agronomy

MDPI

*Article*

# Genetic and Molecular Characterization of a New EMS-Induced Mutant without the Third Glucose Moiety at the C-3 Sugar Chain of Saponin in *Glycine max* (L.) Merr.

**Jeongyun Ko [1], Jagadeesh Sundaramoorthy [1,2], Chigen Tsukamoto [3], Junbeom Park [1], Jinwon Lee [1], Hyun Jo [1], Jeong-Dong Lee [1], Hak Soo Seo [4] and Jong Tae Song [1,*]**

[1] Department of Applied Biosciences, Kyungpook National University, Daegu 41566, Korea; audrynv98@knu.ac.kr (J.K.); jagadeesh@slu.se (J.S.); junfan59@knu.ac.kr (J.P.); jwlee2018@knu.ac.kr (J.L.); johyun@knu.ac.kr (H.J.); jdlee@knu.ac.kr (J.-D.L.)
[2] Department of Plant Breeding, Swedish University of Agricultural Sciences, 23053 Alnarp, Sweden
[3] Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka 020-8550, Japan; chigen@iwate-u.ac.jp
[4] Department of Agriculture, Forestry and Bioresources, Seoul National University, Seoul 08826, Korea; seohs@snu.ac.kr
* Correspondence: jtsong68@knu.ac.kr; Tel.: +82-053-950-7753

**Abstract:** Saponin, a secondary metabolite, is produced by various plant species, including soybean (*Glycine max* (L.) Merr.). Soybeans synthesize triterpenoid saponins, which are classified by their aglycone structure and sugar chain composition. Here, we characterized an ethyl methanesulfonate-induced mutant, PE1539, without saponin and with a glucose moiety at the third position of the C-3 sugar chain. The saponin phenotype of PE1539 is described by the accumulation of Ab-γg saponin and deficiency of Ab-αg saponin and DDMP-αg saponin, similar to a previously reported *sg-3* mutant in soybean. Genetic analysis showed that the saponin phenotype of PE1539 is controlled by a recessive mutation. We mapped the gene responsible for the phenotype of PE1539 and the mapped region included *Sg-3* (*Glyma.10G104700*). Further analysis of *Sg-3* in PE1539 using DNA sequencing revealed a single-nucleotide substitution in the exon (G804A), resulting in a premature stop codon; thus, PE1539 produced a PSPG box-truncated protein. Saponin phenotype analysis of the $F_2$ population—from a cross between wild-type Uram and PE1539—showed that the phenotype of saponin was cosegregated with the genotype of *Sg-3*. Quantitative real-time PCR showed reduced expression of *Sg-3* in PE1539 cells. Together, our data indicate that the saponin phenotype of PE1539 results from a mutation in *Sg-3*.

**Keywords:** soybean; saponin; saponin phenotype; *Sg-3*; UDP-glycosyltransferase



**Citation:** Ko, J.; Sundaramoorthy, J.; Tsukamoto, C.; Park, J.; Lee, J.; Jo, H.; Lee, J.-D.; Seo, H.S.; Song, J.T. Genetic and Molecular Characterization of a New EMS-Induced Mutant without the Third Glucose Moiety at the C-3 Sugar Chain of Saponin in *Glycine max* (L.) Merr. *Agronomy* **2022**, *12*, 1598. https://doi.org/10.3390/agronomy12071598

Academic Editor: Hengyou Zhang

Received: 7 June 2022
Accepted: 29 June 2022
Published: 1 July 2022



## 1. Introduction

Soybean (*Glycine max* (L.) Merr.) is a crop of immense economic value owing to its nutritional composition, with soybean seeds containing approximately 40% protein, 21% oil, and 34% carbohydrates on a dry basis [1]. In addition to the primary metabolites, soybean seeds contain phytochemicals, such as isoflavones, phytic acids, phytosterols, and saponins [2]. Saponins of soybean account for 0.2–0.5% of the whole soybean seed [3], and their levels are higher in the hypocotyls than in the cotyledons [4]. Soybean saponins have been a subject of study due to their influence on human health. Saponins are reported to have several effects, such as inhibition of HIV [5], suppression of cancer [6–8], antioxidant activity [9], and restriction of inflammation [10].

Soybean saponins are categorized into group A and 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) saponins (Figure S1) based on their structural differences. Group A saponins have sugar chains at the C-3 and C-22 positions of the soyasapogenol A aglycone structure, whereas DDMP saponins have a DDMP moiety at the C-22 position

and one sugar chain at the C-3 position of the soyasapogenol B aglycone structure. Soyasapogenol A has an additional hydroxyl group at the C-21 position, which is attached to CYP72A69 encoded by *Sg-5* (*Glyma.15g243300*) [11], whereas soyasapogenol B lacks a hydroxyl group at the C-21 position. However, soyasapogenol B interacts with UGT73B4, encoded by *Sg-9* (*Glyma.16G033700*), and conjugates with the DDMP moiety at the C-22 position [12]. This DDMP moiety is an easily degradable residue; therefore, DDMP saponins degrade to form structurally stable group B and E saponins depending on the structure of the C-22 position [13–15].

The diversity of soybean saponins is determined by the combination of sugars in their sugar chains. Previous studies have revealed several glycosyltransferases responsible for the glycosylation of soybean saponins. *Sg-1* (*Glyma.07G254600*) determines the type of terminal sugar at the C-22 position of group A saponins [16]. $Sg\text{-}1^{a}$ and $Sg\text{-}1^{b}$ alleles produce a xylosyltransferase (UGT73F4) and a glucosyltransferase (UGT73F2), respectively, whereas the recessive $sg\text{-}1^{0}$ allele is not able to produce any active glycosyltransferase. *Sg-3*, *Sg-4*, *GmSGT2*, and *GmSGT3* regulate glycosylation at the C-3 position of group A and DDMP saponins [17–20]. At the second position of the C-3 sugar chain, *Sg-4* (*Glyma.01G046300*) and *GmSGT2* (*Glyma.11G053400*) encode arabinosyltransferase (UGT73P10) and galactosyltransferase (UGT73P2), which conjugate arabinose and galactose, respectively. At the terminal of the C-3 sugar chain, *Sg-3* (*Glyma.10G104700*) and *GmSGT3* (*Glyma.08G181000*) determine the residue type. *Sg-3* encodes a glucosyltransferase (UGT91H9), and *GmSGT3* encodes a rhamnosyltransferase (UGT91H4).

