PLOS ONE

RESEARCH ARTICLE

Molecular elucidation of a new allelic variation at the *Sg-5* gene associated with the absence of group A saponins in wild soybean

Jagadeesh Sundaramoorthy[1], Gyu Tae Park[1], Kyosuke Mukaiyama[2], Chigen Tsukamoto[2], Jeong Ho Chang[3], Jeong-Dong Lee[1], Jeong Hoe Kim[4], Hak Soo Seo[5], Jong Tae Song[1]*

1 School of Applied Biosciences, Kyungpook National University, Daegu, Republic of Korea, **2** Faculty of Agriculture, Iwate University, Morioka, Iwate, Japan, **3** Department of Biology Education, Kyungpook National University, Daegu, Republic of Korea, **4** Department of Biology, Kyungpook National University, Daegu, Republic of Korea, **5** Department of Plant Science, Seoul National University, Seoul, Republic of Korea

* jtsong68@knu.ac.kr

Citation: Sundaramoorthy J, Park GT, Mukaiyama K, Tsukamoto C, Chang JH, Lee J-D, et al. (2018) Molecular elucidation of a new allelic variation at the *Sg-5* gene associated with the absence of group A saponins in wild soybean. PLoS ONE 13 (1): e0192150. https://doi.org/10.1371/journal.pone.0192150

Editor: David A Lightfoot, College of Agricultural Sciences, UNITED STATES

Received: December 8, 2017

Accepted: January 17, 2018

Published: January 30, 2018

Data Availability Statement: Relevant data are within the paper and in GenBank accession file in NCBI database (accession number: MF624839).

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

Competing interests: The authors have declared that no competing interests exist.

Abstract

In soybean, triterpenoid saponin is one of the major secondary metabolites and is further classified into group A and DDMP saponins. Although they have known health benefits for humans and animals, acetylation of group A saponins causes bitterness and gives an astringent taste to soy products. Therefore, several studies are being conducted to eliminate acetylated group A saponins. Previous studies have isolated and characterized the *Sg-5* (*Glyma.15g243300*) gene, which encodes the cytochrome P450 72A69 enzyme and is responsible for soyasapogenol A biosynthesis. In this study, we elucidated the molecular identity of a novel mutant of *Glycine soja*, 'CWS5095'. Phenotypic analysis using TLC and LC-PDA/MS/MS showed that the mutant 'CWS5095' did not produce any group A saponins. Segregation analysis showed that the absence of group A saponins is controlled by a single recessive allele. The locus was mapped on chromosome 15 (4.3 Mb) between Affx-89193969 and Affx-89134397 where the previously identified *Glyma.15g243300* gene is positioned. Sequence analysis of the coding region for the *Glyma.15g243300* gene revealed the presence of four SNPs in 'CWS5095' compared to the control lines. One of these four SNPs (G1127A) leads to the amino acid change Arg376Lys in the EXXR motif, which is invariably conserved among the CYP450 superfamily proteins. Co-segregation analysis showed that the missense mutation (Arg376Lys) was tightly linked with the absence of group A saponins in 'CWS5095'. Even though Arg and Lys have similar chemical features, the 3D modelled protein structure indicates that the replacement of Arg with Lys may cause a loss-of-function of the Sg-5 protein by inhibiting the stable binding of a heme cofactor to the CYP72A69 apoenzyme.

Introduction

Saponins are glycosylated compounds that are widely distributed in plants, and have different biological and pharmaceutical properties [1,2]. In addition, saponins are structurally diverse, for example triterpenoid and steroidal saponins [3]. Triterpenoid saponins are widely distributed in higher plants [1,2]. They are composed of triterpene aglycone, with one or more sugar chains. Biosynthesis of saponins is initialized from isopentenyl pyrophosphate in the mevalonate pathway [3]. Triterpene aglycones are derived from the 30-carbon linear 2,3-oxidosqualene precursor. In the first step of saponin synthesis, 2,3-oxidosqualene is cyclized by oxidosqualene cyclases (OSCs) to produce polycyclic triterpene [4,5]. After cyclization of the basic triterpene backbone, the backbone is oxidized to produce a hydrophobic aglycone called sapogenin. The oxidization step is catalyzed by cytochrome P450 (CYP450s) mono-oxygenases [6]. The next step is to synthesize saponins via *O*-glycosylation of the aglycones. Studies on triterpenoid saponins have been undertaken on crops and medicinal plants due to their important commercial uses in cosmetic and pharmaceutical industries. However, the biological roles of triterpenoid saponins in plants remain underexplored.

Soybean seeds are abundant in high-quality proteins and fats. In addition, triterpenoid saponins are the major components of mature seeds [7,8]. Soybean seeds on average contain 0.2–0.5% of saponin, although this value can widely vary depending on cultivars, degree of maturity, and growing locations [9–11]. In the well-known medicinal plant *Panax ginseng*, saponins are abundant in roots [12]. However, soybean saponins are rich in hypocotyls rather than in cotyledons and other plant parts [13,14]. Szakiel et al. [15] reported that oleanolic acid glycoside, a type of saponins in *Calendula officinalis* were synthesized in the cytoplasm and subsequently transported through the vacuolar membrane and finally accumulated in the vacuole. However, the site of saponin biosynthesis in soybean is still unknown.

Soybean saponins are mainly classified into group A and 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) saponins, which are commonly known as bisdemoside and monodesmoside saponins, respectively. Group A saponins contain two sugar chains at the C-3 and C-22 hydroxyl groups on the triterpenoid aglycone (soyasapogenol A; 3β, 21β, 22β, 24-tetrahydroxyolean-12-ene) (Fig 1) [10,16,17]. DDMP saponins contain a single sugar chain at the C-3 hydroxyl position of aglycone (soyasapogenol B; 3β, 22β, 24-trihydroxyolean-12-ene) and a DDMP moiety at the C-22 position (Fig 1) [10,16,17]. DDMP saponins are unstable molecules, which then degrade during food processing and the saponin extraction process into group B and group E saponins after hydrolysis (S1 Fig) [18]. In addition, the C-21 hydroxyl group is specific to soyasapogenol A, which distinguishes it from soyasapogenol B (Fig 1).

Diverse health benefits of soybean saponins have been reported. Saponins showed the antiviral activity against HIV, antioxidant activity, cholesterol-lowering activity, growth inhibition of tumor cell lines and anticancer activity through their anti-inflammatory activity [19–23]. Although soybean saponins are well known for their health benefits, the terminal acetylated sugar at the C-22 position of group A saponins in soybean seeds causes bitterness and astringent taste in soy products [7,24]. Therefore, several studies have been focused on removing the astringent taste of soy foods products through genetic identification and modification of genes involved in the biosynthesis of saponins [8,17,25,26].

In previous studies, many UDP-glycosyltransferase (UGT) genes, namely the *Sg-1*, *Sg-3*, *Sg-4*, *GmSGT2*, and *GmSGT3* involved in soybean saponin biosynthesis, have been identified [7,16,27–31]. Cytochrome P450 (CYP) 72A subfamily proteins are involved in the legume triterpenoid saponin biosynthesis [32]. Very recently, a hydroxylase gene, *Sg-5*

