Plant Science 267 (2018) 148–156

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

# Functional characterization of naturally occurring wild soybean mutant (*sg-5*) lacking astringent saponins using whole genome sequencing approach

Hafiz Mamoon Rehman[a], Muhammad Amjad Nawaz[a], Zahid Hussain Shah[b], Seung Hwan Yang[a], Gyuhwa Chung[a,*]

[a] Department of Biotechnology, Chonnam National University, Yeosu, Chonnam, 550-749, South Korea
[b] Department of Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

*Keywords:*
Astringent saponins
*CYP72A69*
Soybean
*G. soja*
*CYP450*

ABSTRACT

Triterpenoid saponins are one of the most highly accumulated groups of functional components in soybean (*Glycine max*) and the oxidative reactions during their biosynthesis are required for their aglycone diversity. Natural mutants of soyasaponins in wild soybean (*Glycine soja*) are valuable resources for establishing the soyasaponin biosynthesis pathway and breeding new soybean varieties. In this study, we investigated the genetic mechanism behind the absence of group A saponins in a Korean wild soybean mutant, CWS5095. Whole genome sequencing (WGS) of CWS5095 identified four point mutations [Val6 → Asp, Ile231 → Thr, His294 → Gln, and Arg376 → Lys] in *CYP72A69* (*Glyma15g39090*), which oxygenate the C-21 position of soyasapogenol B or other intermediates to produce soyasapogenol A, leading to group A saponin production. An *in vitro* enzyme activity assay of single-sited mutated clones indicated that the Arg376 > Lys mutation (a highly conserved mutation based on a nucleotide change from G → A at the 1,127th position) may lead to loss of gene function in the *sg-5* mutant. A very high normalized expression value of 377 reads per kilo base per million (RPKM) of *Glyma15g39090* in the hypocotyl axis at the early maturation seed-development stage confirmed their abundant presence in seed hypocotyls. A molecular dynamics analysis of the Arg376 > Lys mutation based on the *CYP3A4* (a human CYP450) protein structure found that it was responsible for the increase in axis length toward the heme (active site), which is critically important for biological activity and ligand binding. Our results provide important information on how to eradicate bitter and astringent saponins in soybean by utilizing the reported mutation in *Glyma15g39090*, and its importance for seed hypocotyl development based on transcript abundance.

## 1. Introduction

Naturally occurring triterpenoid saponins are the most abundant secondary metabolites in soybean seeds, which consist of group A and DDMP (2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one) saponins, based on their aglycone structure [1]. The aglycone moieties of group A and DDMP saponins consist of soyasapogenol A and B types, respectively. DDMP saponins (soyasapogenol B and E), are highly desirable because of their health-promoting activities, such as the prevention of dietary hypercholesterolemia, and anti-colon-cancer and liver-protective activities, unlike group A saponins, which are bitter in taste and are undesirable in soybean [2–5]. Previously, our lab found that soyasaponins Aa and Ab exert an anti-obesity effect on 3T3-L1 adipocytes through the downregulation of adipogenesis-related transcription factor peroxisome proliferator-activated receptor γ (PPARγ) [6]. Although saponins are largely produced in plants as natural products, their biological functions remain unknown. They play very important roles in plant defense mechanisms against pathogens, pests, and herbivores due to their antimicrobial, antifungal, antiparasitic, insecticidal, and antifeedant properties [7–10]. During normal growth and development, many plants synthesize and accumulate saponins [11].