A natural mutant with the recessive *sg-3* allele, *Glycine max* 'Mikuriya-ao', has been previously reported [18,20–22]. Mikuriya-ao is unable to produce Ab-αg saponin ($C_{67}H_{104}O_{33}$) and DDMP-αg saponin ($C_{54}H_{84}O_{22}$), which have a glucose moiety at the third sugar of the C-3 position. Instead, it accumulates Ab-γg saponin ($C_{61}H_{94}O_{28}$) and produces Ab-βg saponin ($C_{67}H_{104}O_{32}$) and DDMP-βg saponin ($C_{54}H_{84}O_{21}$) with a rhamnose residue at the third sugar of the C-3 sugar chain. Moreover, Mikuriya-ao has a large deletion in the genomic DNA around the *Sg-3* locus.

In this study, we isolated a mutant line, PE1539, which lacks saponins with glucose at the third sugar position of the C-3 sugar chain from the ethyl methanesulfonate (EMS)-induced mutant population of the soybean cultivar Pungsannamul. We found that the PE1539 phenotype is caused by a single recessive allele. The mapped region of the gene responsible for the PE1539 phenotype contains the *Sg-3* locus. A single-nucleotide polymorphism was detected in the coding region of the *Sg-3* gene of PE1539, and we designated this allele as *sg-3a*. The expression of *Sg-3* in PE1539 was significantly lower than that in wild-type Pungsannamul. The saponin phenotype of PE1539 and the genotype of *Sg-3* showed cosegregation, strongly indicating that the *sg-3a* allele of PE1539 was responsible for its saponin phenotype.

## 2. Materials and Methods

### *2.1. Plant Materials*

An EMS-induced mutant population was developed by treating the seeds of soybean cultivar Pungsannamul with 0.3% EMS [23]. The population progressed to the $M_4$ generation, and seeds were obtained from each of the 892 $M_3$ lines. The saponin phenotype of the seeds was analyzed using thin-layer chromatography (TLC), and a mutant line, PE1539, was isolated. The soybean cultivar Uram and PE1539 were crossed for a segregation analysis and physical mapping, and $F_2$ individuals were harvested from the crossing and used for cosegregation analysis. The plants were grown in the experimental fields of Kyungpook National University (Gunwi, 36°07′ N, 128°38′ E, Korea).

### *2.2. TLC of Saponin Composition*

To analyze the saponin composition of the Pungsannamul, Uram, PE1539, and $F_2$ populations, saponins were extracted from the hypocotyl of mature seeds. A TLC analysis was performed following the protocol described by Krishnamoorthy et al. [24] and Sun-

daramoorthy et al. [25] with slight modifications. Hypocotyl tissues were incubated in a 10-fold volume (*v*/*w*) of 80% (*v*/*v*) methanol (aqueous) at 25 °C for 12 h. Three microliters of extract was loaded onto TLC silica gel plates (Merck, Darmstadt, Germany) and air-dried, and then the TLC chamber was saturated with 50 mL of the lower phase of a mixed solution of chloroform, methanol, and water (65:35:10, *v*/*v*), and the dried TLC plates were developed in the TLC chamber at room temperature for 40 min. The plates were dried at 90 °C for 5 min and developed using 10% (*v*/*v*) $H_2SO_4$ for 10 min in a chamber. Saponins were colored by drying at 100 °C for 15 min.

### *2.3. Liquid Chromatography with Photodiode Array and Tandem Mass Spectrometry (LC-PDA-MS/MS) of Saponin Composition*

LC-PDA-MS/MS was performed to separate and identify saponins in PE1539 based on the protocols established by Takada et al. [26] and Son et al. [27] with a few modifications at each step. The analysis was conducted using a Prominence UFLC system (Shimadzu Corporation, Kyoto, Japan) with a C30 reversed-phase column (Develosil C30-UG-3, 2.0 mm I.D. × 250 mm, Nomura Chemical, Seto, Okayama, Japan) and a tandem mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, Santa Clara, CA, USA). Saponins were extracted by soaking the hypocotyl in a 10-fold volume of 80% aqueous ethanol (*v*/*v*). The extracted samples were diluted 10 times and analyzed by injecting 2 mL of diluted solution. Solvent A was acetonitrile with 0.1% (*v*/*v*) formic acid and solvent B was 0.1% (*v*/*v*) aqueous formic acid. The flow rate used for the analysis was 0.15 mL/min. The analysis began with 20% (*v*/*v*) solvent A and 80% solvent B, and the concentration of solvent A was increased to 65% (*v*/*v*) for 45 min. The concentration of solvent A was immediately increased to 100% (*v*/*v*) and maintained for 5 min. Subsequently, the solvent composition was recovered to 20% (*v*/*v*) of solvent A with 80% solvent B, which is the primary status for analysis, and maintained for 15 min. The saponins were detected using ultraviolet (UV) light at a wavelength of 205 nm. The tandem mass spectrometer was set in positive (+)-ion mode and automatic full-scan mode of mass-to-charge ratio (m/z) from 300 to 2000. The UV and MS spectral data were recorded and analyzed using the Xcalibur software, version 2.1 (Thermo Fisher Scientific, Santa Clara, CA, USA).

### *2.4. Extraction of Genomic DNA*

Genomic DNA from the Pungsannamul, Uram, and $F_2$ populations was extracted for physical mapping, sequencing, and cosegregation analyses. Genomic DNA samples from $F_2$ mutant individuals for physical mapping were isolated using the HiGene™ Genomic DNA Prep Kit (Biofact, Daejeon, Korea). Other genomic DNA samples were extracted using the cetyltrimethylammonium bromide (CTAB) DNA isolation method [28].

### *2.5. Bulked Segregant Analysis and Sequencing of Sg-3*

The same amount of genomic DNA from 20 $F_2$ mutant individuals was combined and analyzed using the Axiom® Soybean Genotyping Array (Affymetrix, Waltham, MA, USA). Bulked segregant analysis (BSA) was performed to detect the locus associated with the phenotype of PE1539 by detecting recombinations in $F_2$ mutant bulk data using Microsoft Excel. Primers were designed to verify the sequence of *Sg-3* in PE1539 cells (Table S1). Genomic DNA from Pungsannamul and PE1539 was used as a template for the PCR. The PCR mixture was 30 µL volume of mixture consisting of 1 µL of template DNA, 1 µL of 5 mM dNTP mix, 1 µL of 10 pmol/µL forward and reverse primers, 3 µL of 10X e-Taq Reaction Buffer (25 mM $MgCl_2$ mixed, Solgent, Daejeon, Korea), 22.9 µL of $ddH_2O$ and 0.1 µL of Solg™ e-Taq DNA Polymerase (5 U/µL, Solgent, Daejeon, Korea). The PCR reaction was proceeded with 5 min of initial denaturation at 95 °C, 30 s of denaturation at 95 °C, 30 s of annealing at 56 °C, 1 min of elongation at 72 °C and 5 min of final elongation at 72 °C. The PCR products were purified using the HiGene™ Gel & PCR Purification System (Biofact, Daejeon, Korea) and sequenced (Solgent, Daejeon, Korea).