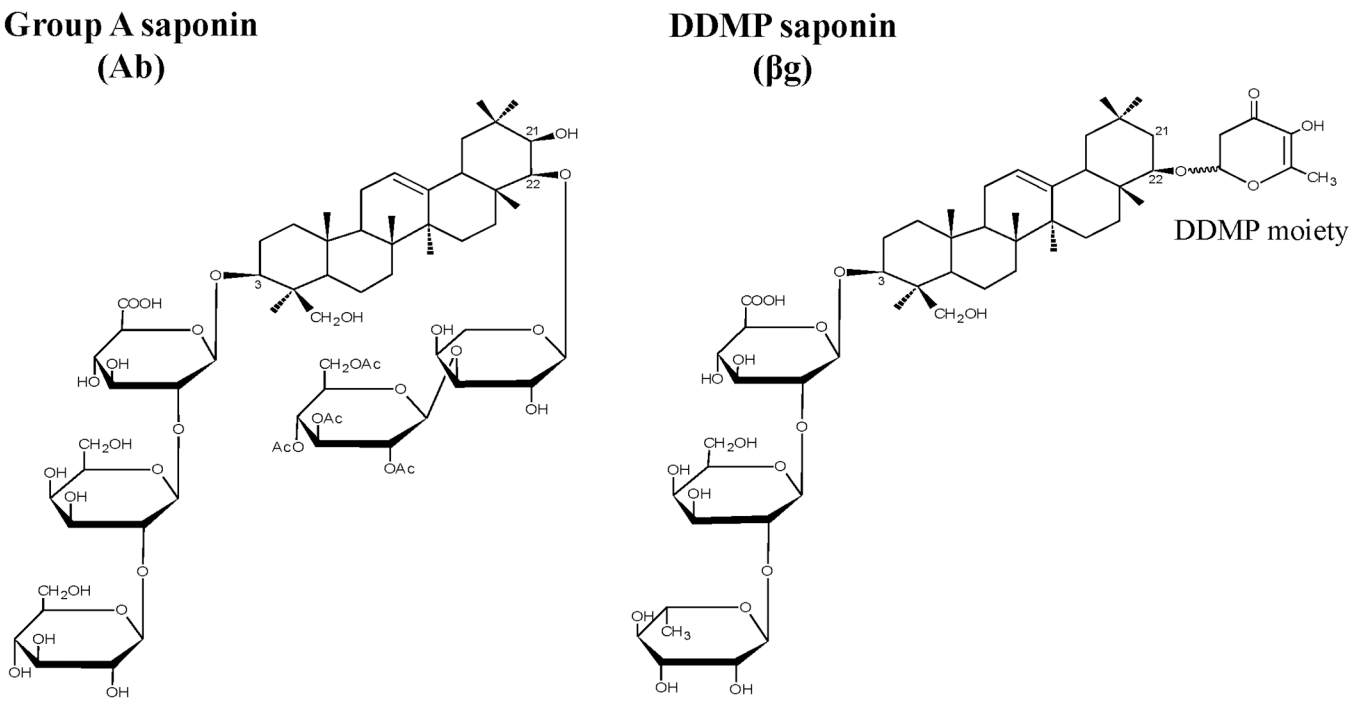

Fig 1. Chemical structure representation of group A (Ab) and DDMP (βg) triterpenoid saponins. The group A saponin (Ab) is bisdesmoside glycoside that contains two sugar chains attached at the C-3 and C-22 hydroxyl positions of soyasapogenol A. The DDMP saponin (βg) is monodesmoside glycoside that contains a single sugar chain attached to the C-3 and the 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) moiety at the C-22 hydroxyl positions of soyasapogenol B.

https://doi.org/10.1371/journal.pone.0192150.g001

(*Glyma.15g243300*; previously annotated as *Glyma15g39090* in the Phytozome database v1.1), that encodes the CYP72A69 enzyme, has been characterized in soybean [17]. The natural mutation was first identified in wild soybean, ′B01082′, which does not accumulate soyasapogenol A; therefore, biosynthesis of group A saponins does not occur [8,26]. The *sg-5* mutant in ′B01082′ was found to have a premature stop codon (L164*) in *Glyma.15g243300* [17]. Yano et al. [17] developed two induced mutations (ENT-1376, R44* and ENT-1339, S348P) in *Glyma.15g243300* that lack or contain 4-fold lower concentration of group A saponins compared to the wild-type, respectively.

Krishnamurthy et al. [33] discovered a new mutant (′CWS5095′) that contained no group A saponins among 3025 wild soybean accessions collected from nine regions of Korea. This study aimed to determine the molecular basis of a new genetic component influencing the biosynthesis of group A saponins in the new mutant, ′CWS5095′. Analysis of saponin contents in both seed hypocotyl and cotyledon showed the complete absence of group A saponins in ′CWS50595′. We found that the biosynthesis of group A saponins in ′CWS5095′ is regulated by a single recessive allele of the *Sg-5* gene, designated *sg-5a*. Among the four SNPs detected in *sg-5a*, an SNP led to the unique amino acid change, R376K, in the EXXR motif, which is absolutely conserved among CYP450s in the plant kingdom. Even though Arg and Lys possess similar chemical feature, the 3D modelled protein structure showed that the mutation R376K may inhibit the stable binding of the heme cofactor and subsequently leads to loss-of-function of Sg-5 in the ′CWS5095′ mutant. This suggests that that EXXR motif is very essential for the functional activity of the CYP450s, especially in plants.

Results

Phenotypic and genetic elucidation of a new *G. soja* mutant that lacks group A saponins

The wild soybean accession from Korea, 'CWS5095', which does not contain any group A saponins, was discovered by Krishnamurthy et al. [33]. In the present study, we confirmed the phenotype of the mutant 'CWS5095' by chromatography. The thin layer chromatography (TLC) patterns of the saponins in the hypocotyl extracts of 'CWS5095' lacked group A saponins when compared to two cultivated ('Pungsannamul' and 'Uram') and three common wild ('CW12048', 'CW16078', and 'CW13613') soybeans (Fig 2A). The LC-photodiode array (PDA)/MS/MS analysis also revealed that the hypocotyl extracts of 'CWS5095' lacked peaks with the retention time expected for group A saponins, compared to cultivar 'Pungsannamul' (Fig 2B). Furthermore, cotyledon extracts of 'CWS5095' and 'Pungsannamul' were also

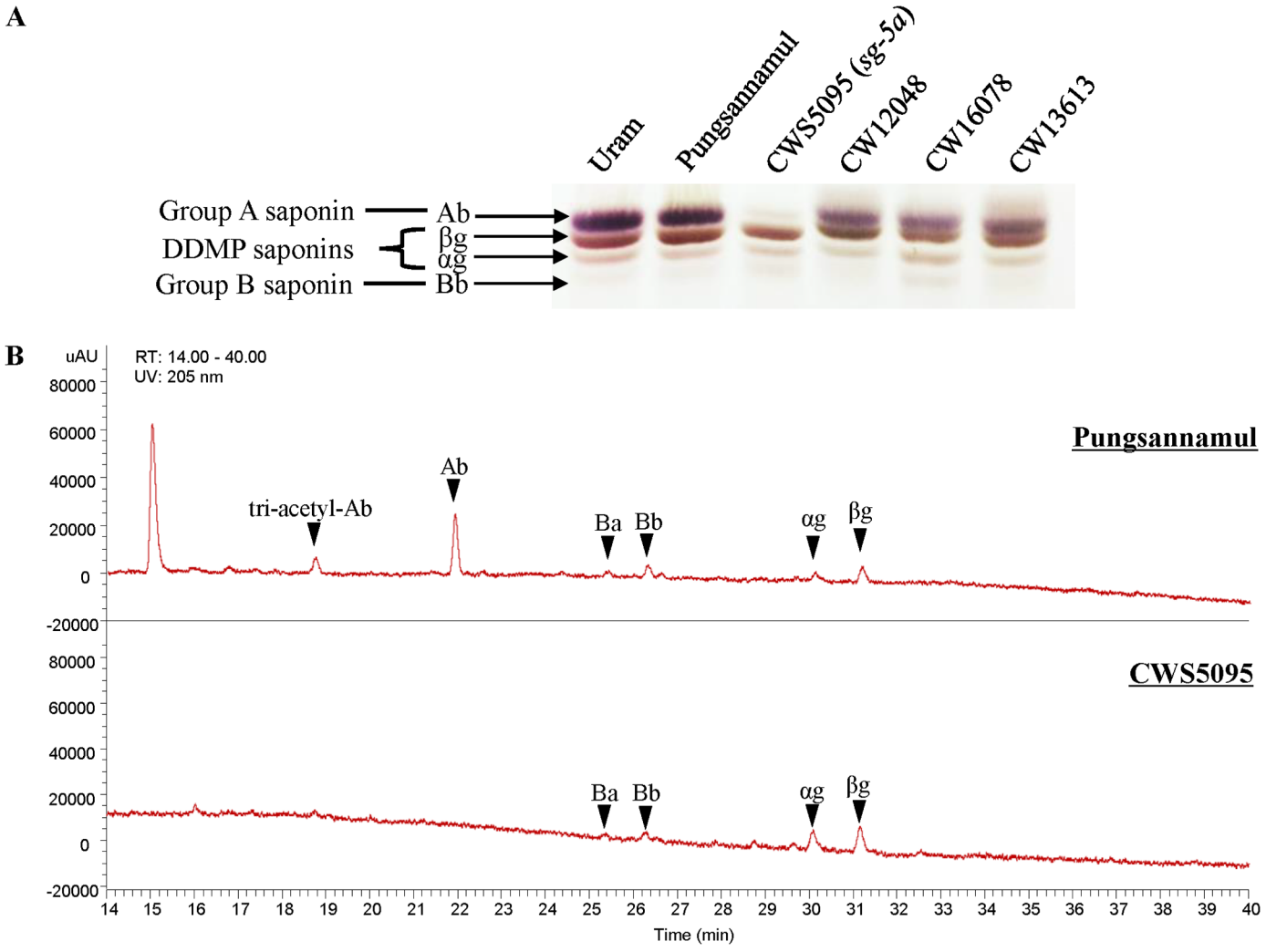

Fig 2. Separation and detection of saponins in seed hypocotyls. (A) The saponin phenotypes of wild and cultivated soybean accessions, along with new mutant, 'CWS5095' in hypocotyl extracts (in 80% aqueous methanol), were detected by thin layer chromatography (TLC). (B) Seed hypocotyl extracts (in 80% aqueous methanol) of 'Pungsannamul' (wild-type) and 'CWS5095' (mutant) were separated by high-performance liquid chromatography (HPLC) and detected by UV absorption at 205 nm. The chemical structures of saponins corresponding to each peak were identified using LC-PDA/MS/MS. The chemical structures and corresponding names of the saponins are shown in Fig 1 and S1 Fig.

https://doi.org/10.1371/journal.pone.0192150.g002

analyzed by LC-PDA/MS/MS and showed that group A saponins were absent in ′CWS5095′ (S2 Fig). These results confirm that ′CWS5095′ did not produce group A saponins.

