1*^{b} allele (see Supplemental Table 2 online) and cloned into pCR4Blunt-TOPO vector (Invitrogen) to generate pCR4:*Sg-1*^{b}. The nucleotide sequences were confirmed with a 3130xL genetic analyzer (Applied Biosystems) using a BigDye-Terminator version 3.1 cycle sequencing kit (Applied Biosystems). The cloned ~5.0-kb fragment was retrieved by digestion with *Not*I and *Spe*I existing in the cloning vector and inserted into the pPZPPD, which had been digested with *Psp*OMI and *Spe*I, to generate pPZPPD:*Sg-1*^{b}.

To generate the transgenic soybean plants containing pPZPPD:*Sg-1*^{b}, *Agrobacterium tumefaciens*–mediated transformation was employed according to the procedure of Paz et al. (2006) with some modifications. In brief, mature soybean seeds of 'Kinusayaka' (*sg-1*^{0}) maintained in a humid plastic container for a week were subjected to overnight imbibition on wet filter paper. The cotyledonary node section of the imbibed seeds was wounded using a microbrush (Yamada et al., 2010) soaked with pPZPPD:*Sg-1*^{b}-harboring *Agrobacterium* (EHA105 strain) suspended in a cocultivation medium containing 0.02% of Silwet L-77 (Bio Medical Science). After inoculation, explants were placed adaxial side down on filter paper laid over cocultivation medium solidified with 0.425% agar. After 5 d of cocultivation, explants inoculated with *Agrobacterium* were washed in liquid shoot induction medium (SIM) containing 25 mg/L Meropen (Dai Nippon Sumitomo Pharma). The explants were set with their adaxial side up on solid SIM containing 0.7% agar and 25 mg/L Meropen for 2 weeks and then transferred onto SIM containing 6 mg/L glufosinate-ammonium (Sigma-Aldrich). After a total of 4 weeks of culture on SIM, explants were transferred to shoot elongation medium containing 25 mg/L Meropen and 6 mg/L glufosinate and were subcultured every 2 weeks. Transgenic shoots emitting DsRed2 fluorescence were dipped in 1 mg/L indole butyric acid then transferred to rooting medium. Rooting plantlets were transferred to soil in pots and grown under greenhouse conditions. Transgene presence in T_1 seeds was identified using PCR analysis with the Sg-1:151del primer set (see Supplemental Table 2 online).

Molecular Cloning of Sg-1 Proteins

Total RNA was extracted from the respective soybean varieties and accessions (see Supplemental Table 1 online) using an RNeasy Plant Mini Kit (Qiagen). cDNAs were synthesized from 1 μg of total RNA using a first-strand synthesis system for RT-PCR (Invitrogen). To obtain cDNA of *Sg-1* genes, PCR with KOD plus DNA polymerase (Toyobo) was run at 94°C for 3 min followed by 30 cycles at 94°C for 1 min, 55°C for 1 min, and 68°C for 2 min using the XhoI-Sg-1-Fw and BamHI-Sg-1-Rv primers (see Supplemental Table 2 online). An amplified fragment was subcloned into pCR-TOPO blunt II (Invitrogen). Sequencing reactions were done using a BigDye-Terminator version 3.1 cycle sequencing kit (Applied Biosystems) and then analyzed using a 3100 genetic analyzer to confirm the nucleotide sequence (Applied Biosystems). Site-directed mutagenesis for Sg-1^{a}-S138G and Sg-1^{b}-G138S was performed according to our previous reports (Noguchi et al., 2007, 2009; Ono et al., 2010). *Sg-1* genes were mutated in vitro using recombinant PCR with XhoI-Sg-1-Fw and BamHI-Sg-1-Rv and following PCR primer sets indicated in Supplemental Table 2 online.