The sequenced data of Pungsannamul and PE1539 were compared using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on 12 January 2021).

### *2.6. Multiple Alignment of UDP-glycosyltransferases*

The amino acid sequences of UDP-glycosyltransferases were aligned. Amino acid sequences were collected from NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/, accessed on 3 April 2022) and analyzed using ClustalW (http://www.genome.jp/tools-bin/clustalw, accessed on 3 April 2022).

### *2.7. Segregation and Cosegregation Analysis of Sg-3*

The $F_2$ individuals crossing Uram and PE1539 were used for segregation and cosegregation of the saponin phenotype and *Sg-3* genotype. The $F_2$ population consisted of 173 individuals. Hypocotyl tissues were used for the TLC analysis to investigate the saponin phenotype. A derived cleaved amplified sequence (dCAPS) marker was designed to target a 1-base pair substitution at the 804-nucleotide position, and the amplicon of the mutant contained the restriction site of *Mbo*I (Table S1). The composition of PCR mixture for the dCAPS analysis was the same as that of sequencing analysis. The PCR reaction was proceeded with 5 min of initial denaturation at 95 °C, 30 s of denaturation at 95 °C, 30 s of annealing at 56 °C, 30 s of elongation at 72 °C and 5 min of final elongation at 72 °C. The PCR products were digested with *Mbo*I (Takara Bio, Kusatsu, Shiga, Japan) and separated on 2% agarose gel using electrophoresis. The segregation and cosegregation ratios were calculated using the chi-square test.

### *2.8. Expression Analysis of Sg-3*

Hypocotyl tissue samples were collected from developing seeds 50 days after flowering, and total RNA was extracted using the QIAzol® Lysis Reagent (Qiagen, Germantown, MD, USA). The DNA remaining in the extracted RNA samples was removed using recombinant DNase I (Takara Bio, Kusatsu, Shiga, Japan). The cDNA samples were synthesized using the DiaStar™ RT Kit (Solgent, Daejeon, Korea). The expression of *Sg-3* was quantified using primers designed for quantitative real-time PCR (qRT-PCR) (Table S1) and WizPure™ qPCR Master (SYBR; Wizbiosolutions, Seongnam, Korea) with a LightCycler® 480 Real-Time PCR System (Roche, Basel, Switzerland).

## 3. Results and Discussion

### *3.1. Phenotypic Characterization of a New EMS-Induced Mutant*

To isolate saponin mutants in an EMS-induced population of Pungsannamul, TLC analysis was conducted. TLC analysis of PE1539 showed the accumulation of Ab-γg and a deficiency of Ab-αg compared to that in Pungsannamul and Uram (Figure 1a). LC-PDA-MS/MS analysis was performed to elucidate the saponin composition of PE1539. The LC-PDA-MS/MS results revealed that PE1539 lacked Ab-αg saponin, DDMP-αg saponin, and B-αg, which have glucose as the third sugar at the C-3 position. Instead, PE1539 had increased amounts of Ab-γg, Ab-βg, and DDMP-βg saponins (Figure 1b,c). The investigated saponin phenotype of PE1539 is identical to the previously reported mutant Mikuriya-ao [20]; therefore, we hypothesized that the phenotype of PE1539 is caused by the disabled Sg-3 UDP-glucosyltransferase. The accumulation of Ab-γg saponins can be explained by the fact that the substrates of Sg-3 remained. The increased amounts of Ab-βg saponin and DDMP-βg saponin could be the result of an increased likelihood of GmSGT3 interacting with the precursor saponins (Figure S2).