For segregation analysis, crosses were performed between two common cultivars (′Pungsannamul′ and ′Uram′) and ′CWS5095′. A total of 120 F_2 plants from the cross between ′Uram′ and ′CWS5095′ was segregated into 96 individuals with group A saponins and 24 without group A saponins (Table 1). The segregation fitted a 3:1 ratio. A similar pattern of segregation was observed in the population of the F_3 progeny (from F_2 heterozygous plants) derived from the cross between ′Pungsannamul′ and ′CWS5095′ (Table 1). The segregation analysis indicated that a single recessive allele controlled the mutant phenotype in ′CWS5095′.

Physical mapping of the *Sg-5* locus

We performed linkage analysis using an Affymetrix Axiom® SNP array to determine the gene involved in soyasapogenol A biosynthesis in mutant ′CWS5095′. A physical map was constructed using 20 F_2 individual lines derived from crosses between ′Pungsannamul′ and ′CWS5095′, and between ′Uram′ and ′CWS5095′. The locus was mapped to a 4.3 Mb region between the Affx-89193969 and Affx-89134397 SNP arrays on chromosome 15 (E) (S3 Fig). In a previous study, the *Sg-5* locus that controls the presence/absence of group A saponins in soybean was mapped genetically; it was found to be 1.2 cM distant from the SSR marker Satt117 on chromosome 15 (E) [8]. Yano et al. [17] characterized the *Sg-5* gene (*Glyma.15g243300*). The previously identified *Sg-5* locus (*Glyma.15g243300*) was located in the region that we mapped.

Molecular analysis of the *sg-5a* allele

We analyzed coding sequence for *Glyma.15g243300* (position +1 to 1536; GenBank accession number MF624839) to determine whether *Glyma.15g243300* is responsible for the absence of group A saponins in the ′CWS5095′ mutant. The sequence analysis revealed that ′CWS5095′ contained four single-nucleotide polymorphisms (SNPs) compared to the common cultivars ′Pungsannamul′ and ′Williams 82′ (Fig 3A). The first SNP (T17A), detected in the first exon, converted valine into an aspartic acid (V6D). The second (C882G) and third (A910C) SNPs in the fourth exon led to a glutamine substitution for histidine (H294Q) and a leucine substitution for isoleucine (I304L), respectively. The fourth SNP (G1127A) was identified in the fifth exon and converted arginine into lysine (R376K).

Among the four SNPs detected in ′CWS5095′, the second and third SNPs were also detected in the coding sequences of two common *G. soja* accessions that contained group A saponins (′CW13613′ and ′CW16078′) (S4 Fig). This indicated that, among the four SNPs, the first and fourth were important mutations that might be involved in the allelic variation seen in ′CWS5095′.

Table 1. Segregation and co-segregation of F_2 individuals for group A saponin phenotypes on soybean hypocotyl.

Cross	No. of F_2 plants	F_3 population (from F_2 heterozygous plants)	Group A saponin phenotype					Genotype					
			Present	Absent	Expected	χ^2	*p* value†	W*	H*	M*	Expected	χ^2	*p* value†
Pungsannamul x CWS5095	-	304	233	71	3:1	0.22	**0.63**	71	162	71	1:2:1	0.65	**0.72**
Uram x CWS5095	120	-	96	24	3:1	0.86	**0.35**	29	67	24	1:2:1	1.07	**0.58**

†Not significant (*p* value>0.05)

*W: wild homozygote, H: heterozygote, M: mutant homozygote.

https://doi.org/10.1371/journal.pone.0192150.t001

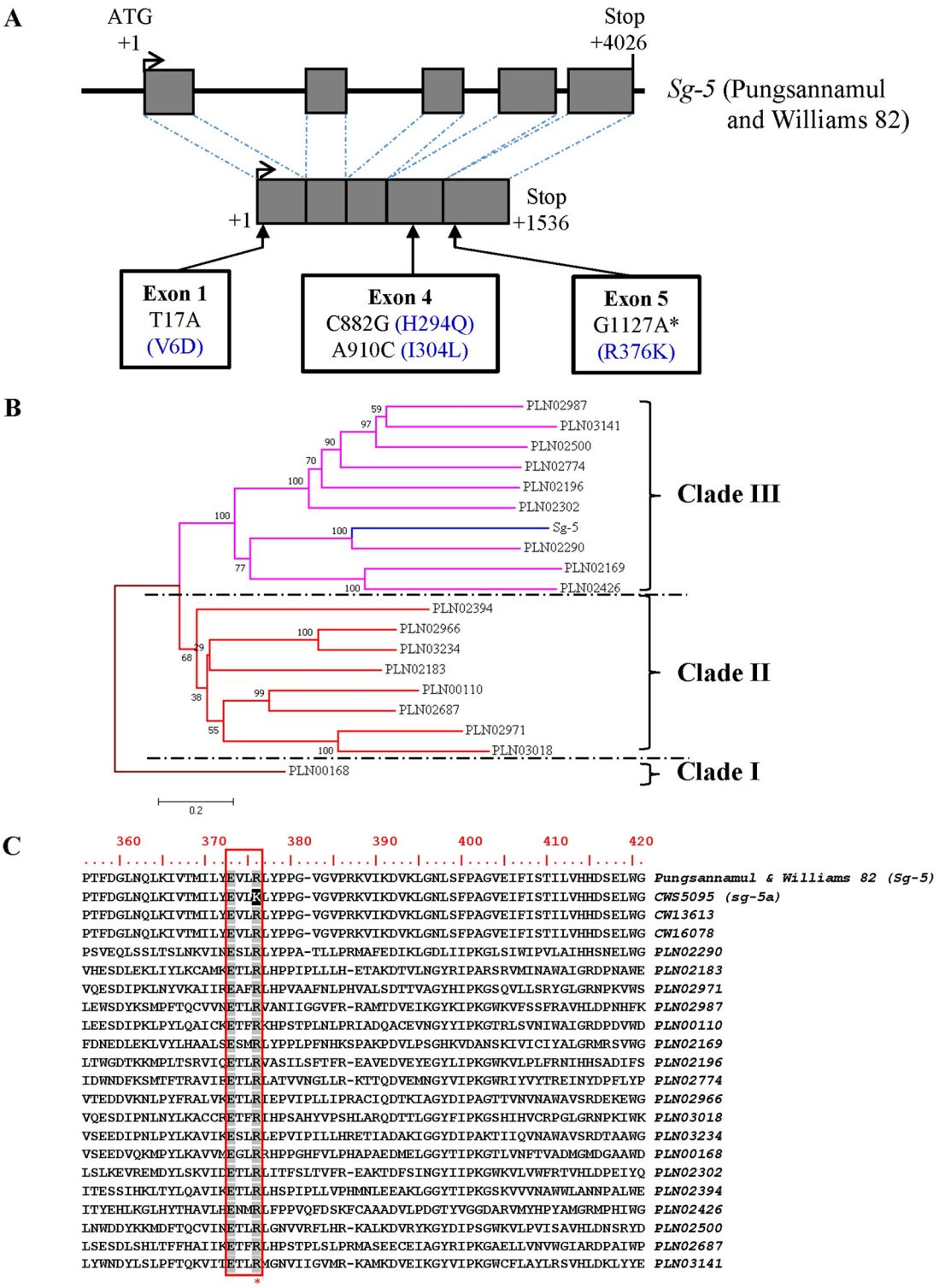

Fig 3. Gene structure of *Sg-5* (*Glyma.15g243300*) and the polymorphisms between 'Pungsannamul' and 'CWS5095'. (A) Mutational changes in the recessive alleles at the *Sg-5* locus are indicated in the mutant by the white boxes. Exons are indicated by gray boxes. The asterisk indicates the most influential SNP out of the four SNPs detected in 'CWS5095'. (B) Phylogenetic tree constructed using neighbor-joining clustering of 18 CYP450 superfamily proteins belong to different plant species. CYP450 superfamily proteins were classified into three groups (clades I to III). The Sg-5 protein was positioned in clade III and was distantly related to the clade I CYP450 superfamily. (C) Comparison of amino acid sequence Sg-5 proteins with 18 CYP450

superfamily proteins from plants (GenBank accession numbers are provided in S4 Fig). An important amino acid substitution detected in 'CWS5095' is highlighted in black and its position is marked with an asterisk. The EXXR motif is denoted by a red box and identical amino acids are shown in gray.