Heterologous Expression of Sg-1 Proteins

Plasmids of *Sg-1* cDNAs were digested using *Xho*I and *Bam*HI. The resulting DNA fragments were ligated with a pET-15b vector (Merck) that had previously been digested with *Xho*I and *Bam*HI. The resultant plasmids were transformed into *Escherichia coli* BL21(DE3). The transformant cells were precultured at 37°C for 16 h in a Luria-Bertani broth containing 50 μg/mL ampicillin. Twenty milliliters of the preculture was then inoculated into 800 mL of the same medium. After incubation at 37°C until the OD_{600} reached 0.5, isopropyl 1-β-D-thiogalactoside was added to the broth at a final concentration of 0.4 mM, followed by further incubation at 22°C for 20 h. All subsequent operations were conducted at 0 to 4°C. The recombinant *E. coli* cells were harvested by centrifugation (7000*g*, 15 min), washed with distilled water, and resuspended in buffer A (20 mM sodium Pi, pH 7.4, containing 14 mM 2-mercaptoethanol and 0.5 M NaCl) containing 20 mM imidazole. The cells were disrupted at 4°C by five cycles of ultrasonication (where one cycle corresponds to 10 kHz for 1 min followed by an interval of 1 min). The cell debris was removed by centrifugation (7000*g*, 15 min). Polyethyleneimine was slowly added to the supernatant solution to a final concentration of 0.12% (v/v). The mixture was allowed to stand at 4°C for 30 min, followed by centrifugation (7000*g*, 15 min). The supernatant was applied to a HisTrapTM HP column (1 mL; GE Healthcare UK) that had been equilibrated with buffer A containing 20 mM imidazole. The column was washed with buffer A containing 20 mM imidazole and the enzyme was eluted with buffer A containing 200 mM imidazole. The active column-bound fractions were concentrated and desalted using Vivaspin 30,000 MWCO (GE Healthcare UK), followed by substitution with buffer B (20 mM potassium Pi, pH 7.5, containing 14 mM 2-mercaptoethanol). The protein concentration was determined using the Bradford method (Bradford, 1976) with BSA as a standard. SDS-PAGE was performed according to the method of Laemmli (1970), and the proteins in the gels were visualized by Coomassie Brilliant Blue R 250.

Enzyme Assays of Sg-1 Proteins

The standard reaction mixture (50 μL) consisted of 100 μM glycosyl acceptor (A0-αg), 2 mM glycosyl donor, 50 mM potassium phosphate buffer, pH 7.5, and enzyme. After a 10-min preincubation of the mixture without the enzyme at 30°C, the reaction was initiated by addition of the enzyme. After incubation at 30°C for 10 min, the reaction was stopped by freezing in liquid N_2. In assays for sugar donor specificity, reaction mixtures were incubated for 5 min. The substrates and glycosylated

products were analyzed using LC-MS. Several additional potential substrates were tested: soyasapogenol A and soyasapogenol B (soybean saponins); hederagenin (*Medicago truncatula* saponin); and genistein (soybean isoflavone). Soyasapogenol A and B were purchased from Koshiro, whereas hederagenin and genistein were from Extrasynthese Chemical. The reaction solution was mixed in equal amount of acetonitorile containing 0.1% formic acid, and the supernatant was filtered through a 0.45-μm filter. The filtered samples (5 μL) were injected into a Develosil C30-UG-5 column. Elution was performed with a linear-gradient system consisting of two mobile phases (A, acetonitorile-formic acid = 99.9:0.1; B, water-formic acid = 99.9:0.1). The gradient was initiated at 20% (v/v) A and maintained for 65 min, increased to 100% (v/v) A for 5 min, and then decreased to 20% (v/v) A for 20 min. The flow rate was kept at 0.15 mL/min for a total run time of 90 min. The MS analyses were performed under the same setting conditions as the saponin composition analysis described above.

To determine the initial velocity of Sg-1 proteins, the assays were performed under steady state conditions using the standard assay system (see above) with various substrate concentrations. The apparent K_m and V_{max} values for glycosyl donors and the sugar acceptor (A0-αg) in the presence of a saturating concentration of the counter substrate were determined by fitting the initial velocity data to the Michaelis-Menten equation using nonlinear regression analysis (Segel, 1975; Leatherbarrow, 1990).

Homology Modeling of Sg-1 Proteins

Multiple alignment analysis was performed using a ClustalW 1.83 installed in Genetyx 9.0. The crystal structure of Vv_GT1 (Protein Data Bank code: 2c1z) was used to construct homology models of Sg-1^a and Sg-1^b using the Modelor module installed in the Discovery Studio 2.5 (Accelrys). Before optimization, the UDP-2F-Glc bound in Vv_GT1 was inserted into the constructed Sg-1^a and Sg-1^b models and the sugar moiety of UDP-2F-Glc was replaced with Xyl or Glc moieties as needed. Structure optimization of each model was performed using the molecular mechanics and molecular dynamics simulation with CHARMm force field in the Discovery Studio 2.5 (Accelrys).