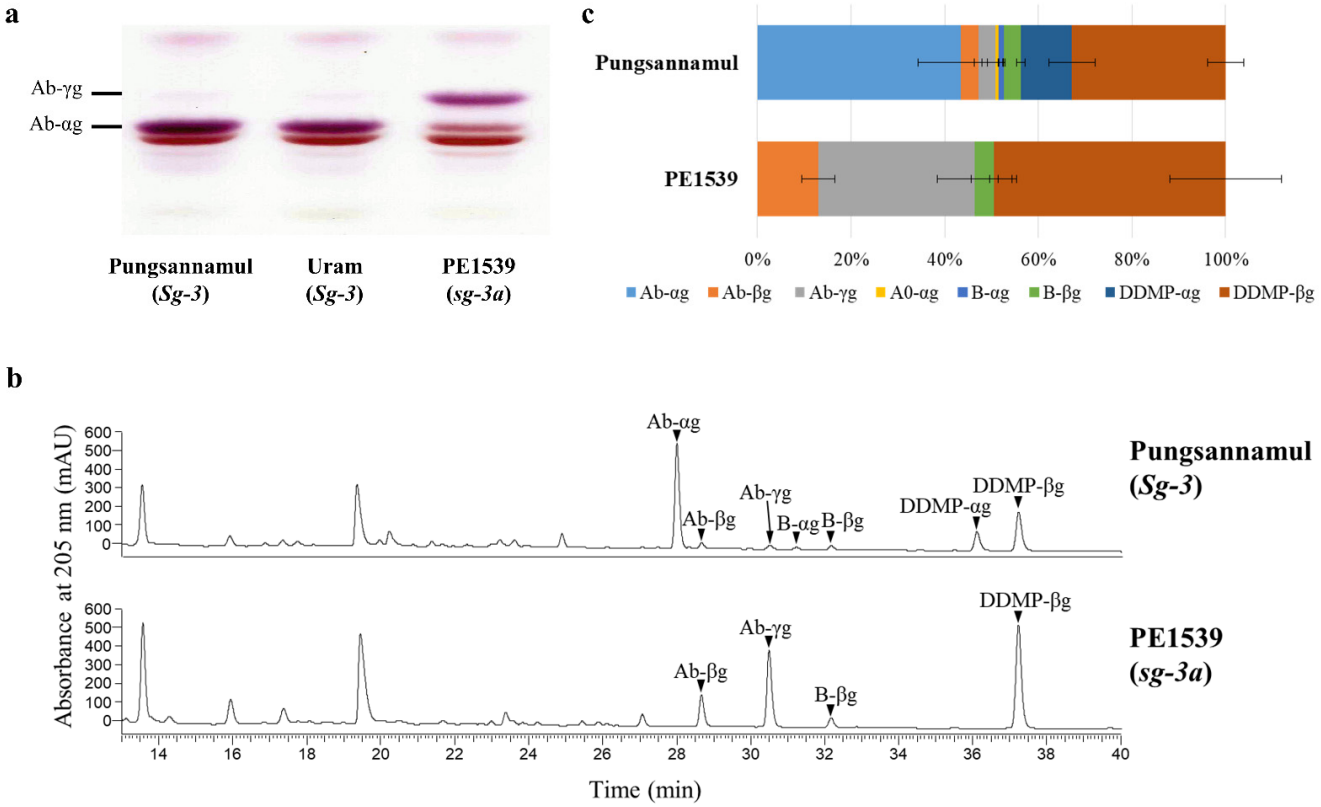


**Figure 1.** Phenotypes of saponin in wild-type soybeans and PE1539. (**a**) Thin-layer chromatography (TLC) patterns of Pungsannamul, Uram, and PE1539. The TLC pattern of PE1539 shows accumulation of Ab-γg and deficiency of Ab-αg in comparison with the typical wild-type saponin TLC patterns of Pungsannamul and Uram. (**b**) Liquid chromatography with photodiode array and tandem mass spectrometry (LC-PDA-MS/MS) results for Pungsannamul and PE1539. Black triangles indicate the peaks at corresponding retention times (RTs) of saponins. Ab-αg (RT = 28 min), Ab-βg (RT = 28.65 min), Ab-γg (RT = 30.48 min), B-αg (RT = 31.20 min), B-βg (RT = 32.15 min), DDMP-αg (RT = 36.10 min), and DDMP-βg (RT = 37.20 min). (**c**) Proportions of individual saponins to total saponins in Pungsannamul and PE1539. Each type of saponin is distinguished by a different color. The error bars indicate the standard deviation.

### *3.2. Physical Mapping and Molecular Characterization of PE1539*

For the segregation analysis, a cross was made between the soybean cultivar Uram and PE1539. The cross obtained 173 $F_2$ individuals, which was segregated into 132 individuals of the typical saponin phenotype (Uram type) and 41 individuals of the mutant phenotype (PE1539 type). According to the chi-square test, the segregation ratio was 3:1, which means that the phenotype of PE1539 is controlled by a single recessive allele (Table 1).

**Table 1.** Segregation and cosegregation analysis of $F_2$ population.

| Parents and Their Progenies | Saponinphenotype | Observed | (Expected) | $\chi^2$ Value | Probability * | Genotype † | Observed | (Expected) | $\chi^2$ Value | Probability * |
|---|---|---|---|---|---|---|---|---|---|---|
| P1: Uram | Typical | | | | | W | | | | |
| P2: PE1539 | Ab-γg rich | | | | | M | | | | |
| $F_2$ population | Typical | 132 | (129.75) | 0.156 | 0.69 | W | 41 | (43.25) | 0.468 | 0.79 |
| | Ab-γg rich | 41 | (43.25) | | | H | 91 | (86.5) | | |
| | | | | | | M | 41 | (43.25) | | |

* Not significant ($p > 0.05$). † W, wildtype homozygote; H, heterozygote; M, mutant homozygote.

To determine the candidate gene responsible for the PE1539 phenotype, we performed BSA by applying the Axiom® soybean genotyping array on bulked genomic DNA of 20 mutant $F_2$ individuals from the cross between Uram and PE1539. The locus was mapped to a 45-Mb region between Affx-89235864 and Affx-89101343 on chromosome 10 (Figure 2a). We anticipated that the saponin phenotype of PE1539 was a result of a dysfunctional Sg-3 because the mapped region included the *Sg-3* locus. Therefore, the *Sg-3* gene of PE1539 was sequenced to identify any polymorphisms. The sequencing results revealed that a single-nucleotide substitution occurred from G to A at the 804 nucleotide position in the exon of *Sg-3* in PE1539 compared to that in Pungsannamul. This mutation results in a premature stop codon. We designated this mutated allele *sg-3a* (Figure 2b).