https://doi.org/10.1371/journal.pone.0192150.g003

CYP450s are mono-oxygenase hemoproteins that contain heme as a cofactor. In the NCBI-Conserved Domain Database (NCBI-CDD) database, CYP450 family proteins are clustered into 29 superfamilies based on the conserved domain models that generate overlapping annotations on the same protein sequence (https://www.ncbi.nlm.nih.gov/cdd). The protein sequences of 18 superfamilies registered for plant species were obtained from the NCBI-CDD database and were used to construct a phylogenetic tree (Fig 3B). Phylogenetic analysis showed that the Sg-5 protein belongs to clade III and is distantly related to the clades I and II CYP450 superfamily proteins. We performed multiple alignment analysis of the Sg-5 protein, along with the 18 CYP450 superfamilies in all three clades, to investigate the impact of the single-nucleotide substitutions on the allelic variations in 'CWS5095' (Fig 3C and S4 Fig). The results of the multiple alignments showed that the first amino acid change (V6D) was not conserved among the CYP450 superfamily proteins (S4 Fig). However, the fourth amino acid substitution (R376K) in the EXXR motif was absolutely conserved among the CYP450 family proteins in plants, animals, fungi, and bacteria (Fig 3C) [34] with a few rare exceptions in some microorganisms [35,36]. The EXXR motif is known to be involved in heme binding and the stabilization of the tertiary structure of CYP450 proteins [37].

We performed a 3D model of Sg-5 using Phyre2, a web-based 3D structure modeling program [38] to obtain further insights into the effect of the noteworthy amino acid substitutions on Sg-5 protein function. The highest score in the 3D model output for Sg-5 suggested that the crystal structure of human CYP450 3a4 was the most appropriate model protein (PDP id: 1W0E_A) [39]. Based on the structures of apo and either the substrate or inhibitor complexes, the 3D model was able to show the binding sites of the substrate, inhibitor, a cofactor heme, and the EXXR motif (Fig 4A). The detailed view of the 3D model indicated that the Arg376 that appeared at the fourth position of the EXXR motif could be a key factor that mediates the interaction between the EXXR motif and the meander loop by hydrogen bonding with Glu373, Arg430, His413, Ala423, and Glu425 (Fig 4B). Zhao et al. [40] stated that Arg376 also has a hydrophobic interaction with the side chain of Trp419 (Fig 4B). In 'CWS5095', the amino acid change Arg376Lys leads to multiple rotamers, which significantly reduces the number of hydrogen bonds (Fig 4C). In the mutant, the EXXR motif may not stably interact with the meander loop, which means that there is structural perturbation in CYP450 and inhibited heme binding. In summary, the noteworthy amino acid change (Arg376Lys) may lead to loss-of-function of Sg-5 protein. Therefore, there is no biosynthesis of group A saponins in 'CWS5095'. The mutant allele of 'CWS5095' was designated as *sg-5a*.

Co-segregation of the group A saponin phenotype with the *G. soja* mutant

A derived cleaved amplified polymorphic sequence (dCAPS) analysis was performed to determine the co-segregation pattern of the *sg-5a* allele and the group A saponin phenotype (Fig 5). A pair of dCAPS primers amplified 525-bp DNA products. The products from the wild-type control soybeans ('Pungsannamul' and 'Uram') were digested with *Mbo*I to generate 476-bp DNA fragments, whereas the products from *sg-5a* ('CWS5095') remained uncut (Fig 5B). F_2 individuals that did not produce group A saponins only contained the longer fragment, whereas F_2 individuals that did produce group A saponins either had only the shorter fragment or both the longer and shorter fragments (Fig 5B). These results indicate that the dCAPS marker co-segregated with the saponin phenotype of 120 F_2 plants derived from a cross

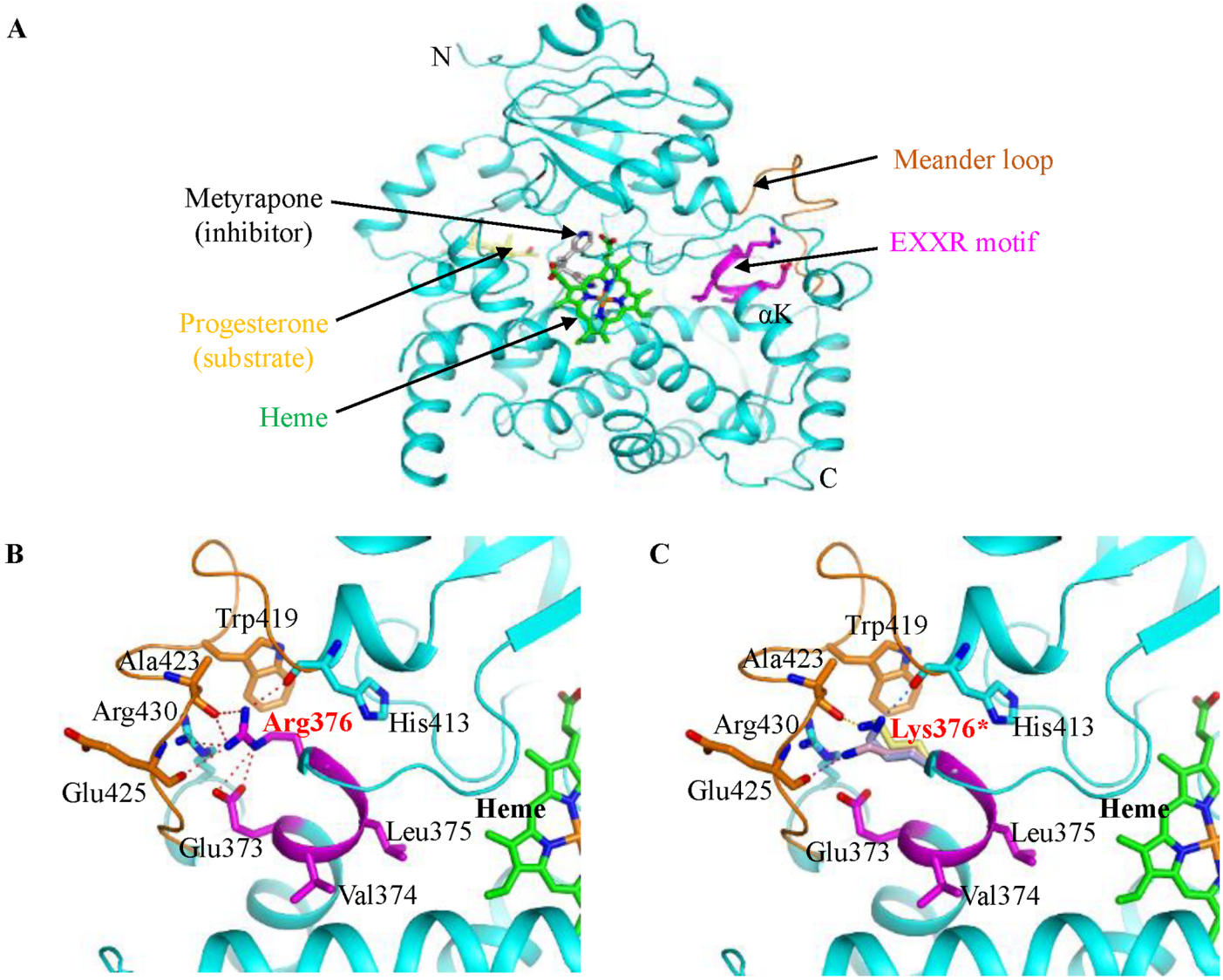

Fig 4. Three-dimensional model simulation of the amino acid substitution in the Sg-5 protein structure. (A) 3D model structure of the Sg-5 protein showing the heme cofactor (green), meander loop (brown), the EXXR motif (pink), and the αK helix that contains the EXXR motif and stabilizes heme by hydrophobic interactions. (B) Detailed view of wild-type Sg-5 protein showing that Arg376 in the EXXR motif interacts extensively with residues in the meander loop. Hydrogen bonds are shown in the side chains (red) of Glu373 and Arg430 and the main chains (red) of His413, Ala423, and Glu425. Arg376 also makes a hydrophobic interaction with the side chain of Trp419. (C) Detailed view of the mutant sg-5 protein from ′CWS5095′ showing the three rotamers of Lys376 and their corresponding hydrogen bonds in pink, light blue, and light yellow.

https://doi.org/10.1371/journal.pone.0192150.g004

between ′Uram′ and ′CWS5095′ (*sg-5a*). Similarly, the saponin phenotype co-segregated with a dCAPS marker in 304 F_3 progeny populations (from F_2 heterozygous individual) derived from a cross between ′Pungsannamul′ and ′CWS5095′. The segregation fits a 1:2:1 ratio (Table 1). These results revealed a perfect co-segregation of the *Sg-5* polymorphism with the saponin phenotype, which indicated that the *Sg-5* locus is tightly linked to group A saponin biosynthesis in ′CWS5095′ and that the novel *sg-5a* allele is recessive to *Sg-5*.