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL/DDBJ databases under the following accession numbers: *Sg-1^a*, AB628091; *Sg-1^b*, AB628089; *sg-1$^{0\text{-}a}$*, AB628092; *sg-1$^{0\text{-}b}$*, AB628090; *Sg-1-like* , AB628093, AB649293, AB649294, and AB649295; soybean actin, V00450; soybean glyceraldehyde-3-phosphate dehydrogenase 1, AF061564; Gm IF7GlcT, AB292164; Gm F3GlcT, GU434274; At F7RhaT, NP_563756; Ac UGT73G1, AAP88406; Ac UGT73J1, AAP88407; Fa GT7, ABB92749; Sb UBGlcT, AB031274; Gt 3′GlcT, AB076697; Bv UGT73A4, AY526080; At UGT73C5, AAD20156; At F7GlcT, NM_129234; Mt UGT73P3, FJ477889; Gm SGT2, BAI99584; St SGT3, ABB84472; St SGT1, U82367; St SGT2, DQ218276; Sa GT4A, AB182385; Mt UGT73K1, AAW56091; Mt UGT73F3, FJ477891; Ge IF7GlcT, AB098614; Sl GAME1, HQ293016; Sl GAME2, HQ293018; and Sl GAME3, HQ293017. Gm, *G. max*; At, *Arabidopsis thaliana*; Ac, *Allium cepa*; Fa, *Fragaria* × *ananassa*; Sb, *Scutellaria baicalensis*; Gt, *Gentiana triflora*; Bv, *Beta vulgaris*; Mt, *M. truncatula*; St, *Solanum tuberosum*; Sa, *Solanum aculeatissimum*; Ge, *Glycyrrhiza echinata*; Sl, *Solanum lycopersicum*.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Chemical Structures of Three Group A Saponin Components Found in Seed Hypocotyls of Different Soybean Genotypes.

Supplemental Figure 2. Multiple Alignments of Deduced Amino Acid Sequences of Glyma07g38470 (*Sg-1-like*) of Seven Different Soybean Genotypes.

Supplemental Figure 3. Multiple Alignments of 5′ Genomic Sequences of Glyma07g38470 of Seven Different Soybean Genotypes.

Supplemental Figure 4. Schematic Illustration of Primary Structures of *Sg-1* Genes and Molecular Markers for the Identification of *Sg-1* Genotypes in Soybean.

Supplemental Figure 5. DNA Gel Blot Analysis of the Candidate *Sg-1* Gene (Glyma07g38460).

Supplemental Figure 6. Map of pPZPPD:*Sg-1^b*.

Supplemental Figure 7. Segregation of Transgenes in Seeds Produced by the Primary 'Kinusayaka' (*sg-1$^{0\text{-}b}$*) Plant Transformed with pPZPPD:*Sg-1^b*.

Supplemental Figure 8. SDS-PAGE of Recombinant Sg-1 Proteins.

Supplemental Figure 9. Sugar Donor Specificity of Sg-1 Proteins for Glycosylation of the Terminal Sugar of the C-22 Position.

Supplemental Figure 10. Gene Expression of the Two Predicted Glycosyltransferase Genes, *Sg-1* (Glyma07g38460) and *Sg-1-like* (Glyma07g38470).

Supplemental Figure 11. A Proposed Model of Evolution of the *Sg-1* Locus in Soybean.

Supplemental Table 1. Composition of Group A Saponins in Soybean Seeds Used in This Study.

Supplemental Table 2. Primer Sets Used in This Study.

Supplemental Table 3. Relationship between Group A Saponin Compositions and SSR Marker Genotypes Flanking the *Sg-1* Locus in Two F2 Populations.

Supplemental Table 4. Functional Annotation for the Genes Located between the Two Molecular Markers sc68_160aAT and Satt336.