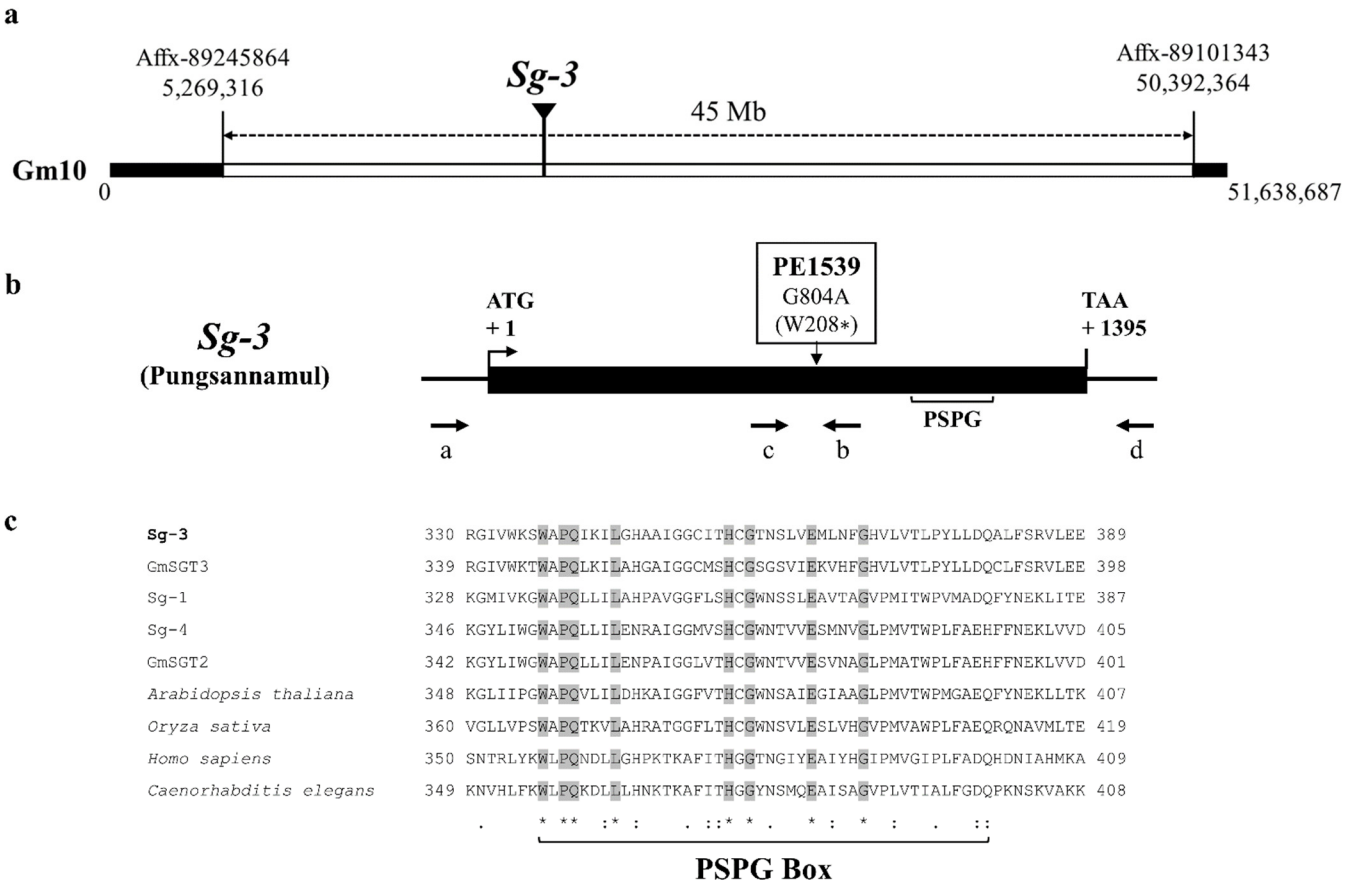


**Figure 2.** Genomic characterization of PE1539. (**a**) Physical map of *Sg-3*. A candidate region of 45 Mb based on the BSA result is shown on chromosome 10 of soybean. Twenty $F_2$ mutant individuals from the cross of Uram and PE1539 were used for the physical mapping. Indicated physical positions are based on the Williams 82 version 4 assembly. (**b**) Polymorphism of *Sg-3* (*Glyma.10G104700*) in Pungsannamul and PE1539. The *sg-3a* allele of PE1539 has a single-nucleotide substitution causing the generation of a premature stop codon before the plant secondary product glycosyltransferase (PSPG) box. The black box indicates exon of *Sg-3.* The region marked with black brackets indicates the PSPG box. Arrows indicate the location and direction of primers used for sequencing. a, Gm-Sg3-F1; b, Gm-Sg3-R1; c, Gm-Sg3-F2; d, Gm-Sg3-R2. (**c**) Multiple amino acid sequence alignment of UDP-glycosyltransferases. Amino acid sequences of UDP-glycosyltransferases from different species were aligned together. The region marked with black brackets indicates PSPG box. The regions marked with gray highlighting and asterisks indicate identical amino acids. Colons indicate conserved substitutions. Dots indicate semi-conserved substitutions. GenBank accession numbers of UGTs: Sg-3, NP_001348424.1; GmSGT3, NP_001240857.1; Sg-1, NP_001237242.2; Sg-4, XP_003517967.1; GmSGT2, NP_001304384.2; *Arabidopsis thaliana*, NP_001189528.1; *Oryza sativa*, XP_015617717.1; *Homo sapiens*, NP_001068.1; *Caenorhabditis elegans*, NP_497918.1.

Among the UDP-glycosyltransferases, a conserved motif named plant secondary product glycosyltransferase (PSPG) box is observed at the C-terminal [29]. This conserved PSPG motif is known to play an important role in recognizing the sugar donor and receptor in the reaction [30,31]. The PSPG boxes are conserved in Sg-3, GmSGT3, Sg-1, Sg-4, and GmSGT2 of soybean, and in UGTs of *Arabidopsis thaliana*, *Oryza sativa*, *Homo sapiens*, and *Caenorhabditis elegans* (Figure 2c). Considering that the mutated protein translated from *sg-3a* lacks the PSPG box, it is expected to lose its original function as a UDP-glucosyltransferase.

### 3.3. Cosegregation of the Sg-3 Polymorphism with the Saponin Phenotype

To determine the role of the *sg-3a* mutant allele in the saponin phenotype of PE1539, we isolated genomic DNA of $F_2$ individuals from the cross of Uram and PE1539 and analyzed cosegregation with the dCAPS marker. The mutant allele *sg-3a* produced digestible amplicons for the *Mbo*I restriction enzyme. Wild-type homozygous individuals showed one 245-bp band, mutant homozygous individuals showed one 197-bp band, and heterozygous individuals showed both 245-bp and 197-bp bands (Figure 3). Of the 173 individuals analyzed, 41 had the wild-type genotype, 90 had the heterozygous genotype, and 41 had the mutant genotype. The genotype of the population was segregated at a 1:2:1 ratio (Table 1). All progenies with the saponin phenotype of PE1539 had a homozygous genotype of *sg-3a*, and the other progenies with the typical saponin phenotype of Pungsannamul were found to have either heterozygous or homozygous genotypes of *Sg-3*. Therefore, we conclude that the *sg-3a* allele is a new recessive allele responsible for the saponin phenotype of PE1539.