Discussion

Triterpenoid saponins are important secondary metabolites in soybean seeds. OSCs, CYP450s, and UGTs are the three main enzymes involved in the biosynthesis of saponins [3]. CYP450s

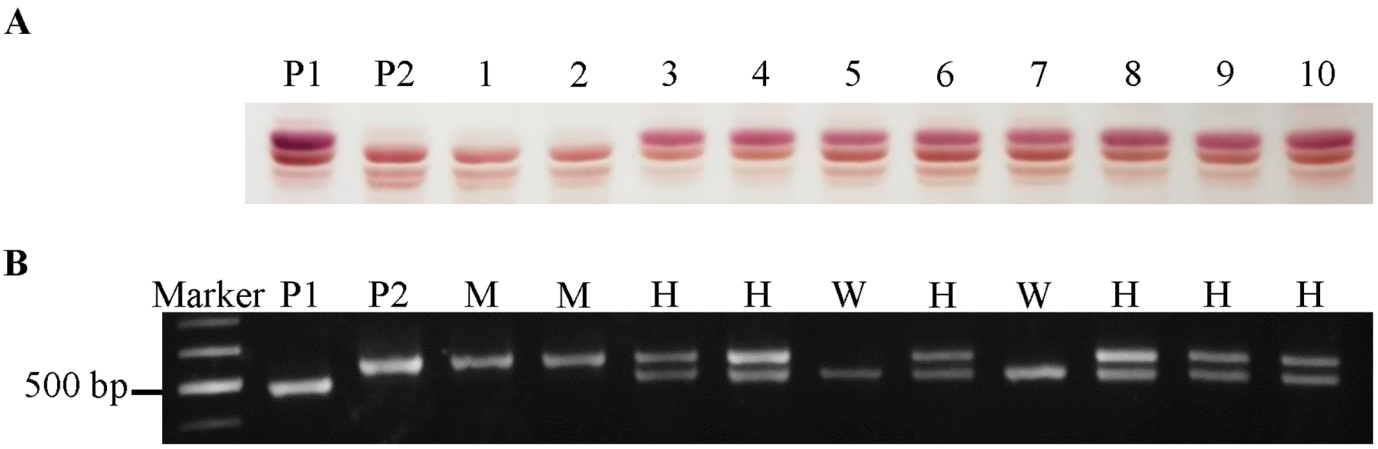

Fig 5. Co-segregation analysis of the *Sg-5* locus that controls group A saponin mutants. (A) Thin layer chromatogram showing the segregation of saponin phenotypes in the F_2 populations derived from a cross between 'Uram' and 'CWS5095' (B) A dCAPS analysis of the *sg-5a* allele showing the co-segregation of *Sg-5* polymorphism and the saponin pattern detected in the F_2 populations from a cross between 'Uram' and 'CWS5095'. P1, 'Uram'; P2, 'CWS5095'; W, wild homozygote; H, heterozygote; M, mutant homozygote.

https://doi.org/10.1371/journal.pone.0192150.g005

are mono-oxygenases that contain heme as a cofactor. In flowering plant genomes, around 300 CYP450s genes have been identified [41]. Plant CYP450s were generally classified into two major clades: A-type and non-A-type clades [42]. Majority of A-type and a few non-A-type CYP450s are involved in the biosynthesis of secondary metabolites such as terpenoids, flavonoids, alkaloids and phytoalexins [42, 43]. One of the non-A-type CYP450s, CYP72, were found to participate in pentacyclic triterpene modifications [43,44]. The CYP72 family consists of three subfamilies, CYP72A, CYP72B and CYP72C [45]. The members of the CYP72A subfamily were found to be involved in the legume-specific triterpene saponin biosynthesis [46]. To date, four CYP72As of *Medicago truncatula* (CYP72A61, CYP72A63, CYP72A67, and CYP72A68) and one CYP72A of *Glycyrrhiza uralensis* (CYP72A154) were identified to be associated in the triterpene saponin biosynthesis [32,46,47]. In soybean, CYP93E1 is first identified as an A-type CYP450 associated with the triterpenoid saponin biosynthesis, which catalyzes the hydroxylation of β-amyrin at C-24 [48]. Recently, CYP72A69, which catalyzes the hydroxylation of C-21 during the biosynthesis of soyasapogenol A, was characterized (*Glyma.15g243300*) [17]. In the present study, we obtained the new mutant accession, 'CWS5095' from the Chung wild legume germplasm collection [33,49]. Phenotypic analyses using TLC and LC-PDA/MS/MS showed that the mutant 'CWS5095' did not produce any group A saponins. Segregation analysis showed that the absence of group A saponins is controlled by a single recessive allele. The locus that controls group A saponin biosynthesis in soybean was mapped to 4.3 Mb between the Affx-89193969 and Affx-89134397 SNP arrays on chromosome 15 (E) where the previously identified gene *Glyma.15g243300* is positioned.

The CYP450s were identified by primary structure analysis of the protein sequence, particularly the two CYP450 signature motifs. They are the FXXGXRXCXG (CXG) motif in the heme-binding domain and the EXXR motif in the K-helix [50]. Although all CYP450s have a similar structural fold, they commonly have less than 20% sequence identities [51]. There are only three amino acid residues that are completely conserved (Cys in the CXG motif, and Glu and Arg in the EXXR motif) in the CYP450 superfamily proteins across almost all the biological kingdom species [34,52]. The amino acids "E" and "R" in the EXXR motif are occasionally reported as being non-conserved [35,36]. The EXXR motif is believed to involved in stabilizing

the core that is associated with the heme prosthetic group of CYP450s [37]. Mutations in either the glutamic acid or the arginine residues of the EXXR motif have led to impaired enzymatic actions in several CYP450s [53–55]. Analysis of the coding sequence for *Glyma.15g243300* in 'CWS5095' revealed the presence of four SNPs compared to the wild-type control soybeans, 'Williams 82' and 'Pungsannamul'. Among them, a significant amino acid change (R376K) was detected in the EXXR motif. The predicted 3D model for Sg-5 showed that the EXXR motif mediated the meander loop and heme. Based on the importance of the meander loop, which has a role in heme binding, EXXR should stably interact with the residues in the meander loop [37,55]. Furthermore, the Arg376 residue showed extensive interactions with several residues in the meander loop. Based on this structural feature, R376K in the mutant 'CWS5095' may lead to structural perturbation and thereby decrease binding affinity for the heme cofactor. Higashimoto et al. [56] stated that replacement of Arg with Lys in the heme oxygenase-1 caused a loss of the heme degradation activity. This suggests that even the Arg and Lys have similar chemical features, the substitution of the Arg to Lys in the EXXR motif of CYP450 that is absolutely conserved in the plant kingdom may lead to a loss-of-function and consequently no biosynthesis of group A saponins in 'CWS5095'. In addition, our results showed a perfect co-segregation of the missense mutation (R376K) in the *sg-5a* allele with the absence of group A saponins in 'CWS5095'. However, we cannot rule out the possibility that other polymorphisms may also play a subtle role in the functional activity of CYP72A69. Additionally, analysis of the enzymatic activity is necessary to confirm the association of the R376K polymorphism with absence of group A saponins.

During the past three decades, several studies for genetic improvement have been conducted to eliminate undesirable components and enhance the quality of soybean seeds [57]. Soybean seeds contain undesirable components like raffinose, stachyose, lipoxygenase, and Kunitz trypsin inhibitor. Raffinose and stachyose accumulation in soybean seeds are indigestible and cause flatulence in poultry and livestock [58]. Qiu et al. [59] reported that a mutation in the stachyose synthase gene led to an extremely low content of stachyose. The undesirable 'beany' flavor in soybean seeds is caused by oxidation of soy products resulting from seed lipoxygenase activities [60]. Mutations in the *lipoxygenase-2* gene led to a reduction in the lipoxygenase activity and thereby a better-quality soybean meal [60–62]. Kunitz trypsin inhibitors (KTi) in soybean seeds also contribute to indigestibility [63], and mutations in the *KTi* gene results in a drastic reduction in the level of Kunitz trypsin inhibitors during the seed development stage [64].