Supplemental Table 5. Genotyping of *Sg-1* Alleles with Diagnostic Markers and Allelic Frequencies for the *Sg-1* Locus in Cultivated and Wild Soybeans.

Supplemental Data Set 1. Text File of the Alignment Used for the Phylogenetic Analysis Shown in Figure 5C.

ACKNOWLEDGMENTS

We thank to Kunihiko Komastu (National Agricultural Research Center for Hokkaido Region), Hiromi Toyonaga (Suntory Ltd.), Nobuo Tsuruoka (Suntory Ltd.), and Yosuke Kawai (Ritsumeikan University) for technical support. This work was supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-Oriented Industry (BRAIN) and, in part, by a grant from the Ministry of Agriculture, Forestry, and Fisheries of Japan (Research Project for Utilizing Advanced Technologies in Agriculture, Forestry, and Fisheries, 18063).

AUTHOR CONTRIBUTIONS

T.S., E.O., K.T., M.H., and C.T. designed and performed the research, analyzed data, and wrote the article. Y.T., Y.N., A.H., H.S., M.O., H.H., T.T., A.K., S.K., and N.T. performed experiments and analyzed data. M.I. designed the entire research and wrote the article.

Received December 24, 2011; revised April 20, 2012; accepted April 30, 2012; published May 18, 2012.

REFERENCES

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W., and Dixon, R.A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant J. **41:** 875–887.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:** 248–254.

Carelli, M. et al. (2011). Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell **23:** 3070–3081.

Confalonieri, M., Cammareri, M., Biazzi, E., Pecchia, P., Fevereiro, M.P.S., Balestrazzi, A., Tava, A., and Conicella, C. (2009). Enhanced triterpene saponin biosynthesis and root nodulation in transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel β-amyrin synthase (*AsOXA1*) gene. Plant Biotechnol. J. **7:** 172–182.

Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., VanToai, T.T., Lohnes, D.G., Chung, J., and Specht, J.E. (1999). An integrated genetic linkage map of the soybean genome. Crop Sci. **39:** 1464–1490.

Dalton, D.A. (1995). Antioxidant defenses of plants and fungi. In Oxidative Stress and Antioxidant Defenses in Biology, S. Ahmad, ed (New York: Chapman and Hall), pp. 298–355.

Ellington, A.A., Berhow, M., and Singletary, K.W. (2005). Induction of macroautophagy in human colon cancer cells by soybean B-group triterpenoid saponins. Carcinogenesis **26:** 159–167.

Ellington, A.A., Berhow, M.A., and Singletary, K.W. (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, G.R., Price, K.R., Tsukamoto, C., and Okubo, K. (1991). Saponins. In Toxic Substances in Crop Plants, J.P.F. D'Mello, C.M. Duffus, and J.H. Duffus, eds (Cambridge, UK: The Royal Society of Chemistry), pp. 285–327.

Ford, C.M., Boss, P.K., and Høj, P.B. (1998). Cloning and characterization of *Vitis vinifera* UDP-glucose:flavonoid 3-*O*-glucosyltransferase, a homologue of the enzyme encoded by the maize *Bronze-1* locus that may primarily serve to glucosylate anthocyanidins *in vivo*. J. Biol. Chem. **273:** 9224–9233.

Hajdukiewicz, P., Svab, Z., and Maliga, P. (1994). The small, versatile *pPZP* family of *Agrobacterium* binary vectors for plant transformation. Plant Mol. Biol. **25:** 989–994.

Huhman, D.V., Berhow, M.A., and Sumner, L.W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. J. Agric. Food Chem. **53:** 1914–1920.

Hwang, T.Y., Nakamoto, Y., Kono, I., Enoki, H., Funatsuki, H., Kitamura, K., and Ishimoto, M. (2008). Genetic diversity of cultivated and wild soybeans including Japanese elite cultivars as revealed by length polymorphism of SSR markers. Breed. Sci. **58:** 315–323.

Hwang, T.Y. et al. (2009). High-density integrated linkage map based on SSR markers in soybean. DNA Res. **16:** 213–225.

Hymowitz, T. (2004). Speciation and cytogenetics. In Soybeans: Improvement, Production, and Uses, 3rd ed, H.R. Boerma and J.E. Specht, eds (Madison, WI: American Society of Agronomy-Crop Science Society of America-Soil Science Society of America), pp. 97–136.