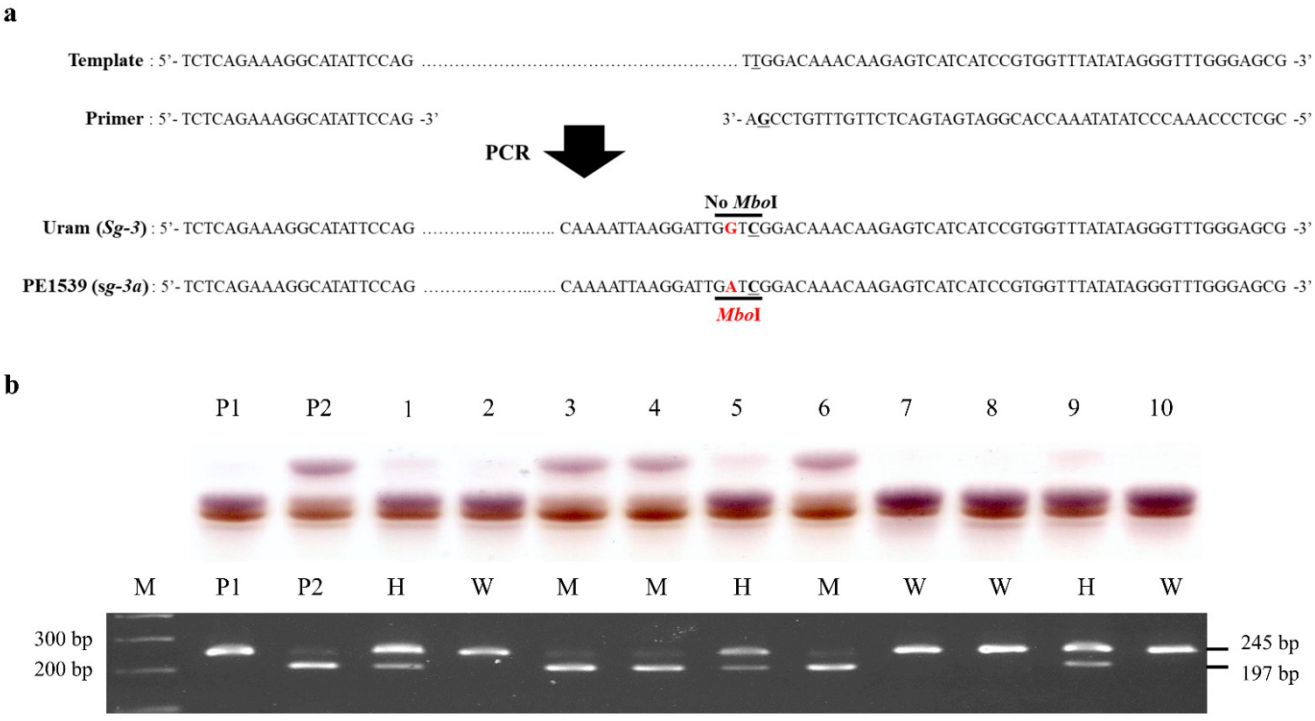


**Figure 3.** Cosegregation analysis of the $F_2$ population. (**a**) Schematic diagram of the dCAPS marker for *sg-3a* polymorphism. Reverse primer of dCAPS marker provides *Mbo*I restriction site only for the amplicon of *sg-3a* template. (**b**) Cosegregation analysis of the $F_2$ population based on a TLC analysis and dCAPS marker. The dCAPS result shows the cosegregation of *Sg-3* polymorphism and saponin phenotype. The PCR amplicon was 245-bp-long, and the amplicons from *sg-3a* allele were digested into 197-bp-long DNA fragments, whereas the amplicons from *Sg-3* were not digested. P1, Uram; P2, PE1539; #1–#10, individuals of $F_2$ population; M, a 1-kb marker; W, wild-type homozygote; H, heterozygote; M, mutant homozygote.

### 3.4. Expression Analysis of *Sg-3*

To investigate the influence of the mutation on the expression of *Sg-3* in PE1539, we monitored the expression levels of *Sg-3* in PE1539 and wild-type Pungsannamul using qRT-PCR. The expression of *Sg-3* was significantly lower in PE1539 cells than in Pungsannamul cells (Figure 4a). We hypothesized that the decreased expression of PE1539 was a result of nonsense-mediated mRNA decay (NMD) derived from the premature stop codon in the *sg-3a* allele. The NMD pathway is a highly conserved pathway in eukaryotes that recognizes and removes mRNAs with premature stop codons. Therefore, it avoids the risk of accumulating truncated proteins that may be potentially harmful [32]. In plants, abnormally long 3′-UTRs (longer than 300 bp) work as *cis*-elements of NMD, and the mRNA is finally degraded by the NMD mechanism [33–35]. For *sg-3a* in PE1539, the length of the mutated 3′-UTR was extended by 591-bp nucleotides after the premature termination codon (Figure 4b). Hence, we speculated that the nonsense mutation in *sg-3a* triggered NMD and reduced the expression of *sg-3a* in PE1539.