In addition to the above mentioned undesirable traits, group A saponins in soybean seeds are well-known for unpleasant taste and bitterness in soybean food [65]. To the contrary, group B and DDMP saponins are less bitter and thus more beneficial to human health [16,24]. If group A saponins are removed genetically from soybean, group B saponins are increased in compensation for the elimination of group A saponins [8]. Therefore, the elimination of group A saponins and enrichment of group B and DDMP saponins in soybean seeds are in limelight in researches to improve flavor and enhance health benefits [8]. For this purpose, we are developing an elite cultivar containing no group A saponins by a series of backcrosses to eliminate the wild genetic background of *G. soja*. Later, the newly developed elite line will be used to prove the hypothesis that the absence of group A saponins lead to an increase of DDMP and group B saponins. Inheritance of the *sg-5a* mutant allele completely corresponding to the absence of group A saponins was detected by the SNP marker that we developed in this study. Therefore, the SNP marker would serve as a useful tool for high-throughput marker-assisted selection to develop a soybean cultivar with no group A saponins.

Materials and methods

Plant materials

A new *G. soja* mutant, 'CWS5095', that does not contain any group A saponins, was discovered in Gyeonggi-do, Republic of Korea [33,49]. To isolate and identify the locus involved in group A saponin mutants, physical mapping was performed using segregating populations derived from crosses of the 'Pungsannamul' and 'Uram' cultivars with 'CWS5095'. In addition, three wild *G. soja* accessions ('CW12048', 'CW16078', and 'CW13613') were also used as wild-type controls. All the experimental populations were grown in the experimental fields at Kyungpook National University (Gunwi, 36˚07'N, 128˚38'E, Republic of Korea).

Saponin extraction and thin layer chromatography analysis of 'CWS5095'

Saponins were extracted from the hypocotyls of mature dry seeds from each accession used in this study. Saponin extractions were kept for 24 h in ten-fold volumes of (v/w) of 80% (v/v) methanol (aqueous) at room temperature. The extracts were directly used for TLC analysis or stored at 4˚C. The TLC was performed according to Krishnamurthy et al. [66], with slight modifications. Briefly, 5 μL of saponin extracts from each sample were directly loaded on pre-silica gel plates (TLC silica gel 60 F_{254}, Merck, Darmstadt, Germany) with a micropipette and air dried. The dried plates were developed in a developing chamber, which was saturated with the lower phase of a chloroform:methanol:water (65:35:10, v/v/v) mix for 50 min. The plates were dried at 90˚C for 5 min and then developed by freshly prepared H_2SO_4 for 8 min in a closed chamber. The saponins were visualized by heating at 100–110˚C for 10 min.

LC-PDA/MS/MS analysis

To obtain the detailed composition of the saponins in 'CWS5095', a LC-PDA/MS/MS on a C30 reverse phase column was performed according to Takada et al. [8]. The saponins were extracted from the hypocotyls and cotyledons of each accession by adding 10 and 20 μL of 50-fold and 7.5-fold (v/w) solutions of 80% aqueous methanol (v/v), respectively. The UV and MS spectra were analyzed using Xcalibur software version 3.1 (Thermo Fisher, Santa Clara, CA, USA).

Physical mapping of the *Sg-5* locus and sequence analysis of the *Sg-5* gene

Genomic DNAs were isolated from trifoliate leaves using the CTAB extraction method [67]. To identify the locus involved in group A saponin mutants, a physical map was constructed using an Affymetrix 180K Axiom® SNP array from the crosses between 'Pungsannamul' and 'CWS5095', and between 'Uram' and 'CWS5095'. A total of 20 F_2 individuals were selected after phenotyping of F_3 populations (mutant-type, nine; wild-type, five; heterozygote, six) for physical map construction. The *Sg-5* locus was demarked by the detection of recombinants in the F_2 individuals through Microsoft Excel software. The coding sequences of the candidate *Sg-5* gene (*Glyma.15g243300*) were amplified using the following PCR conditions: initial denaturation at 94˚C for 5 min, 35 cycles of denaturation at 94˚C for 20 s, annealing at 58˚C for 40 s, extension at 72˚C for 1 min; and a final extension at 72˚C for 5 min. The PCR products were subjected to sequencing (SolGent Co., Daejeon, Republic of Korea). The primers used for the sequencing are listed in Table 2.

Multiple alignment analysis and 3D structure modeling of the Sg-5 protein

Multiple alignment analysis was performed using the ClustalW program (http://www.genome.jp/tools-bin/clustalw). CYP450 superfamily proteins derived from the NCBI-CDD

Table 2. List of primers.

Analysis	Primer name	Sequence (5′-3′)
Sg-5	Sg-5-F1	CACGCTTTTTGCATTTATCC
	Sg-5-R1	CGAACAAAACTTGGTTCCTTG
	Sg-5-F2	AGCTCGCAAAGTTGCTCATT
	Sg-5-R2	GCTGATGTTGCTTGCAGTTG
	Sg-5-F3	CAACTGCAAGCAACATCAGC
	Sg-5-R3	GAAGGGAATGTTCTTTGATGC
	Sg-5-F4	GTGATCAAACTTATTGGATGAGC
	Sg-5-R4	TCCCGTGTTCTTCAATTTCC
	Sg-5-F5	GGTTGGTACCTAAAAGGATGAA
	Sg-5-R5	TTCATATTTCTCCACCTTATG
	Sg-5-F6	ATCATTTCCTGCTGGAGTGG
	Sg-5-R6	ATGGTGACATTCTAACTCCACAA
dCAPS	sg-5-dCAPS-F	GGTTGGTACCTAAAAGGATGAA
	sg-5-dCAPS-R	ACATCTTTGATAACTTTTCGAGGAACACCAACTCCTGGAGGGTATGAT

https://doi.org/10.1371/journal.pone.0192150.t002

(https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) were used to construct a phylogenetic tree and for multiple alignment analysis, which demonstrated the influence of the unique SNPs detected in ′CWS5095′. A 3D model of the Sg-5 protein was developed using Phyre2, a web-based 3D structure modeling program [38].

Derived cleaved amplified polymorphic sequence analysis

Genomic DNAs were isolated from all F_2 individuals derived from crosses between ′Pungsannamul′ ′CWS5095′, and between ′Uram′ and ′CWS5095′. They were used for dCAPS analysis. The PCR primers (Table 2) were designed to detect a single-base substitution at nucleotide position 1127 of *Sg-5* in ′CWS5095′. The nucleotide substitution (G to A) generates a *Mbo*I site (GATC) in the amplified PCR product from the wild parents. The PCR conditions were the same as those mentioned above. The amplified products were digested with *Mbo*I (Takara Bio Inc, Shiga, Japan) and separated on a 1.2% agarose gel.

Supporting information

S1 Fig. Chemical structure and nomenclatures of triterpene saponins in soybean. (A) Representative chemical structures of group A, DDMP, group B and group E saponins. (B) Classification of soybean saponins based on their sugar moieties or the DDMP moiety attached to the C-3 and C-22 positions. Asterisks indicate the saponins analyzed in this study.
(PDF)

S2 Fig. Separation and detection of saponin components from soybean seed cotyledons. Seed cotyledon extracts (80% aqueous methanol) of ′Pungsannamul′ (wild-type) and ′CWS5095′ (mutant) were separated by high-performance liquid chromatography (HPLC) and detected by UV absorption at 205 nm. Chemical structures of saponins corresponding to each peak were identified using LC-PDA/MS/MS analysis. Chemical structures and the corresponding names of the saponins are shown in Fig 1 and S1 Fig.
(PDF)

S3 Fig. Physical map construction of the *Sg-5* locus in soybean. (A) Mapping of the *Sg-5* locus on chromosome 15 (linkage group E) with individual F_2 lines from populations derived

from crosses between ′Pungsannamul′ and ′CWS5095′ and between ′Uram′ and ′CWS5095′. The previously mapped *Sg-5* locus along with the SSR markers (Satt117 and GMES0332) by Takada et al. (2013) was indicated. (B) Phenotype and genotype of the recombinants. The saponin phenotype of each recombinant was determined by progeny testing. Horizontal bars represent the recombinant region for each F_2 individual. Black, white and shaded bars represent, respectively, mutant homozygous, wild homozygous and wild heterozygous F_2 individuals.
(PDF)

S4 Fig. Comparison of amino acid sequence Sg-5 proteins from ′CWS5095′, two wild soybean accessions, and 18 CYP450 superfamily proteins belonging to other plant species. GenBank accession numbers and sources of CYP450 superfamily proteins are listed below. The single nucleotide changes detected in ′CWS5095′ are highlighted by yellow and black boxes. The EXXR motif is shown by a red box. The single nucleotide polymorphisms detected in ′CW13613′ are highlighted in blue.
(PDF)

Author Contributions

Conceptualization: Jagadeesh Sundaramoorthy.