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

Itkin, M. et al. (2011). GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. Plant Cell **23:** 4507–4525.

Kato, S., Yumoto, S., Takada, Y., Kono, Y., Shimada, S., Sakai, T., Shimada, H., Takahashi, K., Adachi, T., Tabuchi, K., and 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., and 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.

Kita, Y., Hanafy, M.S., Deguchi, M., Hasegawa, H., Terakawa, T., Kitamura, K., and Ishimoto, M. (2009). Generation and characterization of herbicide-resistant soybean plants expressing novel phosphinothricin *N*-acetyltransferase genes. Breed. Sci. **59:** 245–251.

Kitagawa, I., Taniyama, T., Nagahama, Y., Okubo, K., Yamauchi, F., and Yoshikawa, M. (1988). Saponin and sapogenol. XLII. Structures of Acetyl-soyasaponins A1, A2, and A3, astringent partially acetylated bisdesmosides of soyasapogenol A, from American soybean, the seeds of *Glycine max* MERRILL. Chem. Pharm. Bull. (Tokyo) **36:** 2819–2828.

Kohara, A., Nakajima, C., Hashimoto, K., Ikenaga, T., Tanaka, H., Shoyama, Y., Yoshida, S., and Muranaka, T. (2005). A novel glucosyltransferase involved in steroid saponin biosynthesis in *Solanum aculeatissimum*. Plant Mol. Biol. **57:** 225–239.

Kosambi, D.D. (1943). The estimation of map distances from recombination walues. Ann. Hum. Genet. **12:** 172–175.

Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004). Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch. Biochem. Biophys. **429:** 198–203.

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

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

Kudou, S., Tonomura, M., Tsukamato, C., Uchida, T., Yoshikoshi, M., and Okubo, K. (1994). Structural elucidation and physiological properties of genuine soybean saponins. In Food Phytochemicals for Cancer Prevention I, M.T. Huang, T. Osawa, C.T. Ho, and R.T. Rosen, eds (Washington, DC: American Chemical Society), pp. 340–348.

Kuroda, M., Kimizu, M., and Mikami, C. (2010). A simple set of plasmids for the production of transgenic plants. Biosci. Biotechnol. Biochem. **74:** 2348–2351.

Kuzina, V., Ekstrøm, C.T., Andersen, S.B., Nielsen, J.K., Olsen, C.E., and Bak, S. (2009). Identification of defense compounds in *Barbarea vulgaris* against the herbivore *Phyllotreta nemorum* by an ecometabolomic approach. Plant Physiol. **151:** 1977–1990.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680–685.

Leatherbarrow, R.J. (1990). Using linear and non-linear regression to fit biochemical data. Trends Biochem. Sci. **15:** 455–458.

Lincoln, S.E., Daly, M.J., and Lander, E.S. (1992). Construction Genetic Maps with MAPMAKER/EXP 3.0. Technical Report, 3rd ed. (Cambridge, MA: Whitehead Institute).

Mackenzie, P.I. et al. (1997). The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. Pharmacogenetics **7:** 255–269.

Masuda, T., and Goldsmit, P.D. (2009). World soybean production: area harvested, yield, and long-term projections. Int. Food Agrib. Manag. Rev. **12:** 143–162.

Matsuno, M. et al. (2009). Evolution of a novel phenolic pathway for pollen development. Science **325:** 1688–1692.

McCue, K.F., Allen, P.V., Shepherd, L.V., Blake, A., Maccree, M.M., Rockhold, D.R., Novy, R.G., Stewart, D., Davies, H.V., and Belknap, W.R. (2007). Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. Phytochemistry **68:** 327–334.

McCue, K.F., Allen, P.V., Shepherd, L.V., Blake, A., Whitworth, J., Maccree, M.M., Rockhold, D.R., Stewart, D., Davies, H.V., and Belknap, W.R. (2006). The primary *in vivo* steroidal alkaloid glucosyltransferase from potato. Phytochemistry **67:** 1590–1597.

McCue, K.F., Shepherd, L.V.T., Allen, P.V., Maccree, M.M., Rockhold, D.R., Corsini, D.L., Davies, H.V., and Belknap, W.R. (2005). Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. Plant Sci. **168:** 267–273.