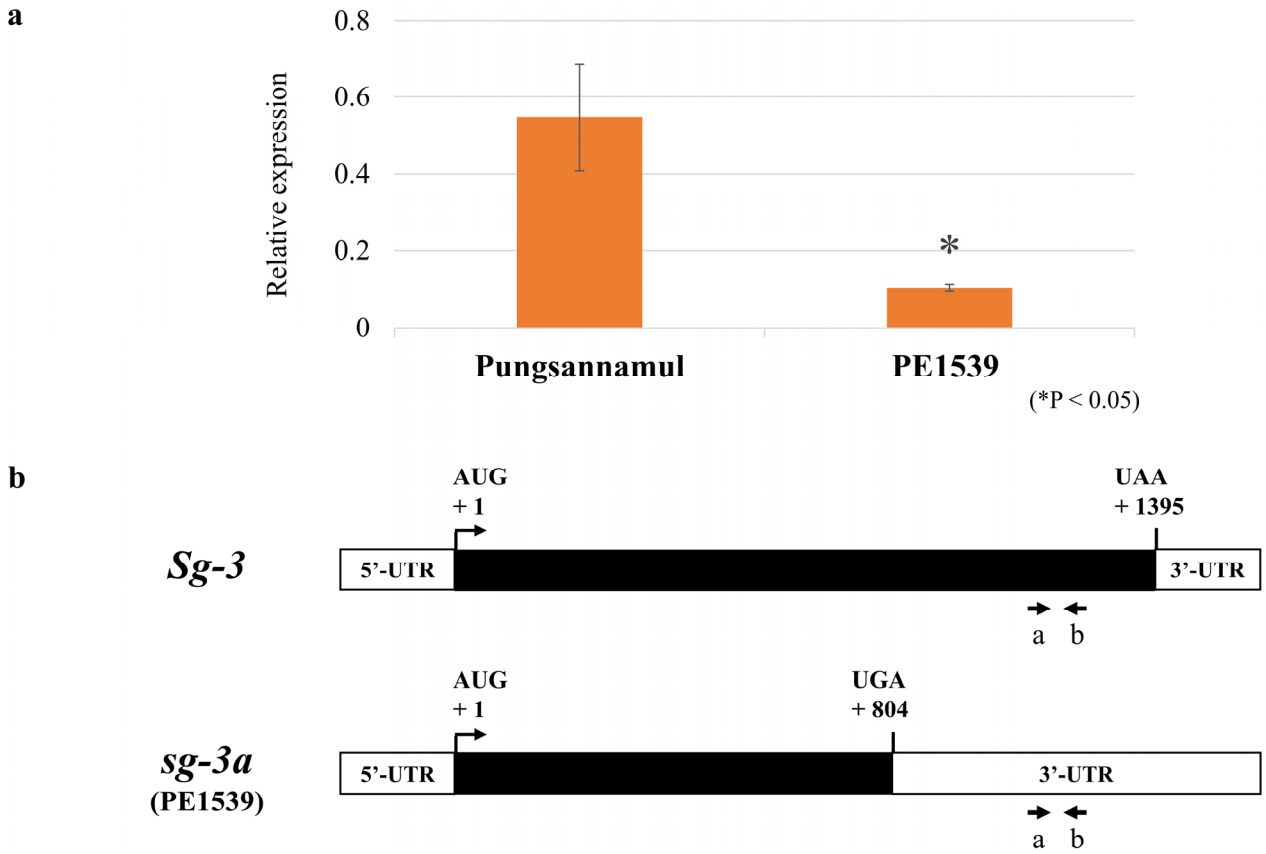


**Figure 4.** Expression of *Sg-3* in PE1539. (**a**) Relative transcript levels of *Sg-3* in Pungsannamul and *sg-3a* in PE1539. The expression of *sg-3a* is significantly reduced in PE1539. Asterisk indicates reduced expression determined using Student's *t* test (* $p < 0.05$). The error bars indicate the standard deviation. (**b**) Schematic diagram of *Sg-3* and *sg-3a* mRNA. Note that 3′-UTR is expanded due to the premature stop codon in PE1539. Arrows a and b indicate location and direction of primers used for qRT-PCR. The black boxes indicate exons. The white boxes indicate 5′-UTR and 3′-UTR. 5′-UTR, 5′ untranslated region; 3′-UTR, 3′ untranslated region.

## 4. Conclusions

The study revealed that the lack of saponins with a third glucose moiety at the C-3 sugar chain in PE1539 was caused by a new allele of *Sg-3*, which was designated as *sg-3a*. The new allele had a single-nucleotide polymorphism, which resulted in a premature stop codon. The cosegregation analysis showed the recessiveness of *sg-3a* to the wild-type *Sg-3*

allele and its relationship with the saponin phenotype of PE1539. These results suggested that dysfunctional Sg-3 protein was produced, and *Sg-3* gene expression was inhibited. We concluded that the recessive *sg-3a* allele was responsible for the absence of saponins with a third glucose moiety at the C-3 sugar chain in PE1539.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12071598/s1, Figure S1: Structure and nomenclature of soybean saponins; Figure S2: Conjugation of third sugar moiety of soybean saponin; Table S1: List of primers for sequencing, dCAPS, and qRT-PCR analyses.

**Author Contributions:** Conceptualization, J.K., J.S. and J.T.S.; data curation, J.K. and J.S.; formal analysis, J.K., J.S., C.T. and J.P.; investigation, J.K., J.S. and J.P.; methodology, J.K., J.S. and J.T.S.; project administration, J.T.S.; resources, J.-D.L. and J.T.S.; supervision, J.T.S.; writing—original draft, J.K.; writing—review and editing, J.K., J.L., H.J., J.-D.L., H.S.S. and J.T.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) fund by the Ministry of Education (No. NRF-2016R1D1A1B01007409).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Medic, J.; Atkinson, C.; Hurburgh, C.R. Current knowledge in soybean composition. *J. Am. Oil Chem. Soc.* **2014**, *91*, 363–384. [CrossRef]
2. Isanga, J.; Zhang, G.-N. Soybean bioactive components and their implications to health—A review. *Food Rev. Int.* **2008**, *24*, 252–276. [CrossRef]
3. Kim, S.-L.; Berhow, M.A.; Kim, J.-T.; Chi, H.-Y.; Lee, S.-J.; Chung, I.-M. Evaluation of soyasaponin, isoflavone, protein, lipid, and free sugar accumulation in developing soybean seeds. *J. Agric. Food Chem.* **2006**, *54*, 10003–10010. [CrossRef]
4. Shimoyamada, M.; Harada, K.; Okubo, K. Saponin composition in developing soybean seed (*Glycine max* (L.) Merrill, cv. Mikuriyaao). *Agric. Biol. Chem.* **1991**, *55*, 1403–1405. [CrossRef]
5. Nakashima, H.; Okubo, K.; Honda, Y.; Tamura, T.; Matsuda, S.; Yamamoto, N. Inhibitory effect of glycosides like saponin from soybean on the infectivity of HIV in vitro. *AIDS* **1989**, *3*, 655–658. [CrossRef]
6. Ellington, A.A.; Berhow, M.; Singletary, K.W. Induction of macroautophagy in human colon cancer cells by soybean B-group triterpenoid saponins. *Carcinogenesis* **2005**, *26*, 159–167. [CrossRef]
7. Ellington, A.A.; Berhow, M.A.; Singletary, K.W. 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* **2006**, *27*, 298–306. [CrossRef]
8. Salyer, J.; Eswaranandam, S.; Lee, S.-O. Soyasaponin I, III, and soyasapogenol B inhibit proliferation and modulate PKC expression in caco-2 human colon cancer cells. *J. Food Res.* **2013**, *2*, 81. [CrossRef]
9. Zhu, L.; Zhang, M.; Liu, X.; Liu, H.; He, Y.; Wang, B.; Ma, T. Evaluation of in vitro antioxidant activities of soyasaponins from soy hypocotyls in human HepG2 cell line. *Chem. Pap.* **2017**, *71*, 653–660. [CrossRef]
10. Zha, L.; Chen, J.; Sun, S.; Mao, L.; Chu, X.; Deng, H.; Cai, J.; Li, X.; Liu, Z.; Cao, W. Soyasaponins can blunt inflammation by inhibiting the reactive oxygen species-mediated activation of PI3K/Akt/NF-kB pathway. *Proc. Nalt. Acad. Sci. USA* **2014**, *9*, e107655. [CrossRef]
11. Yano, R.; Takagi, K.; Takada, Y.; Mukaiyama, K.; Tsukamoto, C.; Sayama, T.; Kaga, A.; Anai, T.; Sawai, S.; Ohyama, K. Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. *Plant J.* **2017**, *89*, 527–539. [CrossRef]
12. Sundaramoorthy, J.; Park, G.T.; Komagamine, K.; Tsukamoto, C.; Chang, J.H.; Lee, J.D.; Kim, J.H.; Seo, H.S.; Song, J.T. Biosynthesis of DDMP saponins in soybean is regulated by a distinct UDP-glycosyltransferase. *New Phytol.* **2019**, *222*, 261–274. [CrossRef]
13. Kudou, S.; Tonomura, M.; Tsukamoto, C.; Uchida, T.; Sakabe, T.; Tamura, N.; Okubo, K. Isolation and structural elucidation of DDMP-conjugated soyasaponins as genuine saponins from soybean seeds. *Biosci. Biotechnol. Biochem.* **1993**, *57*, 546–550. [CrossRef]
14. Kudou, S.; Tonomura, M.; Tsukamoto, C.; Shimoyamada, M.; Uchida, T.; Okubo, K. Isolation and structural elucidation of the major genuine soybean saponin. *Biosci. Biotechnol. Biochem.* **1992**, *56*, 142–143. [CrossRef]