Data curation: Jagadeesh Sundaramoorthy, Jong Tae Song.

Formal analysis: Jagadeesh Sundaramoorthy, Gyu Tae Park, Kyosuke Mukaiyama.

Funding acquisition: Jong Tae Song.

Methodology: Jagadeesh Sundaramoorthy, Kyosuke Mukaiyama, Jong Tae Song.

Project administration: Jong Tae Song.

Resources: Jeong-Dong Lee, Jong Tae Song.

Software: Jagadeesh Sundaramoorthy, Gyu Tae Park, Kyosuke Mukaiyama, Chigen Tsukamoto, Jeong Ho Chang.

Supervision: Jong Tae Song.

Validation: Gyu Tae Park, Chigen Tsukamoto, Jong Tae Song.

Visualization: Jagadeesh Sundaramoorthy, Chigen Tsukamoto, Jeong Ho Chang.

Writing – original draft: Jagadeesh Sundaramoorthy, Gyu Tae Park.

Writing – review & editing: Jagadeesh Sundaramoorthy, Jeong-Dong Lee, Jeong Hoe Kim, Hak Soo Seo, Jong Tae Song.

References

1. Sparg SG, Light ME, van Staden J. Biological activities and distribution of plant saponins. J Ethnopharmacol. 2004; 94: 219–243. https://doi.org/10.1016/j.jep.2004.05.016 PMID: 15325725
2. Vincken J-P, Heng L, de Groot A, Gruppen H. Saponins, classification and occurrence in the plant kingdom. Phytochemistry. 2007; 68: 275–297. https://doi.org/10.1016/j.phytochem.2006.10.008 PMID: 17141815
3. Moses T, Papadopoulou KK, Osbourn A. Metabolic and functional diversity of saponins, biosynthetic intermediates, and semi-synthetic derivatives. Crit Rev Biochem Mol Biol. 2014; 49: 439–462. https://doi.org/10.3109/10409238.2014.953628 PMID: 25286183
4. Abe I, Rohmer M, Prestwich GD. Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. Chem Rev. 1993; 93: 2189–2206. https://doi.org/10.1021/cr00022a009

5. Sawai S, Saito K. Triterpenoid biosynthesis and engineering in plants. Front Plant Sci. 2011; 2: 25. https://doi.org/10.3389/fpls.2011.00025 PMID: 22639586
6. Kahn RA, Durst F. Function and evolution of plant cytochrome P450. In: Romeo JT, Ibrahim R, Varin L, De Luca V, editors. Recent Advances in Phytochemistry: Evolution of Metabolic Pathways. Oxford: Elsevier Science Ltd; 2000. pp. 151–190. https://doi.org/10.1016/S0079-9920(00)80007-6
7. Takada Y, Sayama T, Kikuchi A, Kato S, Tatsuzaki N, Nakamoto Y, et al. 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. 2010; 60: 3–8. https://doi.org/10.1270/jsbbs.60.3
8. Takada Y, Sasama H, Sayama T, Kikuchi A, Kato S, Ishimoto M, et al. 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. https://doi.org/10.1007/s00122-012-2013-5 PMID: 23229125
9. Fenwick GR, Price KR, Tsukamoto C, Okubo K. Saponins. In: Felix D'Mello JP, Duffus CM, Duffus JH, editors. Toxic substances in crop plants. Cambridge: The Royal Society of Chemistry; 1991. pp. 285–327. https://doi.org/10.1533/9781845698454.285
10. Shiraiwa M, Harada K, Okubo K. Composition and content of saponins in soybean seed according to variety, cultivation year and maturity. Agric Biol Chem. 1991; 55: 323–331. https://doi.org/10.1080/00021369.1991.10870575
11. Kim S-L, Berhow MA, 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. https://doi.org/10.1021/jf062275p PMID: 17177534
12. Karu N, Reifen R, Kerem Z. Weight gain reduction in mice fed *Panax ginseng* saponin, a pancreatic lipase inhibitor. J Agric Food Chem. 2007; 55: 2824–2828. https://doi.org/10.1021/jf0628025 PMID: 17367157
13. Shimoyamada M, Kudo S, Okubo K, Yamauchi F, Harada K. Distributions of saponin constituents in some varieties of soybean plant. Agric Biol Chem. 1990; 54: 77–81. https://doi.org/10.1271/bbb1961.54.77
14. 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. https://doi.org/10.1271/bbb1961.55.1403
15. Szakiel A, Ruszkowski D, Janiszowska W. Saponins in *Calendula officinalis* L.–structure, biosynthesis, transport and biological activity. Phytochem Rev. 2005; 4: 151–158. https://doi.org/10.1007/s11101-005-4053-9
16. Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, et al. The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell.2012; 24: 2123–2138. https://doi.org/10.1105/tpc.111.095174 PMID: 22611180
17. Yano R, Takagi K, Takada Y, Mukaiyama K, Tsukamoto C, Sayama T, et al. 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. https://doi.org/10.1111/tpj.13403 PMID: 27775214
18. 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. https://doi.org/10.1271/bbb.56.142 PMID: 1368127
19. Konoshima T, Kokumai M, Kozuka M, Tokuda H, Nishino H, Iwashima A. Anti-tumor-promoting activities of afromosin and soyasaponin I isolated from *Wistaria brachybotrys*. J Nat Prod. 1992; 55: 1776–1778. https://doi.org/10.1021/np50090a011 PMID: 1294698
20. Okubo K, Kudou S, Uchida T, Yoshiki Y, Yoshikoshi M, Tonomura M. Soybean saponin and isoflavonoids: structure and antiviral activity against human immunodeficiency virus in vitro. In: Huang M-T, Osawa T, Ho C-, Rosen RT, editors. Food phytochemicals for cancer protection I: fruits and vegetables. Washington: American Chemical Society; 1994. pp. 330–339. https://doi.org/10.1021/bk-1994-0546.ch026
21. Sugano M, Goto S, Yamada Y, Yoshida K, Hashimoto Y, Matsuo T, et al. Cholesterol-lowering activity of various undigested fractions of soybean protein in rats. J Nutr. 1990; 120: 977–985. PMID: 2398419
22. Yoshiki Y, Kinumi M, Kahara T, Okubo K. Chemiluminescence of soybean saponins in the presence of active oxygen species. Plant Sci. 1996; 116: 125–129. https://doi.org/10.1016/0168-9452(96)04375-0
23. Kim H-Y, Yu R, Kim J-S, Kim Y-K, Sung M-K. Antiproliferative crude soy saponin extract modulates the expression of IκBα, protein kinase C, and cyclooxygenase-2 in human colon cancer cells. Cancer Lett. 2004; 210: 1–6. https://doi.org/10.1016/j.canlet.2004.01.009 PMID: 15172114

24. Okubo K, Iijima M, Kobayashi Y, Yoshikoshi M, Uchida T, Kudou S. Components responsible for the undesirable taste of soybean seeds. Biosci Biotechnol Biochem. 1992; 56: 99–103. https://doi.org/10.1271/bbb.56.99

25. Kato S, Yumoto S, Takada Y, Kono Y, Shimada S, Sakai T, et al. A new soybean cultivar 'Kinusayaka' lacking three lipoxygenase isozymes and group A acetyl saponin. Bull Natl Agric Res Cent Tohoku Reg. 2007; 107: 29–42.

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

27. Shiraiwa M, Yamauchi F, Harada K, Okubo K. Inheritance of "group A saponin" in soybean seed. Agric Biol Chem. 1990; 54: 1347–1352. https://doi.org/10.1080/00021369.1990.10870134

28. Tsukamoto C, Kikuchi A, Harada K, Kitamura K, Okubo K. Genetic and chemical polymorphisms of saponins in soybean seed. Phytochemistry. 1993; 34: 1351–1356. https://doi.org/10.1016/0031-9422(91)80028-Y PMID: 7764284

29. Kikuchi A, Tsukamoto C, Tabuchi K, Adachi T, Okubo K. Inheritance and characterization of a null allele for group A acetyl saponins found in a mutant soybean (*Glycine max* (L.) Merrill). Breed Sci. 1999; 49: 167–171. https://doi.org/10.1270/jsbbs.49.167

30. Takada Y, Tayama I, Sayama T, Sasama H, Saruta M, Kikuchi A, et al. Genetic analysis of variations in the sugar chain composition at the C-3 position of soybean seed saponins. Breed Sci. 2012; 61: 639–645. https://doi.org/10.1270/jsbbs.61.639 PMID: 23136503

31. 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. https://doi.org/10.1016/j.febslet.2010.03.037 PMID: 20350545

32. Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K, et al. Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. Plant Cell Physiol. 2013; 54: 740–749. https://doi.org/10.1093/pcp/pct015 PMID: 23378447