Group A saponins have two sugar chains on the C-3 and C-22 hydroxyl groups, while DDMP saponins contain only one sugar chain at C-3 and a DDMP residue on the C-22 hydroxyl group in their aglycones [12]. Cyclization and oxidation of 2,3-oxidosqualene *via* β-amyrin and cytochrome P450 (CYP450) enzymes produced both soyasapogenol A and B [1,13]. The presence of the C-21 hydroxyl group is highly specific to soyasaponin A, and is considered critical for soyasapogenol A biosynthesis [4]. Previously, *CYP93E1* in soybean, *CYP88D6* and *CYP72A*154 in liquorice, and *CYP716A12* in *Medicago truncatula* were functionally characterized as hydroxylating β-amyrin at various

* Corresponding author.
*E-mail address:* chung@chonnam.ac.kr (G. Chung).

https://doi.org/10.1016/j.plantsci.2017.11.014
Received 17 October 2017; Received in revised form 24 November 2017; Accepted 27 November 2017
Available online 29 November 2017

positions where different types of sugar were attached by UGTs (UDP-glycosyltransferases) [14–19]. The ability of CYP450 to oxygenate β-amyrin has evolved multiple times over time [16]. Biazzi et al. [20] functionally characterized the *CYP72A67* gene, which catalyzes a key oxidative step during hemolytic saponin biosynthesis in *Medicago truncatula*. Additionally, two more CYP450 genes, *CqCYP716A78* and *CqCYP716A79*, were found to be involved in saponin biosynthesis in *Chenopodium quinoa* leaves by Jurado et al. [21].

Naturally occurring saponin mutants of wild soybean provide unique genetic resources for eradicating group A saponins, which are bitter and astringent in taste [13]. In earlier studies, several naturally available saponin mutants of Korean and Japanese origin were identified and functionally characterized. Previously, a group A acetyl saponin-deficient mutant, CWS2133, from wild soybean, was genetically characterized with a point mutation (G → A) at the +948th position in the *Sg-1* locus [22]. The group A acetyl saponin-deficient soybean mutant, PE1515, generated from the cultivar 'Pungsannamul' (containing Ab-series saponins) using ethyl methanesulfonate treatment with a single base substitution from G → A at 1004 bp of the *Sg-1* gene, introduced a premature stop codon Park et al. [41].

Group A saponins in soybean result in a bitter and astringent aftertaste. Therefore, group A acetyl saponin-deficient cultivars may solve the problem that has hindered the broader use of soybeans in the food industry. The objective of this study was to investigate a group A acetyl saponin-deficient wild soybean mutant. Identification of new mutants with new mutation sites that reduce the bitter taste of group A soyasaponins will help to elucidate the genetic and biochemical mechanisms behind saponin diversification in soybean. Here, we report the genetic and biochemical characterization of CWS5095 mutant with a mutation in the *CYP72A*69 gene, which are involved in the biosynthetic pathway of soyasapogenol A.

Though group A saponins are beneficial to human health, they also cause the bitter and astringent taste characteristics of soybeans, and are therefore unwanted components in the food industry. Hence, the CWS5095 mutant is a promising resource for generating new soybean cultivars with improved tastes in soybean breeding programs, especially in Korea.

## 2. Material and methods

### 2.1. Identification of the sg-5 mutant

A naturally occurring group A saponin-deficient mutant CWS5095 (*sg-5*) in wild soybean, *Glycine soja* Sieb. and Zucc., previously reported by Krishnamurthy et al. [23], was obtained from Chung's Wild Legume Germplasm Collection, Chonnam National University, South Korea. Seed coats of the group A saponin-deficient mutant CWS5095 and normal type CWS4792 were removed from mature, dry seeds with a separation of hypocotyl and cotyledons for saponin extraction in 10-fold volumes of 80% (v/v) methanol (aqueous) at 27 °C. Five microliters of saponin extract was directly applied and air dried on silica gel plates, and thin-layer chromatography (TLC) was performed according to Krishnamurthy et al. [23]. For the detailed composition of extracted saponins from CWS5095 and CWS4792, extracts were diluted 10 times with 80% (v/v) aqueous methanol, and analyzed using liquid chromatography photodiode array detection method (LC–PDA/MS/MS), as reported by Krishnamurthy et al. [24].