15. Chitisankul, W.T.; Shimada, K.; Omizu, Y.; Uemoto, Y.; Varanyanond, W.; Tsukamoto, C. Mechanism of DDMP-saponin degradation and maltol production in soymilk preparation. *LWT* **2015**, *64*, 197–204. [CrossRef]
16. Sayama, T.; Ono, E.; Takagi, K.; Takada, Y.; Horikawa, M.; Nakamoto, Y.; Hirose, A.; Sasama, H.; Ohashi, M.; Hasegawa, H. The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell* **2012**, *24*, 2123–2138. [CrossRef]
17. Shibuya, M.; Nishimura, K.; Yasuyama, N.; Ebizuka, Y. Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. *FEBS Lett.* **2010**, *584*, 2258–2264. [CrossRef]
18. Takada, Y.; Tayama, I.; Sayama, T.; Sasama, H.; Saruta, M.; Kikuchi, A.; Ishimoto, M.; Tsukamoto, C. Genetic analysis of variations in the sugar chain composition at the C-3 position of soybean seed saponins. *Breed. Sci.* **2012**, *61*, 639–645. [CrossRef]
19. Takagi, K.; Yano, R.; Tochigi, S.; Fujisawa, Y.; Tsuchinaga, H.; Takahashi, Y.; Takada, Y.; Kaga, A.; Anai, T.; Tsukamoto, C. Genetic and functional characterization of Sg-4 glycosyltransferase involved in the formation of sugar chain structure at the C-3 position of soybean saponins. *Phytochemistry* **2018**, *156*, 96–105. [CrossRef]
20. Yano, R.; Takagi, K.; Tochigi, S.; Fujisawa, Y.; Nomura, Y.; Tsuchinaga, H.; Takahashi, Y.; Takada, Y.; Kaga, A.; Anai, T. Isolation and characterization of the soybean *Sg-3* gene that is involved in genetic variation in sugar chain composition at the C-3 position in soyasaponins. *Plant Cell Physiol.* **2018**, *59*, 797–810. [CrossRef]
21. Tsukamoto, C.; Kikuchi, A.; Kudou, S.; Harada, K.; Kitamura, K.; Okubo, K. Group A acetyl saponin-deficient mutant from the wild soybean. *Phytochemistry* **1992**, *31*, 4139–4142. [CrossRef]
22. Tsukamoto, C.; Kikuchi, A.; Harada, K.; Kitamura, K.; Okubo, K. Genetic and chemical polymorphisms of saponins in soybean seed. *Phytochemistry* **1993**, *34*, 1351–1356. [CrossRef]
23. Chae, J.-H.; Dhakal, K.H.; Asekova, S.; Song, J.T.; Lee, J.-D. Variation of fatty acid composition in soybean 'Pungsannamul' mutation population from EMS treatment. *Curr. Res. Agric. Life Sci.* **2013**, *31*, 45–50.
24. Krishnamurthy, P.; Tsukamoto, C.; Yang, S.H.; Lee, J.D.; Chung, G. An improved method to resolve plant saponins and sugars by TLC. *Chromatographia* **2012**, *75*, 1445–1449. [CrossRef]
25. Sundaramoorthy, J.; Park, G.T.; Mukaiyama, K.; Tsukamoto, C.; Chang, J.H.; Lee, J.-D.; Kim, J.H.; Seo, H.S.; Song, J.T. Molecular elucidation of a new allelic variation at the *Sg-5* gene associated with the absence of group A saponins in wild soybean. *Proc. Nalt. Acad. Sci. USA* **2018**, *13*, e0192150. [CrossRef]
26. Takada, Y.; Sasama, H.; Sayama, T.; Kikuchi, A.; Kato, S.; Ishimoto, M.; Tsukamoto, C. Genetic and chemical analysis of a key biosynthetic step for soyasapogenol A, an aglycone of group A saponins that influence soymilk flavor. *Theor. Appl. Genet.* **2013**, *126*, 721–731. [CrossRef]
27. Son, H.; Mukaiyama, K.; Omizu, Y.; Tsukamoto, C. A fragmentation database of soyasaponins by liquid chromatography with a photodiode array detector and tandem mass spectrometry. *Anal. Sci. Technol.* **2021**, *34*, 172–179. [CrossRef]
28. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
29. Mackenzie, P.I.; Owens, I.S.; Burchell, B.; Bock, K.W.; Bairoch, A.; Belanger, A.; Fournel-Gigleux, S.; Green, M.; Hum, D.W.; Iyanagi, T. The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* **1997**, *7*, 255–269. [CrossRef]
30. Gachon, C.M.; Langlois-Meurinne, M.; Saindrenan, P. Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends Plant Sci.* **2005**, *10*, 542–549. [CrossRef]
31. Wang, X. Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett.* **2009**, *583*, 3303–3309. [CrossRef]
32. Peccarelli, M.; Kebaara, B.W. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot. Cell* **2014**, *13*, 1126–1135. [CrossRef]
33. Muhlrad, D.; Parker, R. Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **1999**, *5*, 1299–1307. [CrossRef]
34. Kertesz, S.; Kerenyi, Z.; Merai, Z.; Bartos, I.; Palfy, T.; Barta, E.; Silhavy, D. Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res.* **2006**, *34*, 6147–6157. [CrossRef]
35. Schwartz, A.; Komarova, T.; Skulachev, M.; Zvereva, A.; Dorokhov, Y.L.; Atabekov, J. Stability of plant mRNAs depends on the length of the 3′-untranslated region. *Biochemistry* **2006**, *71*, 1377–1384. [CrossRef]