33. Krishnamurthy P, Tsukamoto C, Honda N, Kikuchi A, Lee J-D, Yang S-H, et al. Saponin polymorphism in the Korean wild soybean (*Glycine soja* Sieb. and Zucc.). Plant Breed. 2013; 132, 121–126. https://doi.org/10.1111/pbr.12016

34. Syed K, Mashele SS. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: Identification of evolutionarily conserved amino acid patterns characteristic of P450 family. PLoS ONE. 2014; 9: e95616. https://doi.org/10.1371/journal.pone.0095616 PMID: 24743800

35. Sezutsu H, Goff GL, Feyereisen R. Origins of P450 diversity. Phil Trans R Soc B. 2013; 368: 20120428. https://doi.org/10.1098/rstb.2012.0428 PMID: 23297351

36. Sello MM, Jafta N, Nelson DR, Chen W, Yu J-H, Parvez M, et al. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. Sci Rep. 2015; 5: 11572. https://doi.org/10.1038/srep11572 PMID: 26129850

37. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure. 1995; 3: 41–62. https://doi.org/10.1016/S0969-2126(01)00134-4 PMID: 7743131

38. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction, and analysis. Nat Protoc. 2015; 10: 845–858. https://doi.org/10.1038/nprot.2015.053 PMID: 25950237

39. Williams PA, Cosme J, Vinković DM, Ward A, Angove HC, Day PJ, et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science. 2004; 305: 683–686. https://doi.org/10.1126/science.1099736 PMID: 15256616

40. Zhao L, He F, Liu H, Zhu Y, Tian W, Gao P, et al. Natural diterpenoid compound elevates expression of Bim protein, which interacts with antiapoptotic protein Bcl-2, converting it to proapoptotic Bax-like molecule. J Biol Chem. 2012; 287: 1054–1065. https://doi.org/10.1074/jbc.M111.264481 PMID: 22065578

41. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. Plant J. 2011; 66: 194–211. https://doi.org/10.1111/j.1365-313X.2011.04529.x PMID: 21443632

42. Paquette SM, Bak S, Feyereisen R. Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol. 2000; 19: 307–317. https://doi.org/10.1089/10445490050021221 PMID: 10855798

43. Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, et al. Cytochromes P450. Arabidopsis Book. 2011; 9: e0144. https://doi.org/10.1199/tab.0144 PMID: 22303269

44. Hamberger B, Bak S. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Phil Trans R Soc B. 2013; 368: 20120426. https://doi.org/10.1098/rstb.2012.0426 PMID: 23297350

45. Werck-Reichhart D, Bak S, Paquette S. Cytochromes P450. Arabidopsis Book. 2002; 1: e0028. https://doi.org/10.1199/tab.0028 PMID: 22303202

46. Biazzi E, Carelli M, Tava A, Abbruscato P, Losini I, Avato P, et al. CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. Mol Plant, 2015; 8: 1493–1506. https://doi.org/10.1016/j.molp.2015.06.003 PMID: 26079384

47. Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, et al. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. Plant Cell. 2011; 23: 4112–4123. https://doi.org/10.1105/tpc.110.082685 PMID: 22128119

48. Shibuya M, Hoshino M, Katsube Y, Hayashi H, Kushiro T, Ebizuka Y. Identification of β-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J. 2006; 273: 948–959. https://doi.org/10.1111/j.1742-4658.2006.05120.x PMID: 16478469

49. Krishnamurthy P, Lee JM, Tsukamoto C, Takahashi Y, Singh RJ, Lee JD, et al. Evaluation of genetic structure of Korean wild soybean (*Glycine soja*) based on saponin allele polymorphism. Genet Resour Crop Evol. 2014; 61: 1121–1130. https://doi.org/10.1007/s10722-014-0095-4

50. Sirim D, Widmann M, Wagner F, Pleiss J. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. BMC Struct Biol. 2010; 10: 34. https://doi.org/10.1186/1472-6807-10-34 PMID: 20950472

51. Graham SE, Peterson JA. How similar are P450s and what can their differences teach us?. Arch Biochem Biophys. 1999; 369: 24–29. https://doi.org/10.1006/abbi.1999.1350 PMID: 10462437

52. Dorner ME, McMunn RD, Bartholow TG, Calhoon BE, Conlon MR, Dulli JM, et al. Comparison of intrinsic dynamics of cytochrome p450 proteins using normal mode analysis. Protein Sci. 2015; 24: 1495–1507. https://doi.org/10.1002/pro.2737 PMID: 26130403

53. Shimizu T, Tateishi T, Hatano M, Fujii–kuriyama Y. Probing the role of lysines and arginines in the catalytic function of cytochrome $P450_d$ by site-directed mutagenesis: Interaction with NADPH-cytochrome P450 reductase. J Biol Chem. 1991; 266: 3372–3375. PMID: 1899862

54. Chen S, Zhou D. Functional domains of aromatase cytochrome P450 inferred from comparative analyses of amino acid sequences and substantiated by site-directed mutagenesis experiments. J Biol Chem. 1992; 267: 22587–22594. PMID: 1429608

55. Hatae T, Hara S, Yokoyama C, Yabuki T, Inoue H, Ullrich V, et al. Site-directed mutagenesis of human prostacyclin synthase: Alteration of Cys^{441} of the Cys-pocket, and Glu^{347} and Arg^{350} of the EXXR motif. FEBS Lett. 1996; 389: 268–272. https://doi.org/10.1016/0014-5793(96)00600-X PMID: 8766713

56. Higashimoto Y, Sakamoto H, Hayashi S, Sugishima M, Fukuyama K, Palmer G, et al. Involvement of NADP(H) in the interaction between heme oxygenase-1 and cytochrome P450 reductase. J Biol Chem. 2005; 280: 729–737. https://doi.org/10.1074/jbc.M406203200 PMID: 15516695

57. Patil G, Chaudhary J, Vuong TD, Jenkins B, Qiu D, Kadam S, et al. Development of SNP genotyping assays for seed composition traits in soybean. Int J Plant Genomics. 2017; 2017: 6572969. https://doi.org/10.1155/2017/6572969 PMID: 28630621

58. Hagely KB, Palmquist D, Bilyeu KD. Classification of distinct seed carbohydrate profiles in soybean. J Agric Food Chem. 2013; 61: 1105–1111. https://doi.org/10.1021/jf303985q PMID: 23317449

59. Qiu D, Vuong T, Valliyodan B, Shi H, Guo B, Shannon JG, et al. Identification and characterization of a stachyose synthase gene controlling reduced stachyose content in soybean. Theor Appl Genet. 2015; 128: 2167–2176. https://doi.org/10.1007/s00122-015-2575-0 PMID: 26179337

60. Shin JH, Van K, Kim KD, Lee Y-H, Jun T-H, Lee S-H. Molecular sequence variations of the lipoxygenase-2 gene in soybean. Theor Appl Genet. 2012; 124: 613–622. https://doi.org/10.1007/s00122-011-1733-2 PMID: 22083354

61. Wang WH, Takano T, Shibata D, Kitamura K, Takeda G. Molecular basis of a null mutation in soybean lipoxygenase 2: Substitution of glutamine for an iron-ligand histidine. Proc Natl Acad Sci USA. 1994; 91: 5828–5832. PMID: 8016074

62. Reinprecht Y, Luk-Labey S-Y, Yu K, Poysa VW, Rajcan I, Ablett GR, et al. Molecular basis of seed lipoxygenase null traits in soybean line OX948. Theor Appl Genet. 2011; 122: 1247–1264. https://doi.org/10.1007/s00122-011-1528-5 PMID: 21243331

63. Wang K-J, Yamashita T, Watanabe M, Takahata Y. Genetic characterization of a novel *Tib*-derived variant of soybean Kunitz trypsin inhibitor detected in wild soybean (*Glycine soja*). Genome. 2004; 47: 9–14. https://doi.org/10.1139/g03-087 PMID: 15060597

64. Jofuku KD, Schipper RD, Goldberg RB. A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. Plant Cell. 1989; 1: 427–435. https://doi.org/10.1105/tpc.1.4.427 PMID: 2562563

65. Rupasinghe HPV, Jackson C-JC, Poysa V, Di Berardo C, Bewley JD, Jenkinson J. Soyasapogenol A and B distribution in soybean (*Glycine max* L. Merr.) in relation to seed physiology, genetic variability, and growing location. J Agric Food Chem. 2003; 51: 5888–5894. https://doi.org/10.1021/jf0343736 PMID: 13129290

66. Krishnamurthy P, Tsukamoto C, Yang SH, Lee JD, Chung G. An improved method to resolve plant saponins and sugars by TLC. Chromatographia. 2012; 75: 1445–1449. https://doi.org/10.1007/s10337-012-2340-3

67. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987; 19: 11–15.