2.1.2 WGS of the *sg-5* mutant

To identify the single nucleotide polymorphisms (SNPs) in the already mapped *Sg-5* locus positioned between the GM15.45648k and GM15.45786k marker regions on chromosome 15, as reported by Yano et al. [25], we performed WGS of CWS5095 DNA with the Illumina Hiseq 4000 platform, at a depth of 30× that commercially available at Theragen Etex (http://www.theragenetex.com/bio/), South Korea. The obtained data were analyzed and compared with the William 82 reference genomic sequences (v2.0) according to Kim et al. [42]. The DNA was extracted from mature dry seeds using the same methods as in previous studies [22,26].

### 2.2. Amplification, cloning, and sequencing of sg-5 gene

The SNPs identified in *Glyma15g39090* were further verified by amplifying the coding sequences. Total RNA was extracted from immature seeds of CWS5095 using TRIzol™ reagent (Invitrogen, USA) and reverse transcribed to cDNA according to the instructions of the MLV reverse transcriptase kit (Invitrogen, USA). The following primer pair and PCR conditions were used to amplify the coding sequences of *Glyma15g39090*: *Sg-5*F TGTGCATAGCAATGGAGATTC and *Sg-5*R CAAAATAGTTGATGACACAAACAAG; an initial denaturation at 95 °C for 4 min, 35 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 35 s, extension at 72 °C for 1.7 min, and a final extension at 72 °C for 7 min. Targeted amplicons were visualized on 0.8% agarose gel, and eluted using the AccuPrep Gel purification kit (Bioneer Inc., Daejeon, Korea). Eluted amplicons were cloned into a pGEM-T easy vector system (Promega, USA), and transformants were selected on ampicillin-resistant medium. The AccuPrep Plasmid isolation kit (Bioneer Inc., Daejeon, South Korea) was used to extract plasmids, and amplified products were sequenced *via* the BigDye™ terminator method in an ABI 3730xl DNA analyzer (Bioneer, Inc. Daejeon, South Korea) with enlisted sequencing primers (Table S1).

### 2.3. Alignment and phylogenetic analysis of CYP72A genes

Putative *CYP72A* genes were identified in the genomes of *Arabidopsis thaliana*, *Medicago truncatula* (Barrel medic), and soybean using the UniProt BLASTP (http://www.uniprot.org/blast/) tool. Alignments were performed with the MUSCLE alignment tool in MEGA, and a neighbor-joining tree was constructed using MEGA 7.0 Kumar et al. [43].

### 2.4. Heterologous expression of single-sited mutant clones

The single-sited mutant clones were synthesized artificially using the services of General Biosystems, USA. The single-sited mutated coding sequences, RC1 (R376 > K), RC2 (Val6 > Asp), RC3 (Ile231 > Thr), and RC4 (His294 > Gln), of *Glyma15g39090* from CWS5095 were expressed individually by cloning them into a PET15-B *E. coli* expression vector using *Nde*I-BamHI endonucleases. The genes carrying individual mutation sites were synthesized from General Biosystems (https://generalbiosystems.com/), USA. The *E. coli* BL21 (DE3) component cells were transformed by adding 1 μL of plasmid into the 100 μL competent cell following heat shock treatment. The selection was based on 50 μg/mL of ampicillin in the lysogeny broth (LB) plate. Four single-sited clones were picked up and inoculated into a test tube containing 4 mL of terrific broth (TB) medium with 50 μg/mL ampicillin shaken at 220 rpm at 37 °C until the OD600 of the culture reached 0.6–0.8 (approximately 2 h). In 80 μL of 1 × phosphate-buffered saline (PBS) buffer and an additional 20 μL of 5 × loading buffer, 1 mL of culture was drawn out, centrifuged, and resuspended. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the rest of culture, with a final concentration of 0.2 mM to induce the expression of the target protein at 15 °C, shaken at 220 rpm for 16 h. The targeted proteins were analyzed using 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

### 2.5. Western blot analysis

Protein samples were run on the SDS-PAGE gel at 220 V for 50 min, and blotting was performed using a semi-dry electrophoretic transfer system with transfer conditions of 25 V for 30 mins. Mouse anti-His (1:2000) antibodies were used to detect the His-tagged *CYP72A69* protein. The detected His-tagged proteins were stained and compared