Meesapyodsuk, D., Balsevich, J., Reed, D.W., and Covello, P.S. (2007). Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol. **143:** 959–969.

Moehs, C.P., Allen, P.V., Friedman, M., and Belknap, W.R. (1997). Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. Plant J. **11:** 227–236.

Montefiori, M., Espley, R.V., Stevenson, D., Cooney, J., Datson, P. M., Saiz, A., Atkinson, R.G., Hellens, R.P., and Allan, A.C. (2011). Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (*Actinidia chinensis*). Plant J. **65:** 106–118.

Murata, M., Houdai, T., Morooka, A., Matsumori, N., and Oishi, T. (2005). Membrane interaction of soyasaponins in association with their antioxidation effect. Soy Prot. Res. **8:** 81–85.

Murata, M., Houdai, T., Yamamoto, H., Matsumori, N., and Oishi, T. (2006). Membrane interaction of soyasaponins in association with their antioxidation effect–Analysis of biomembrane interaction. Soy Prot. Res. **9:** 82–86.

Naoumkina, M.A., Modolo, L.V., Huhman, D.V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W., and Dixon, R.A. (2010). Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. Plant Cell **22:** 850–866.

Nishizawa, K., and Ishimoto, M. (2009). Maturation of somatic embryos as a model for soybean seed development. Plant Biotechnol. **26:** 543–550.

Nishizawa, K., Kita, Y., Kitayama, M., and Ishimoto, M. (2006). A red fluorescent protein, DsRed2, as a visual reporter for transient expression and stable transformation in soybean. Plant Cell Rep. **25:** 1355–1361.

Noguchi, A., Horikawa, M., Fukui, Y., Fukuchi-Mizutani, M., Iuchi-Okada, A., Ishiguro, M., Kiso, Y., Nakayama, T., and Ono, E. (2009). Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales. Plant Cell **21:** 1556–1572.

Noguchi, A., Fukui, Y., Iuchi-Okada, A., Kakutani, S., Satake, H., Iwashita, T., Nakao, M., Umezawa, T., and Ono, E. (2008). Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. Plant J. **54:** 415–427.

Noguchi, A., Saito, A., Homma, Y., Nakao, M., Sasaki, N., Nishino, T., Takahashi, S., and Nakayama, T. (2007). A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (*Glycine max*) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. J. Biol. Chem. **282:** 23581–23590.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. **25:** 1396–1405.

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

Oleszek, W., and Stochmal, A. (2002). Triterpene saponins and flavonoids in the seeds of *Trifolium* species. Phytochemistry **61:** 165–170.

Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., Imai, H., Takahashi, S., Kawai, Y., Ishiguro, M., Fukui, Y., and Nakayama, T. (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant Cell **22:** 2856–2871.

Osbourn, A. (1996). Saponins and plant defence — A soap story. Trends Plant Sci. **1:** 4–9.

Osmani, S.A., Bak, S., Imberty, A., Olsen, C.E., and Møller, B.L. (2008). Catalytic key amino acids and UDP-sugar donor specificity of a plant glucuronosyltransferase, UGT94B1: Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant Physiol. **148:** 1295–1308.

Palmer, R.G., Hymowitz, T., and Nelson, R.L. (1996). Germplasm diversity within soybean. In Soybean: Genetics, Molecular Biology and Biotechnology, D.P.S. Verma and R.C. Shoemaker, eds (Wallingford, UK: CAB International), pp. 1–35.

Papadopoulou, K., Melton, R.E., Leggett, M., Daniels, M.J., and Osbourn, A.E. (1999). Compromised disease resistance in saponin-deficient plants. Proc. Natl. Acad. Sci. USA **96:** 12923–12928.

Paz, M.M., Martinez, J.C., Kalvig, A.B., Fonger, T.M., and Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient *Agrobacterium*-mediated soybean transformation. Plant Cell Rep. **25:** 206–213.

Price, K.R., Curl, C.L., and Fenwick, G.R. (1986). The saponin content and sapogenol composition of the seed of 13 varieties of legume. J. Sci. Food Agric. **37:** 1185–1191.

Sasama, H., Takada, Y., Ishimoto, M., Kitamura, K., and 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 Chemistry, Texture, and Flavor of Soy, K.R. Cadwallader and S.K.C. Chang, eds (Oxford, UK: Oxford University Press), pp. 91–102.

Schmutz, J. et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature **463:** 178–183.

Segel, I.H. (1975). Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. (New York: John Wiley & Sons).

Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A., and Wang, X. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell **17:** 3141–3154.

Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y. (2006). Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J. **273:** 948–959.

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

Shimoyamada, M., Kudou, S., Okubo, K., Yamauchi, F., and Harada, K. (1990). Distribution of saponin constituents in some varieties of soybean plant. Agric. Biol. Chem. **54:** 77–81.

Shiraiwa, M., Harada, K., and Okubo, K. (1991a). 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., and Okubo, K. (1991b). Composition and structure of "group B saponin" in soybean seed. Agric. Biol. Chem. **55:** 911–917.

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

Sugano, M. (2006). Soy in Health and Disease Prevention. (New York, USA: CRC Press, Taylor & Francis Group).

Takada, Y., Sayama, T., Kikuchi, A., Kato, S., Tatsuzaki, N., Nakamoto, Y., Suzuki, A., Tsukamoto, C., and Ishimoto, M. (2010). Genetic analysis of variation in sugar chain composition at the C-22 position of group A saponins 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., and 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.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:** 2731–2739.

Taniyama, T., Nagahama, Y., Yoshikawa, M., and Kitagawa, I. (1988b). Saponin and sapogenol. XLIII. Acetyl-soyasaponins A4, A5, and A6, new astringent bisdesmosides of soyasapogenol A, from Japanese soybean, the seeds of *Glycine max* MERRILL. Chem. Pharm. Bull. (Tokyo) **36:** 2829–2839.

Taniyama, T., Yoshikawa, M., and Kitagawa, I. (1988a). [Saponin and sapogenol. XLIV. Soyasaponin composition in soybeans of various origins and soyasaponin content in various organs of soybean. Structure of soyasaponin V from soybean hypocotyl]. Yakugaku Zasshi **108:** 562–571.

Tava, A., Pecetti, L., Romani, M., Mella, M., and Avato, P. (2011). Triterpenoid glycosides from the leaves of two cultivars of *Medicago polymorpha* L. J. Agric. Food Chem. **59:** 6142–6149.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Tohge, T. et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J. **42:** 218–235.

Tsujino, Y., Tsurumi, S., Yoshida, Y., and Niki, E. (1994). Antioxidative effects of dihydro-γ-pyronyl-triterpenoid saponin (chromosaponin I). Biosci. Biotechnol. Biochem. **58:** 1731–1732.

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

Tsurumi, S., and Ishizawa, K. (1997). Involvement of ethylene in chromosaponin-induced stimulation of growth in lettuce roots. Plant Cell Physiol. **38:** 668–675.

Tsurumi, S., Ishizawa, K., Rahman, A., Soga, K., Hoson, T., Goto, N., and Kamisaka, S. (2000). Effects of choromosaponin I and brassinolide on the growth of roots in etiolated Arabidopsis seedlings. J. Plant Physiol. **156:** 60–67.

Umezawa, T. et al. (2008). Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length-enriched cDNA library. DNA Res. **15:** 333–346.

Voorrips, R.E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. **93:** 77–78.

Wang, X. (2009). Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Lett. **583:** 3303–3309.

Yamada, T., Watanabe, S., Arai, M., Harada, K., and Kitamura, K. (2010). Cotyledonary node pre-wounding with a micro-brush increased frequency of *Agrobacterium*-mediated transformation in soybean. Plant Biotechnol. **27:** 217–220.

Yamori, Y. (2006). Soy for "health for all": Message from WHO CARDIAC study and dietary intervention studies. In Soy in Health and Disease Prevention, M. Sugano ed (New York: Taylor & Francis Group), pp. 107–122.

Yonekura-Sakakibara, K. et al. (2012). Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. Plant J. **69:** 154–167.

Yoshiki, Y., and Okubo, K. (1995). Active oxygen scavenging activity of DDMP (2,3-dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one) saponin in soybean seed. Biosci. Biotechnol. Biochem. **59:** 1556–1557.